

# PERCEPTION OF PROSODIC SOCIAL AFFECTS IN FRENCH: A FREE-LABELING STUDY

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

<sup>1</sup>CLLE ERRSàB UMR 5263, France, marine.guerry@u-bordeaux-montaigne.fr

<sup>2</sup>CLLE ERRSàB UMR 5263, LaBRI 5800, France, takaaki.shochi@labri.fr

<sup>3</sup>LIMSI-CNRS UPR3251, France, rilliard@limsi.fr

<sup>4</sup>Kanazawa Medical University, Sophia University, Japan, ericksondonna2000@gmail.com

## ABSTRACT

This work aims at using a free labeling method as a reliable way to show the complexity of the concepts induced in listeners by audiovisual prosodic attitude performances in French. The experiment was conducted on 16 French social affective expressions in audiovisual context. 27 native French listeners participated in the experiment. Subjects were asked to write down one word that best describes the expression of each stimulus. Results show that among these 16 communication situations, listeners are able to find consistent labels to name the expressivity and form coherent clusters, as far as the intended expression is concerned.

**Keywords:** multi-modal perception, social affects, free labeling

## 1. INTRODUCTION

Social affects are part of the communication process [21] and expressed through audio-visual cues [25, 19, 2]. Variations of these expressions might occur between and within languages and cultures [1]. Such expressive variations may be observed in the different (prosodic) strategies used to express the same conceptual content (e.g. doubt, irony) in two different languages, as variations may also arise from conceptual differences across cultures [29, 30].

A number of studies describe prosodic strategies used to perform given sets of social affects in various languages [26, 6, 27, 15, 14, 9], as well as a few proposing cross-cultural comparisons [24]. Understanding the concept underneath these affective expressions seems crucial when studying social affects - and is particularly difficult in cross-cultural contexts, as the description of the prosodic performances of given conceptual communicative acts

may be biased by the translation of their folk concepts [31]. Most studies on social affective prosody are based on descriptions and categorical perception tests using closed (forced) choices of concepts, and in the case of cross-cultural studies, direct translations of these concepts. Widen and Russell [28] pointed out that, in such cases, subjects "can be forced to choose a label they would not have thought of spontaneously, or would have rejected" (p.114). Several approaches are possible to avoid this problem of inaccurate translation and the biases they might incur [16, 7, 17].

The current paper proposes to use a free labeling approach (as e.g. [28, 8]) to allow subjects to express what they perceived based on the non-verbal behaviour of several speakers of French. A first aim is to study the cognitive variability of attitudinal concepts for a given set of expressions, and then to look at similarities across prosodic expressions used in different communication situations. A long term aim is to compare this free-labeling in French with similar labeling in other languages and cultures.

## 2. FREE-LABELING EXPERIMENT

### 2.1. Corpus

In a multilingual corpus of prosodic attitudes [18], social affective prosodic performances are recorded according to interactional situations to induce given expressive prosodic changes in the speakers. We used two target sentences: "*Une banane*" (a banana); "*Marie dansait*" (Mary was dancing). The interactive situations were inspired by preceding works on different languages, with some of the targeted expressivities not necessarily corresponding to conventionalized affects in all languages (and typically not in French). For instance, the attitude of *Kyoshuku*, conventionalized in the Japanese culture,

is described by [20] 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" (p. 34). This concept does not correspond directly to any convention in France (nor the in USA) [22]. Nevertheless, French speakers express themselves according to their cultural standards in a situation somewhat corresponding to the *kyoshuku* expression.

The 16 situations may relate to the following labels, that will be used for convenience in the remainder of the paper: admiration (ADMI), arrogance (ARRO), authority (AUTH), contempt (CONT), doubt (DOUB), irony (IRON), irritation (IRRI), declaration (DECL), question (QUES), obviousness (OBVI), politeness (POLI), seduction (SEDU), sincerity (SINC), surprise (SURP), uncertainty (UNCE), and "walking-on-eggs" (WOEG). The "walking on eggs" situation corresponds to Japanese *kyoshuku*, and was designed to study the expressive strategies of non-Japanese speakers [18]. Ten speakers having French as their first language (L1) were recorded (6 females / 4 males) for a total of 320 audio-visual stimuli (on the 16 situations x 2 sentences).

The performances were evaluated by 24 L1 French listeners, who were asked to rate the adequacy of each recorded behaviour in the targeted situation. The best performances of 2 female and 2 male speakers for each attitude and each target sentence were selected – i.e. each of the 16 attitudes is represented by 8 different performances to have variability among the prosodic expressions.

## 2.2. Experimental paradigm

A total of 128 stimuli were audiovisually presented in random order to 27 French native listeners (7 females / 20 males). Subjects were asked to write down (through a computer interface) one noun or adjective that best described what each stimulus expressed. Subjects were not limited in time, and were allowed to watch each stimulus several times.

A total of 3456 responses were recorded, normalized for proper analysis. In the case subjects wrote down several labels, the first was kept, as they were instructed to write only one word. Words derived from another were regrouped – e.g. "tristesse" (sadness) and "triste" (sad) are grouped under "tristesse". A total of 276 different normalized labels were obtained.

## 2.3. Data analysis

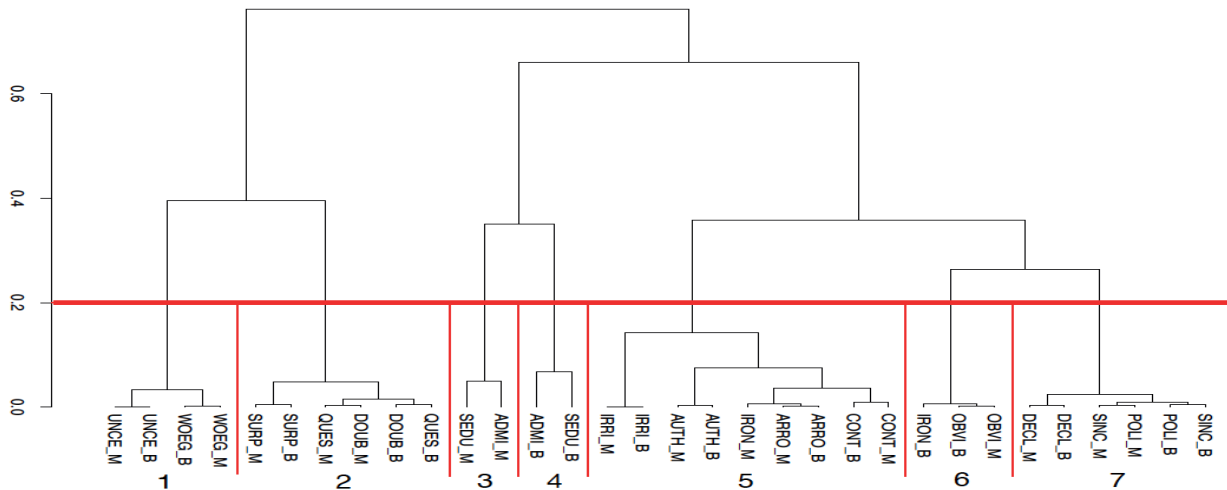
The number of times each of the 276 labels was used by subjects to describe each of the 16 situations, in each of the two target sentences, was aggregated in a matrix of 32 rows x 276 columns. This contingency matrix was analysed using a correspondence analysis (CA), in order to extract the main relations structuring the attitudes and the labels used to describe them.

The first six dimensions of the CA explain more than 50% of the variance; the first 10 dimensions about 70%. We base this analysis on the first six dimensions, in order to remove noisy data. Due to space limitation, we cannot enter the details of the CA. From these six dimensions, a hierarchical clustering algorithm was applied in order to aggregate the rows of the matrix (i.e. the 32 attitudes x sentences), according to their distribution in the space obtained from the labels [10]. The tree obtained from this clustering (cf. fig. 1) can be optimally clustered into 7 clusters (according to a criterion of reduction of the within-cluster inertia [10]).

The main separation in the tree separates clusters 1 and 2 from the others - two clusters containing dubitative speech acts (plus WOEG), as opposed to the more assertive speech acts in the second half of the tree [3]. Cluster 2 contains the neutral questions, plus the expressions of DOUB and SURP. Cluster 1 contains all expressions of UNCE plus WOEG.

The second half of the tree is based on 5 clusters with assertive speech acts, separated in four main groups. Clusters 3 and 4 mix expressions of SEDU and ADMI, respectively for the "Mary" and the "banana" sentences. Cluster 5 contains speech acts where the speakers seek to impose his view or feeling on the interlocutor (IRRI, AUTH, ARRO and CONT), plus IRON on the "Mary" sentence. Cluster 6 groups all expressions of OBVI, with expression of IRON on the "banana" sentence. Cluster 7 groups the neutral declarative sentences with two kinds of polite behaviours (POLI and SINC), that may be viewed as polite assertions.

The labels that are mostly used to describe each cluster are listed in table 1. Only labels with a percentage of occurrences above 10% are reported. This criterion allows a selection of labels shared by most listeners, as well as a removal of the peculiar ones, even if their frequency of occurrences inside the cluster may be greater than their global frequency. Note that under this 10% criterion, no label is specifically used for cluster 5. This cluster is the most complex one with four types of attitudes, plus IRON performed on the "Mary" sen-



**Figure 1:** Tree representing the hierarchical classification of the presented attitudes x sentences based on the CA.

Cluster	Label	Translation	Internal %	Global %	p value	v test
1	Hésitation	Hesitation	19,44	2,86	<0.0001	16,75
	Incertitude	Uncertainty	14,58	2,66	<0.0001	12,66
	Doute	Doubt	18,75	4,60	<0.0001	12,10
2	Surprise	Surprise	20,52	4,46	<0.0001	18,73
	Interrogation	Interrogation	16,82	3,99	<0.0001	15,70
	Etonnement	Astonishment	16,05	4,02	<0.0001	14,69
	Question	Question	12,19	2,66	<0.0001	14,13
	Doute	Doubt	10,03	4,60	<0.0001	6,56
3	Admiration	Admiration	29,17	2,11	<0.0001	17,53
4	Séduction	Seduction	14,81	1,22	<0.0001	11,58
	Envie	Longing	12,96	1,04	<0.0001	10,89
	Joie	Joy	16,20	2,69	<0.0001	9,01
5						
6	Evidence	Obviousness	48,46	7,00	<0.0001	22,82
7	Affirmation	Statement	19,29	5,70	<0.0001	14,28

**Table 1:** List of the labels used to describe each of the 7 clusters, with their English translation, their percentage of occurrence inside the cluster and globally, as well as a statistical test of the significance between both percentages.

tence. The complete list of labels used in this cluster are listed in terms of their frequency of occurrence and decreasing order of specificity: *énervement* (irritation), *agacement* (annoyance), *colère* (anger), *exaspération* (exasperation), *neutre* (neutral), *irritation* (irritation), *autoritaire* (authoritarian), *fermeté* (firmness), *désapprobation* (disapproval), *mépris* (contempt), *direct* (direct), *blasé* (blase), *agressivité* (agressiveness), *arrogance* (arrogance), *hautain* (haughty), *lassitude* (weariness), *insistance* (insistence), *sec* (curt), *mécontentement* (dissatisfaction), *frustration* (frustration), *jalousie* (jealousy), *ennui* (boredom), *réprobation* (disapproval), *contrariété* (vexation), *condescendance* (condescension), *déception* (disappointment), *dédain* (disdain), *reproche* (reproach), *impatience* (impatience), *fatigué* (tiredness), *pédant* (pedantic), *ordre* (command), *grognon*

(grumpy), *contraint* (forced), *conviction* (conviction).

### 3. DISCUSSION

Clusters 1 and 2 are linked to dubitative judgements, with cluster 1 perceived as hesitation or doubt, and cluster 2 mixing surprise, question and doubt. A cue that links both attitudes in cluster 1 is the presence of pauses in most of the performances. Cluster 2 shows mostly melodic rise at the end of the sentence, thus leading to interrogative judgements - with various illocutory strength.

Clusters 3 and 4 are separated according to the type of sentence (“Marie” or “banane”). Performances grouped in cluster 3 are mostly described as *admiration*, and it may fit the communicative situ-

ation constructs for SEDU with this sentence, that is, speakers have to praise the sexy dancing performance of Mary. The attitudinal performances grouped in cluster 4 are labeled as *seduction*, *longing* and *joy* – and are rather appropriate to describe situations where speakers have to express their strong positive desire for something (be it a banana, or the person holding that banana).

For cluster 5, most of the labels have a negative valence, and show the dominant role of the speaker. Note that authoritative (AUTH, IRR) and scornful (CONT, ARRO) expressions show strong ties in our data, even if they tend to be separated inside this cluster: the dominant aspect of these expressions is clearly expressed. Authoritative expressions are mostly valued negatively by our French listeners – which is not necessarily the case in other cultures (e.g. [24, 13, 12]).

Cluster 6 is identified as obviousness. The presence of ironic “banane” sentences (here, the irony meant the speaker obviously does not want a banana, but something more desirable) shows that if the prosodic expression of obviousness is easy to spot, its ironic value needs an interaction context to be decoded. Similarly, the ironic performance grouped with cluster 5 (on the “Marie” sentence), was performed with a negative emphasis - and this negative aspect of prosody may explain the grouping with cluster 5 negative and dominant expressions.

Expressions in cluster 7 are mostly labeled as statement. Valence does not seem to be a prominent cue in the prosodic performances presented in this cluster – except that clusters 3 and 4 seems to be positive and cluster 5, negative. However, the assertive nature of cluster 7 is clearly labeled as a dominant speech act.

#### 4. CONCLUSIONS

This work uses a free labeling method as a reliable way to show the complexity of the concepts induced in listeners by audiovisual prosodic attitude performances. It aims at developing this method as a way for cross-cultural comparisons: this experiment is a first step toward comparison of free-labeling in several languages (US English, Japanese, Brazilian Portuguese, German, and French data based are already available). Results show that among 16 communication situations, listeners can find reliable labels to name the expressivity. Listeners formed coherent clusters, in labeling the expressions intended by the performers. Among the 7 clusters, a main distinction between the linguistic speech acts of assertion and interrogation is observed. Inside the group

of interrogative expressions, a difference between hesitant and dubitative expressions is also observed. Among assertive expressions, positive and negative expressions are identified. The positive expressions address either something that is admirable or something that is desirable (the two showing some confusions). Negative expressions are strongly linked with the imposition of the speaker’s will. This is reminiscent of Culpeper’s work on impoliteness [5] seen, in line with Brown and Levinson’s theory of politeness [4], as a threat to either the positive or negative face of the listeners – which are respectively performed by authoritative and contemptuous behaviours. The expression of obviousness is also well-identified and isolated from others, and it may be a typical melodic *cliché* [6] of French prosody [11]. Finally, among simple assertions, polite expressions do not show many signs of prosodic identification in the behaviour of our French speakers. Perhaps this is also a *cliché*? The typically Japanese expression of *kyoshuku*, defined in the corpus as a situation of “walking-on-eggs” is not recognized as a specific behaviour. It is mixed with uncertainty, and labeled as such. Note that there is no lexical entry to define this expressive behaviour in French (cf. [20]). This lack of conceptual entry in the lexicon may explain partly this grouping. We need to check if the Japanese specific prosodic cues described by [23] may be also found in the performances of the French speakers, or if the French performances are recognized by Japanese listeners, having the lexical apparatus to adequately label the behaviour.

It seems crucial to examine expressiveness as a combination of features (linguistic, conceptual, prosodic, visual, gestural) in order to better understand the mechanism of expressive and perceptual behaviour. Since the free labeling shows encouraging results on French attitudinal conceptual behaviour, we intend to apply it to other languages and cultures in order to better understand the variations in labels between languages and cultures.

#### 5. ACKNOWLEDGMENTS

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) and the ANR PADE Grant. We thank French listeners from Bordeaux University who participated in the perceptual experiment.

## 6. REFERENCES

- [1] Abelin, A., Allwood, J. 2000. Cross linguistic interpretation of emotional prosody. *Proc. ISCA workshop on speech and emotion*. Newcastle, Northern Ireland 110–113.
- [2] Barkhuysen, P., Kraemer, E., Swerts, M. 2010. Cross-modal and incremental perception of audio-visual cues to emotional speech. *Language and Speech* 53, 3–30.
- [3] Brandt, P. A. 2008. Thinking and language. A view from cognitive semio-linguistics. *4th Int Conf Sp. Prosody*, Campinas, Brasil. 649–654.
- [4] Brown, P., Levinson, S. C. 1987. *Politeness: Some Universals in Language Usage*. Studies in Interactional Sociolinguistics. Cambridge University Press.
- [5] Culpeper, J., Bousfield, D., Wichmann, A. 2003. Impoliteness revisited: with special reference to dynamic and prosodic aspects. *Journal of Pragmatics* 35, 1545–1579.
- [6] Fónagy, I., Bérard, E., Fónagy, J. 1983. Clichés mélodiques. *Folia linguistica* 17(1-4), 153–186.
- [7] Goddard, C. 2007. A culture-neutral metalanguage for mental state concepts. In: Schalley, A. C., Khlentzos, D., (eds), *Mental States. Volume 2: Language and Cognitive Structure*. Amsterdam: John Benjamins Publishing Company 11–35.
- [8] Greenberg, Y., Shibuya, N., Tsuzaki, M., Kato, H., Sagisaka, Y. 2009. Analysis on paralinguistic control in perceptual impression space using multiple dimensional scaling. *Speech Communication* 51 585–592.
- [9] Gu, W., Zhang, T., Fujisaki, H. 2011. Prosodic analysis and perception of mandarin utterances conveying attitudes. *Interspeech 2011 - 12th Annual Conf of the Int Sp. Com Association*, Florence, Italy. 1069–1072.
- [10] Husson, F., Lê, S., Pagès, J. 2010. *Exploratory multivariate analysis by example using R*. CRC press.
- [11] Lu, Y. 2015. *Etude contrastive de la prosodie audio-visuelle des affects sociaux chinois mandarin vs. français vers une application pour l'apprentissage de la langue étrangère ou seconde*. PhD thesis Université Stendhal Grenoble 3.
- [12] Lu, Y., Aubergé, V., Rilliard, A. 2012. Do you hear my attitude? Prosodic perception of social affects in Mandarin. *6th Int Conf Sp. Prosody*, Shanghai, China. 685–688.
- [13] Mac, D.-K., Aubergé, V., Rilliard, A., Castelli, E. 2009. Audio-Visual prosody of social attitudes in Vietnamese: building and evaluating a tones balanced corpus. *Interspeech 2009 - 10th Annual Conf of the Int Sp. Com Association*, Brighton, UK. 2263–2266.
- [14] de Moraes, J. A. 2008. The pitch accents in Brazilian Portuguese: analysis by synthesis. *4th Int Conf Sp. Prosody*, Campinas, Brasil. 389–397.
- [15] Morlec, Y., Bailly, G., Aubergé, V. 2001. Generating prosodic attitudes in french: data, model and evaluation. *Speech Communication* 33(4), 357–371.
- [16] Osgood, C. E., May, W. H., Miron, M. S. 1975. *Cross-cultural universals of affective meaning*. University of Illinois Press.
- [17] Rilliard, A., Erickson, D., de Moraes, J. A., Shochi, T. 2014. Cross-cultural perception of some Japanese politeness and impoliteness expressions. In: Baider, F., Cislaru, G., (eds), *Linguistic Approaches to Emotions in Context*. Amsterdam: John Benjamins Publishing Company 251–276.
- [18] Rilliard, A., Erickson, D., Shôchi, T., Moraes, J. A. 2013. Social face to face communication - american english attitudinal prosody. *Proc. Interspeech* Lyon, France.
- [19] Rilliard, A., Shochi, T., Martin, J.-C., Erickson, D., Auberge, V. 2009. Multimodal indices to japanese and french prosodically expressed social affects. *Language and Speech* 52, 223–243.
- [20] Sadanobu, T. 2004. A natural history of japanese pressed voice. *Journal of the Phonetic Society of Japan* 8, 29–44.
- [21] Scherer, K. R., Ellgring, H. 2007. Multimodal expression of emotion: Affect programs or componential appraisal patterns? *Emotion* 7(1), 158.
- [22] Shochi, T. 2008. *Prosodie des affects socioculturels en japonais, français et anglais: à la recherche des vrais et faux-amis pour le parcours de l'apprenant*. PhD thesis Université Stendhal Grenoble 3.
- [23] Shochi, T., Auberge, V., Rilliard, A. 2006. How prosodic attitudes can be false friends: Japanese vs. french social affects. *3rd Int Conf Sp. Prosody*, Dresden, Germany. 692–696.
- [24] Shochi, T., Rilliard, A., Aubergé, V., Erickson, D. 2009. Intercultural perception of English, French and Japanese social affective prosody. In: Hancil, S., (ed), *The role of prosody in affective speech*. Bern, Germany: Linguistic Insights 97, 189–220.
- [25] Swerts, M., Kraemer, E. 2005. Audiovisual prosody and feeling of knowing. *Journal of Memory and Language* 81–94.
- [26] Uldall, E. 1960. Attitudinal meanings conveyed by intonation contours. *Language and Speech* 3(4), 223–234.
- [27] Wichmann, A. 2000. The attitudinal effects of prosody, and how they relate to emotion. *ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion*, Newcastle, Northern Ireland. 143–148.
- [28] Widen, S. C., Russell, J. A. 2003. A closer look at preschoolers' freely produced labels for facial expressions. *Developmental Psychology*, Vol.39 114–128.
- [29] Wierzbicka, A. 1985. Different cultures, different languages, different speech acts: Polish vs. english. *Journal of pragmatics* 9(2-3), 145–178.
- [30] Wierzbicka, A. 1992. Defining emotion concepts. *Cognitive Science* 16(4), 539–581.
- [31] Wierzbicka, A. 2005. Empirical universals of language as a basis for the study of other human universals and as a tool for exploring cross-cultural differences. *Ethos* 33(2), 265–291.