



## Learning effect of social affective prosody in Japanese by French learners

Takaaki Shochi<sup>1,3</sup>, Amandine Brousse<sup>2</sup>, Marine Guerry<sup>3</sup>, Donna Erickson<sup>4</sup>, Albert Rilliard<sup>5</sup>

<sup>1</sup> LaBRI – CNRS UMR5800, Bordeaux, France

<sup>2</sup> Bordeaux Montaigne University, Bordeaux, France

<sup>3</sup> CLLE-ERSSàB – UMR5263, Bordeaux, France

<sup>4</sup> Kanazawa Medical University, Kanazawa, Japan

<sup>5</sup> LIMSI, CNRS, Université Paris-Saclay, F-91405 Orsay

takaaki.shochi@labri.fr, marine.guerry@etu.u-bordeaux-montaigne.fr,  
ericksondonna2000@gmail.com, albert.rilliard@limsi.fr

### Abstract

In this paper, we investigate how French listeners with various levels of knowledge in the Japanese language, as well as Japanese native speakers, recognize the social affective meanings of utterances expressed during face-to-face interactions. A lexically neutral sentence consisting of 3 morae uttered with 8 different social affects by 6 native Japanese speakers during a conversation was used as stimulus for this experiment. Listeners had to recognize the expressions among the 8 possibilities. The perceptual results of the three groups of French listeners (levels 0 to 2) and of the group of native listeners are compared, to see if there may be different perceptual behaviors due to L1 or L2. Results show that L1 listeners' recognition rate was the most accurate, followed by French groups of level 2, 1 and 0. It was also found that the expression of surprise does not need any particular training, while the Japanese arrogant expression was quite difficult to be understood by French listeners. Further investigations of the visual information are suggested.

**Index Terms:** Perception, prosody, social affects, learning effect

### 1. Introduction

Expression of social affects by means of prosodic variations is described in many languages [1, 2, 3, 4, 5] and the language-specific aspect of such expressions has been stressed in such studies. As a consequence, training language learners to use such cues in dialogue interactions is important. Handbooks have been developed to tackle such aims – e.g. [3] for French. But what do foreigners perceive from a language's prosodic variation, when they don't know anything in this language, or when they are learning this language? Is there a shift in perception? Are there changes in the recognition abilities? These questions are important to deepen the understanding of what should be taught to foreign language learners vs. what they already know from their L1 language abilities.

Building on a recording paradigm targeting cross-cultural recordings [6], the paper presents results comparing the perception of 8 Japanese social affects. These 8 social affects were selected by previous research in linguistics, phonetics and psychology [1, 6, 8, 18].

These affective expressions were perceived by French listeners with or without knowledge of Japanese, and by Japanese listeners. The evolution of perception capacities in

these groups of listeners was analyzed in order to check which of these 8 social affects were perceived without knowing Japanese, and which required the French listeners to have some knowledge of Japanese to accurately recognize them.

## 2. Perception experiment

### 2.1. Corpus

Following the research done by [7] on 16 Japanese social affects, a lexically neutral sentence consisting of the 3 morae “Banana” was uttered with 8 different social affects selected with respect to their salient acoustic differences: *Admiration* (ADMI), *Arrogance* (ARRO), *Doubt* (DOUB), *Irritation* (IRRI), *Obviousness* (OBVI), *Politeness* (POLI), *Surprise* (SURP), *Walking on eggs* (WOEG). “Walking-on-eggs” was used to denote a situation corresponding, to some extent, to a situation where Japanese speakers would express “*kyoshuku*”, a Japanese-specific concept defined as “*corresponding to a mixture of suffering ashamedness and embarrassment, which comes from the speaker's consciousness of the fact his/her utterance of request imposes a burden to the hearer*” [8], p. 34.

Based on a perceptual evaluation of the 19 speakers' performances in each of these situations [7], the three best female and the three best male speakers, for each sentence in each situation, were selected. This results in 48 stimuli.

### 2.2. Listeners

4 groups of listeners participated in the experiment:

- Level 0: 15 French native speakers who have no knowledge about Japanese language (mean age = 25)
- Level 1: 18 native French learners of Japanese who studied Japanese for 3 years (mean age = 21)
- Level 2: 13 native French learners of Japanese who have spent at least one year in Japan (mean age = 24)
- Native: 21 Japanese native speakers of Tokyo dialect (mean age = 32, 42)

### 2.3. Experimental paradigm

48 stimuli (6 speakers x 8 social affective expressions) were presented in audio alone condition using a high-quality headphone in a graphical application specifically developed using livecode software. Subjects listened to each stimulus in

a random order. They were required to listen to each stimulus only once and had to select one affective expression among 8 in a forced choice. At the end of the experiment, subjects had to give feedback about their impressions on the experiment.

### 3. Results

#### 3.1. Identification rate

Figure 1 describes the comparison of the identification rate for 8 social expressions among the 4 different groups. According to this figure, native listeners' identification rates were the most accurate, followed by French learners of level 2, who spent more than 1 year in Japan.; then lesser accuracy from French learners of level 1, beginners of Japanese language; then French listeners of level 0, without any knowledge of Japanese language and culture. This result confirms the native effect [14] and the language learning effect [15] for the perception of social affective expressions. Moreover, the native listeners perceived very well most of the presented affective expressions: ADMI, DOUB, IRRI, POLI, SURP and WOEG. The identification rate for expressions of ARRO and OBVI was low in comparison with other expressions, showing confusions inside the dominant affective expressions (ARRO, OBVI, IRRI).

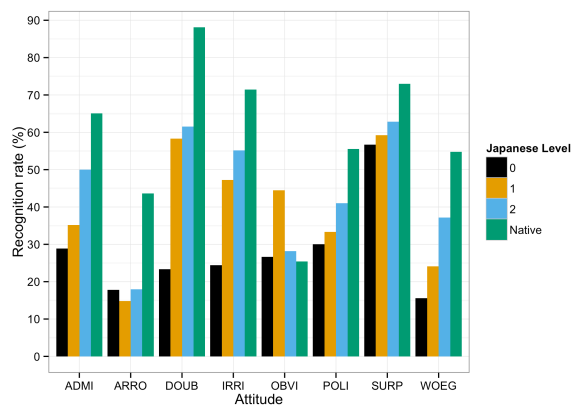


Figure 1: Recognition rate for 8 attitudinal expressions in 4 different listeners' groups (0: Fr0, 1: Fr1, 2: Fr2 and Japanese native listeners).

Listeners of level 0 showed the lowest identification rate among the 4 groups. The recognition score for all expressions was recorded as less than 30% except for SURP (57%).

Listeners of level 1 showed the higher identification rate in comparison with the listeners of level 0. Especially they showed the higher recognition for DOUB, IRRI, OBVI and SURP. However, the expression of ARRO was still difficult to be perceived. As for the group of level 0, they perceive well SURP (59%).

Finally, the listeners of level 2 showed the best identification rate among French listeners, as evidenced by an overall improvement in recognition scores. Especially, the identification rate for politeness expressions like AMDI, POLI and WOEG increased at this stage of language learning process. In addition, as for other French listeners' groups, they perceived well SURP (63%) and they showed difficulty in perceiving ARRO.

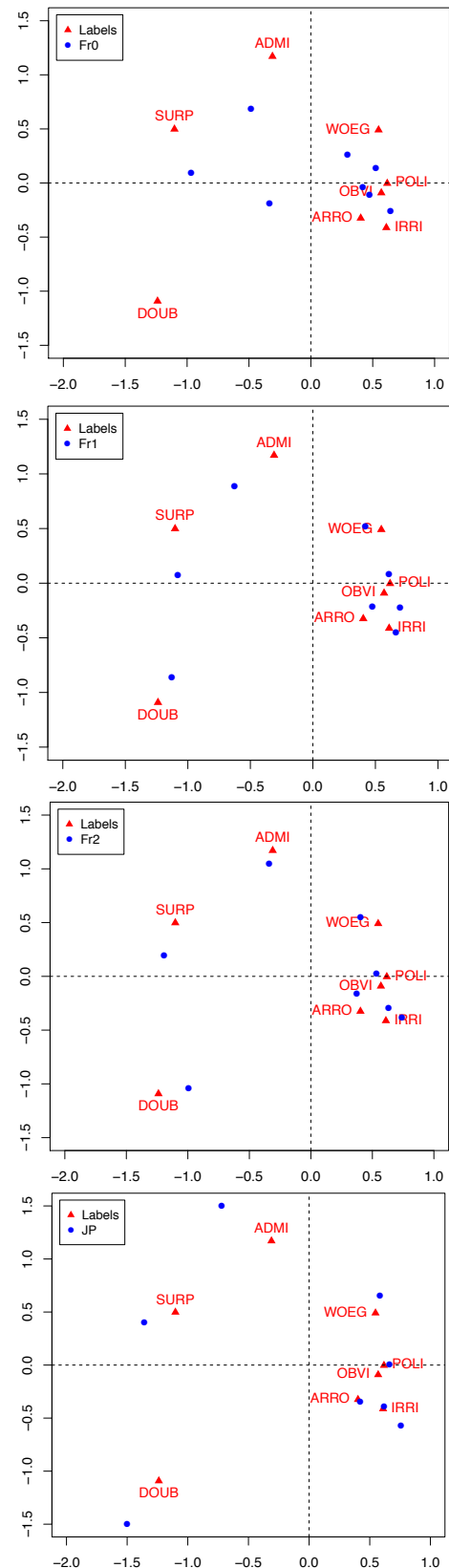


Figure 2: distribution of the distribution of presented (red point) and perceived (Blue point) 8 attitudinal expressions on 1<sup>st</sup> and 2<sup>nd</sup> dimensions in 4 different groups (from the top: Fr0, Fr1, Fr2, and Native).

### 3.2. Correspondence analysis

Figure 2 describes the distribution of perceptual points for both presented attitudes and subjects' perceptual behaviors of all 4 groups of listeners (from level 0 to native) on the 1st and 2nd dimensions extracted by Correspondence Analysis. Firstly, for the results of native listeners, all of the 8 presented social affective expressions were well perceived (i.e. the presented stimuli points were very close to the perceptual points). In addition, 3 expressions (ADMI, SURP and DOUB) were well discriminated from other affective expressions. Moreover, 2 cognitive categories (i.e. a category of politeness expressions: POLI and WOEG, and an another category of dominant expressions: ARRO, IRRI, OBVI) were identified by this psychometric analysis.

Contrary to the perceptual behaviors of native listeners, French listeners of level 0, who do not have any knowledge about Japanese language and culture showed some difficulties to perceive all of the presented 8 social affective expressions. However, the expression of SURP and ADMI were discriminated from other expressions. In addition, the other 6 affective expressions were located in the same perceptual zone showing a fusion of 2 different cognitive categories: politeness and dominant category. Unlike native listeners, the expression of DOUB was included in the category of dominant, and WOEG, which is a type of Japanese politeness expression, was perceived as IRRI. It implies clearly that French listeners perceive this politeness expression as a dominant expression [17].

Concerning the perceptual behaviors of French learners (level 1), who have no experience living in Japan more than 1 year, their identification rates for each of the presented social expressions were still low compared to the native listeners. However, ADMI, SURP and DOUB were discriminated as well as native listeners. At this learning stage, they are able to discriminate DOUB from other affective expressions. In addition, the stimuli of WOEG were generally perceived as politeness expressions. Such a difference between the listeners of level 0 and 1 seems to be related to the language learning effect. The perceptual behaviors of French learners (level 2), who have experience of living in Japan more than 1 year showed that their identification rate for each of the presented social expressions was more than French listeners of level 0 and 1, but lower than native listeners. Unlike Japanese listeners, they are still not able to distinguish the politeness category from the dominant category. Therefore all affective expressions except 3 expressions (i.e. DOUB, SURP and ADMI) are still located in the same perceptual cluster. Among the attitudes of this cluster, ARRO was not well perceived and this social expression was confused with POLI. A similar difficulty with the perception of ARRO was observed for all groups of French listeners. It may be due to the difference of cognitive notion of this affect or to the difference of conventional prosodic features for this affect in the two languages [17].

Figure 3 describes interpolated perceptual points from Fr0 to JP for the 8 investigated attitudinal expressions. According to this complementary analysis of the progression of learning effect for subjects' perceptual behavior, the perceptual points tend to spread toward the exterior of the perceptual structure as the listener's ability in Japanese language increase. This is best seen, according to the quality of representation of each attitude on the first 4 dimensions of the Correspondence Analysis by looking either at the 1<sup>st</sup> and 2<sup>nd</sup> dimension plane,

or at the 3<sup>rd</sup> and 4<sup>th</sup> dimension plane. More specifically, ADMI shows clear expanses on the 2<sup>nd</sup> dimension, DOUB along the 1<sup>st</sup> dimension, IRRI and WOEG along the 3<sup>rd</sup> dimension (in opposite directions), and POLI on the 4<sup>th</sup>. For both ARRO and OBVI, there is a no such effect of language learning on the distinctiveness of the perception.

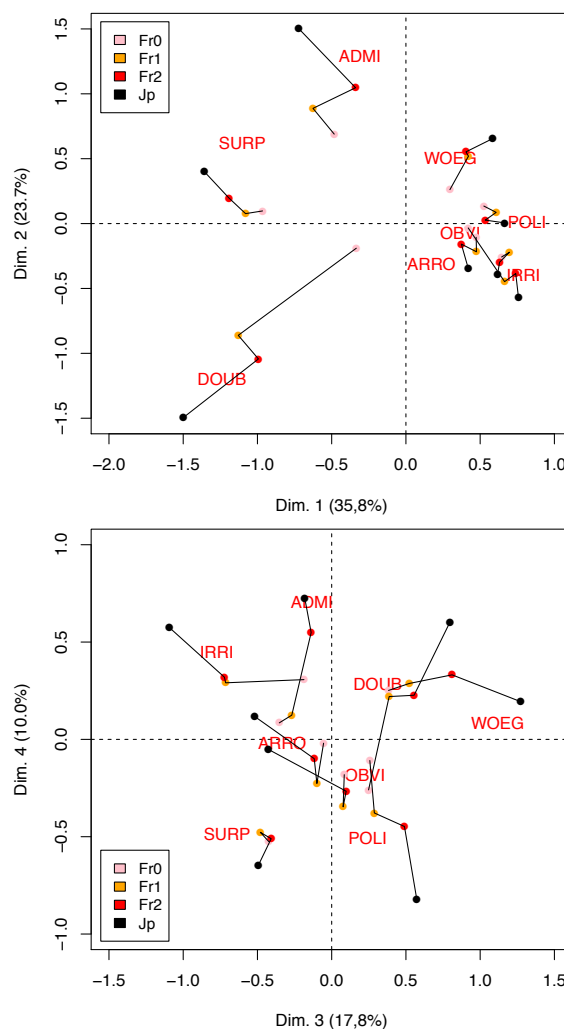


Figure 3: Mapping of the distribution of attitudes on : (top) the 1<sup>st</sup> and 2<sup>nd</sup> dimensions, and (bottom) 3<sup>rd</sup> and 4<sup>th</sup> dimensions of the CA. Each lines follow the progressive perception change from the Fr0 group up to the Japanese one (through Fr1 and Fr2). Red labels indicate the attitudes.

### 4. Conclusions

The current work investigated how French listeners with various levels of knowledge in the Japanese language, as well as Japanese native speakers, recognize the social affective meanings of utterances extracted from the social interaction database. A semantically affectively neutral utterance-type expressed in 8 various social affects by 6 native Japanese speakers who were recognized as best performers in our previous recognition experiment was used as stimuli for this experiment. Listeners had to recognize the speaker's intended

social affective expressions among the 8 possibilities. The perceptual results of the three groups of French listeners (levels 0 to 2) and of the group of native listeners are compared for possible different perceptual behaviors due to L1 or L2 knowledge.

Statistical analysis of the results indicates that L1 listeners' recognition rates were the most accurate, followed by French groups of level 2, 1 and 0. In addition, a clear progression of learning effect for listeners' perceptual behavior for the presented 8 attitudes, except for ARRO and OBVI, was identified by psychometric analysis. These results confirm the native effect and the language learning effect for the perception of social affective expressions like [14, 15]. Especially, the recognition for politeness expressions like POLI and WOEG was particularly difficult for Fr0 or Fr1 level, and it seems to be possible only at the Fr2 level of language learning process. Moreover, it was also found that the expression of surprise does not need any particular training, while the Japanese arrogant expression was quite difficult to be understood by all French listeners. Further work will consist of the investigation of the contribution of visual information to the recognition of social affects for foreign language learners.

## 5. Acknowledgements

This study has been carried out with financial support from the French State, managed by the French National Research Agency (ANR) in the frame of the "Investments for the future" Programme IdEx Bordeaux - CPU (ANR-10-IDEX-03-02), the ANR PADE Grant and the Social affEcts Discrimination Using Combined acousTic and visual informatiON project (SEDUCTION), also Japan Society for the Promotion of Science, Grants-in-Aid for Scientific Research (A) #25240026. We thank Japanese and French listeners from Waseda University and Bordeaux Montaigne University who participated in the perceptual experiment.

## 6. References

- [1] E. Uldall, "Attitudinal meanings conveyed by intonation contours", *Language and Speech*, 3(4), 223-234, 1960.
- [2] H. Fujisaki, and K. Hirose, "Analysis and perception of intonation expressing paralinguistic information in spoken Japanese", In *ESCA Workshop on Prosody*, 1993.
- [3] M. Martins-Baltar, "De l'énoncé à l'énonciation: une approche des fonctions intonatives," Paris: Didier, 1977.
- [4] J.A. de Moraes, "The pitch accents in Brazilian Portuguese: Analysis by synthesis", in *Proc. Speech Prosody 2008*, Campinas, 389-397, 2008.
- [5] W. Gu, T. Zhang, and H. Fujisaki, "Prosodic Analysis and Perception of Mandarin Utterances Conveying Attitudes", in *Proc. Interspeech*, Firenze, Italy, 1069-1072, 2011.
- [6] A. Rilliard, D. Erickson, T. Shochi, and J. A. Moraes, "Social face to face communication - American English attitudinal prosody," in *Proc. Interspeech*, Lyon, France, 2013.
- [7] D. Fourer, T. Shochi, J.-L. Rouas, J.-J. Aucouturier, and M. Guerry, "Going ba-na-nas: Prosodic analysis of spoken Japanese attitudes", in *Proc. Speech Prosody 2014*, Dublin, 2014.
- [8] T. Sadanobu, "A natural history of Japanese pressed voice", *Journal of the Phonetic Society of Japan* 8(1): 29-44, 2004.
- [9] A. Pavlenko, *Emotions and multilingualism*. Cambridge University Press, 2007
- [10] T. Shochi, D. Fourer, A. Rilliard, J.-L. Rouas, and M. Guerry, "Perceptual evaluation of spoken Japanese attitudes," in *International Congress of Phonetic Sciences (ICPhS)*, Glasgow, 2015.
- [11] T. Shochi, G. Gagné, A. Rilliard, D. Erickson, and V. Aubergé, "Learning effect of prosodic affects for Japanese learners of French language", in *Proceeding of Speech Prosody 2010*, 2010.
- [12] J.A. de Moraes, A. Rilliard, "Illocution, attitudes and prosody", in Raso, T. & Mello, H. (eds.), *Spoken Corpora and Linguistic Studies*, Amsterdam : John Benjamins, 233-270, 2014.
- [13] A. Rilliard, T. Shochi, J.-C. Martin, D. Erickson, and V. Aubergé, "Multimodal indices to Japanese and French prosodically expressed social affects". *Language and Speech* 52 (2/3): 223-243, 2009.
- [14] D. Erickson, Ohashi, S., Makita, S., Kajimoto, N., and Mokhtari, P., "Perception of naturally-spoken expressive speech by American English and Japanese listeners", *Proc. of CREST International Workshop on Expressive Speech*, 31-36, 2003.
- [15] D. Erickson, and T. Shochi., "Perception of verbal affect by Japanese children". *Transactions on Technical Committee of Psychological and Physiological Acoustics*, The Acoustical Society of Japan, Vol. 36, 5, 433-438, 2006.
- [16] A. Rilliard, D. Erickson, T. Shochi, and J. A. de Moraes, "Social face to face communication - American English attitudinal prosody," in *Proc. Interspeech*, Lyon, France, 2013.
- [17] T. Shochi, A. Rilliard, and V. Aubergé, and D. Erickson, "Intercultural perception of English, French and Japanese social affective prosody", in S. Hancil (ed.), *The role of prosody in affective speech*, *Linguistic Insights* 97, Bern: Peter Lang, AG, Bern, 31-59, 2009.
- [18] K.R. Scherer, R. Banse, and H.G. Wallbott, "Emotion inferences from vocal expression correlate across languages and cultures", *Journal of Cross-cultural psychology* 32(1), 76-92, 2001.
- [19] N.G. Thorsen, "A study of the perception of sentence intonation - evidence from Danish", *The Journal of the Acoustical Society of America* 67(3), 1014-1030, 1980.