ISAPh 2016 International Symposium on Applied Phonetics
25-28 March 2016, Nagoya, Japan

Perception of prosodic social affects in Japanese by American learners of Japanese

Marine Guerry[1], Takaaki Shochi[1,2], Albert Rilliard[3], Donna Erickson[4]

[1]CLLE-ERRSaB UMR 5263, France, [2] LaBRI UMR 5800, France, [3]LIMSI, CNRS, Université Paris-Saclay, F-91405 Orsay, France, [4]Kanazawa Medical University, Japan

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

Abstract

This study investigates whether people can infer a speaker's social affective expressions in a language that is not their native language, by examining the variable lexical labels used to identify them. A free-labeling paradigm shows evidence of cultural-specificity in the perceptual behavior of listeners. Subjects were 16 US English listeners, asked to name the affective expressions carried by a set of stimuli. The results showed that their perceptual behavior was quite coherent relative to the expressions intended by the speakers. Also, despite the fact that *"kyoshuku"* is not a conventionalized expression in US culture, listeners' terms to describe it were in accordance with the intended expression. The Japanese prosodic expression of seduction was described as "happy" or "joy" by native listeners whereas "flirtatious" was a term used by US English listeners; a so-called "seductive" expression may be more conventionalized in western cultures than in Japan.

Keywords: perception, social affects, free-labeling, L2 listeners.

1. Introduction

Prosodic variations are known to convey affective expressions [1, 2]. Researchers have defended the cultural specificity of such prosodic changes, particularly by investigating language-specific affects— either being conveyed through a given cliché [3, 4], or defining a concept that is not conventionalized in other cultures [5, 6]. But such approaches are language specific, and results obtained in one language are not comparable to those in another language, mostly due to bias involved in label translation, as it has been advocated for a long time by Wierzbicka [7, 8]. Approaching the meaning encoded and perceived in prosodic attitudes beyond the use of labels, was already the goal of Uldall who, based on Osgood's semantic differential technique, observes three dimensions of prosodic meaning in US English (pleasantness, interest, dominance) [9]. Nevertheless, the method she uses is based on predefined labels, and still falls under Wierzbicka translation problem. The use of a free labeling (cf. [10]) allows for an approach of subjects' perception unbiased by such labels. This paradigm has been shown to bring adequate results about the judgments of such attitudes that match previous descriptions. Moreover, this approach allows for a finer description of the feelings evoked in subjects by the speakers' performances [11, 12].

This paper seeks to apply the same free-labeling paradigm in a cross-linguistic framework, asking US English speakers, learners of Japanese, to judge the expressivity of a set of Japanese prosodic expressions. The results will then be compared to descriptions of the same expression, done for L1 speakers of Japanese [12].

2. Material and method

2.1. Stimuli

The selected social affects were recorded on the basis of the methodology proposed in [13]. 19 L1 Japanese speakers produced two target sentences (B: "Banana"—"A banana", and M: "*Mari wa dansu wo shiteimashita*"—"Mary was dancing"), in 16 interactional situations, which are detailed in [13]: 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). Among this set of expressions, some of the expressivities are considered as culturally shared (eg. surprise, question), some might correspond to culture-specific situations, such as "walking-on-eggs" that represents a typical Japanese speech act where speakers express "*kyoshuku*" which corresponds to "*a mixture of suffering ashamedness and embarrassment, which*

10.21437/ISAPh.2016-17

comes from the speaker's consciousness of the fact his/her utterance of request imposes a burden to the hearer" ([5], p. 34) but has no lexical equivalent in English.

The corpus is a subset of a larger set of recordings from 22 L1 Japanese speakers, from which the two best performances (in each of the 2 sentences and each of the 16 expressions) among the male speakers, and among the female speakers, was selected. This represents a total of 128 stimuli: the performances by 2 females and 2 males, for the 16 attitudes based on the two target sentences.

2.2. Procedure

US English speakers, learners of Japanese language (11 females / 5 males) were asked to watch each of the 128 clips. They had to write on a computer device in their native language, one noun or one adjective that best describes what they perceived. Listeners were allowed to replay clips without time constraint. Stimuli were presented in a random order for each listener and the instructions were provided to the listeners on the computer interface.

3. Data analysis

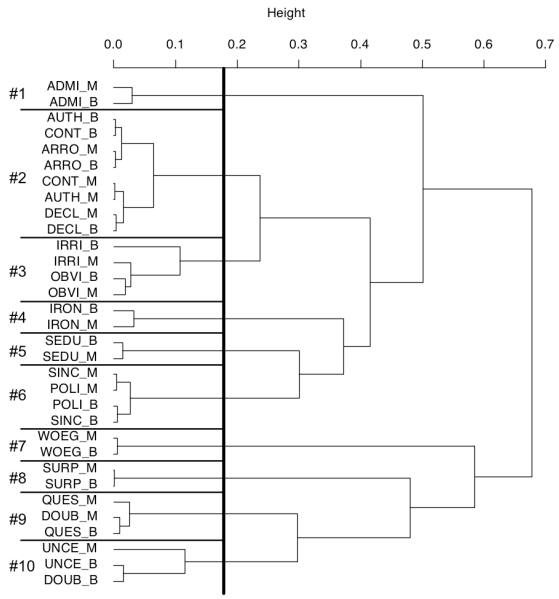
A total of 575 different labels were given by the subjects. These raw labels were reduced to 358 after correcting for spelling and keeping only the first word in those cases subjects wrote down a number of labels (as they were instructed to write only one label for each stimulus). A contingency matrix was created from this data, with 32 rows (for each of the 16 affects in each of the 2 sentences) x 358 columns (the labels). This matrix was used as the input of a Correspondence Analysis (CA) using R's FactoMineR package [14]. The CA regroups the 32 expressions (in columns) according to their descriptions by labels (in lines). The first 9 dimensions (explaining 52% of the variance) were kept according to an elbow criterion. A hierarchical clustering was applied on the resulting distribution of affective expressions (for each sentence), to analyze their distribution. The solution with 10 clusters maximizes a criterion of inertia gain [14], and was kept for this analysis.

3.1. Distribution of the expressions

The hierarchical agglomeration of expressions is represented on Figure 1. The highest separation splits the tree between clusters #1 to #6 and clusters #7 to #10: it opposes assertive and interrogative expressions—an opposition that matches two of the four enunciation modes proposed by [15] and consists of one of the main functions of prosody: the expression of modes.

Among the assertive expressions, the clusters #1 and #5 show that US English subjects do coherently identify the Japanese expressions of admiration and seduction. This differs from the labeling made by the Japanese subjects, for which these two expressions were mixed together. The expression of irony forms a unique cluster (#4) for the US English listeners, as it did for the Japanese listeners. Expressions conveying negative behaviors or impositions of the speaker are grouped in two clusters (#2 and #3) showing their perceptual proximity. However, where Japanese listeners grouped declaration with polite ones (politeness, sincerity) [11], US English listeners group it with authority, arrogance and contempt. Finally, US English listeners regroup the two polite expressions of sincerity and politeness together.

Figure 1: Dendrogram representing the hierarchical classification of the 16 presented attitudes on each of the 2 sentences, based on the CA's first 9 dimensions.



Among the interrogative expressions, US English listeners form four clusters that each contain one of the following expressions (by decreasing levels of prototypicality): "walking-on-eggs", surprise, question, uncertainty—the expression of doubt being mixed between question and uncertainty. This shows the relative similarity between these three expressions. The grouping obtained from US English listeners is also similar to that of Japanese for "walking-on-eggs" and surprise (i.e. well-clustered expressions, without confusions with other expressions, and regrouped at a higher level to an interrogative category). This is especially interesting

for “walking-on-eggs” that is a typically Japanese expression [5], while surprise, as already was shown, is cross-culturally coherent [16].

3.2. Analysis of labels

The analysis of the labels given by the US English listeners to each of these clusters gives more hints about L2 learners’ perceptual understanding of such expressions. The labels that were significantly used more often to describe each cluster are listed in Table 1 (only labels that represent more than 10% of the observed labels in each cluster are reported). Admiration is mostly described as an “excited” behaviour (20%).

Cluster #2 is a complex one with three types of affects, which present a dominant role of the speaker: arrogance, contempt, authority—plus declaration. Similar negative clusters were found in previous experiments on Japanese and French language [11, 12]. Instead of one predominant label, this complex cluster is described by this long list of terms: “neutral, bored, indifferent, aggressive, declarative, uncaring, direct, confidence, smug, annoyed, informative, apathetic, stern, hurried, authoritative, aloof, masculine, threatening, punkish, harsh, bold, tired, normal”. The expression of declaration is grouped within this specific cluster, and this differs from the grouping made by the Japanese and French L1 subjects [11,12]. The expression of declaration is mostly labeled as “neutral” and “bored” by the US English listeners; these labels being also used for the other expressions within the cluster, it explains why the algorithm regroups declaration with the negative expressions. Hence, this list of labels reveals some confusion in the listeners’ interpretation, since if all share a negative interpretation it mixes terms with various degrees of activation,.

The expressions of obviousness and irritation (cluster #3) are both predominantly described as “annoyed” (11%) which is congruent with the second intended expression. Yet, there is no occurrence of the label “obviousness” in the list. A similar description was made by the Japanese listeners. The distinction between the two expressions might be difficult to catch without information on e.g. the interaction context, or may be a matter of degree rather than quality.

Cluster #4 isolates irony for both L1 and L2 groups of subjects, but the labels used to qualify cluster #4 indicate a perceptual discrepancy between US English and Japanese subjects. For the same stimuli of irony, Japanese described it as a negative mocking whereas US English listeners do not perceive it as negative but rather, as an expression of “amusement” (13%). The following labels, used by the US subjects to describe irony, confirm this difference: “funny, humor, laughable, joking”. None of the two groups of subjects used the specific term “irony”, but the terms they used show that ironic expressions are received differently in the two cultural groups, although accurately recognized in both cases.

Cluster #5 is described as “flirtatious” (18%), “happy” (17%), “friendly” (10%): US English listeners name the expression more accurately than Japanese native listeners. Seduction is not a conventionalized expression in Japanese whereas it corresponds to a specific behavior in Western cultures [17].

US English listeners coherently label the two expressions of cluster #6 (politeness and sincerity) using terms that denote careful intentions: “concern, calm, polite, approving”—but they did not distinguish between the two.

The Japanese-specific expression of “*kyoshuku*” (“walking-on-eggs”) is clearly recognized in a single cluster by US English subjects. This cluster shares some similarities with other interrogative expressions, but it is separated from the others in terms of the labels that were used to describe it. Although this expression might not have a lexical equivalent in the US culture, the set of labels used to describe it (“regretful, uncomfortable, disgusted, embarrassed, shy, hesitant, nervous”), shows similarities with the terms used by Sadanobu to define “*kyoshuku”* (cf. the definition given above).

Cluster #8 isolates the expression of surprise and was clearly categorized as such, with about 60% of the answers being labelled “surprise”.

The last two clusters are separated as follows: cluster #9 groups question, and doubt for the “Banana” sentence; cluster #10 groups uncertainty, and doubt for the “Mary” sentence. Cluster #9 is mostly identified as “question” (22%), which shows that the Japanese prosodic interrogation is well identified by US listeners. The labels “unsure” (18%) and “confused” (15%) participated in the distribution of cluster #10. The expression of doubt seems to be blended into two different types of interrogative enunciation modes.

Table 1: List of the labels used significantly more often to describe a cluster, as compared to their global distribution – and that also represent more than 10% of the observed labels for this cluster

Cluster	Labels
#1	Excited
#2	
#3	Annoyed
#4	Amusement
#5	Flirtatious Happy Friendly
#6	
#7	
#8	Surprised
#9	Question
#10	Unsure Confused

4. CONCLUSION

The free-labelling method is a reliable way to catch the variety of meanings received by listeners from a speaker's audiovisual prosodic performances. The clustering analysis led to a distribution of the expressions into 10 clusters according to the terms used to describe the stimuli. Results show that L2 listeners were able to label Japanese expressivities with coherence. A primary distinction opposes between assertive and interrogative modes, the most classical cross-culturally shared dimension. Such a distinction is also reported in previous studies based on a forced-choice paradigm as well as on a free-labeling one [11, 12, 18]. Positive and negative (clusters #2, #3, #4) expressions are clearly separated. However, listeners show confusions between the non-affective expressions and expressions linked with imposition of the speaker's will. Japanese and US English speakers may not share the same prosodic strategies to express authoritative behavior. Hence, studies report that US English subjects' recognition rate for negative expressions is higher than Eastern subjects. Since Japanese culture is linked with the affective suppression display [19], their expressive strategies may have an impact on US English perception.

The expression of seduction by Japanese speakers was more accurately labeled and isolated by the US English listeners than by Japanese ones in their annotation. US English did coin the seductive strategy of the speakers with a term close to the actual prosodic expression's aims while native listeners didn't. This emphasized two things. First, even if Japanese speakers used an unconventional way to express seduction, they did somehow succeed to convey their intention. Second, US English listeners may have relied upon their own cultural affective experience/background in order to conceptualize this expression in a language that is not their native language. Irony also showed two different perceptual behaviors between L1 and L2 subjects: US English listeners consider this expression as a humorous one while Japanese listeners coin it as a negative mocking one. It might be an example of what Shochi [20] describes as "false-friend".

With a closer look on the terms used by US English subjects, we are able to compare their perceptual behavior with native listeners in terms of concepts and see which affects should be taught in order to enhance the pedagogical material in language teaching classes, especially for the expressions of irony and *"kyoshuku"*. The next step of this research will be cross-cultural comparisons between native and non-native listeners of French, English and Japanese. How do French interpret the typical Japanese *"kyoshuku"* expressions? How will Japanese listeners interpret French and US English displays of seduction?

6. Acknowledgements

This study has been carried out with financial support of the mobility project 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 (JSPS), Grants-in-Aid for Scientific Research (A) #25240026. We thank US English listeners of the University of Arizona who participated in the perceptual experiment.

7. REFERENCES

[1] Wichmann, A. 2000. The attitudinal effects of prosody, and how they relate to emotion. ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion, 143–148.

[2] de Moraes, J. A. 2008. The pitch accents in Brazilian Portuguese: Analysis by synthesis. Proc. 4th International Conference on Speech Prosody, Campinas, Brazil, 389–397

[3] Martins-Baltar M. 1997. De l'énoncé à l'énonciation: une approche des fonctions intonatives, Paris: Didier.

[4] Fónagy, I., Bérard, E. and Fónagy, J., 1984. Clichés mélodiques, Folia Linguistica 17,153–185

[5] Sadanobu, T. 2004. A natural history of Japanese pressed voice, Journal of the Phonetic Society of Japan 8(1), 29–44.
[6] Shochi, T., Erickson, D., Sekiyama, K., Rilliard, A., & Aubergé, V. 2009. Japanese children's acquisition of prosodic politeness expressions. Proc. Annual Conference of the International Speech Communication Association, Interspeech, 1743–1746.
[7] Wierzbicka, A. 1985. Different cultures, different languages, different speech acts: Polish vs. english. Journal of pragmatics 9(2-3), 145–178.
[8] Wierzbicka, A. 1992. Defining emotion concepts Cognitive Science 16(4), 539–581.
[9] Uldall, E. 1960. Attitudinal meanings conveyed by intonation contours. Language and Speech 3(4), 223–234.
[10] Widen, S.C. & J. A. Russell, J.A. 2003. A closer look at preschoolers' freely produced labels for facial expressions. Developmental Psychology Vol.39, 114–128.
[11] Guerry, M., Shochi, T., A. Rilliard, A., & D. Erickson,.D. 2015. Perception of prosodic social affects in french: a free-labeling study. Proc. 18th of International Congress of Phonetic Sciences, Glasgow, Scotland UK.
[12] Guerry M., Rilliard A, , Erickson D., Shochi T. 2016. Perception of prosodic social affects in Japanese: a free-labeling study. Proc. Speech Prosody 8, Boston, USA.
[13] Rilliard, A., Erickson, D., Shochi, T., Moraes, J. A. 2013. Social face to face communication - American English attitudinal prosody. Proc. Interspeech Lyon, France.
[14] Husson, F., Lê, S., Pagès, J. 2010. Exploratory multivariate analysis by example using R. CRC press.
[15] Bonaventura, P, Brandt, P. A. 2008. Sentence modes and prosodic phrasing. Proc. 4th International Conference on Speech Prosody, Campinas, Brazil, 655–660.
[16] 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.
[17] Rilliard A., Erickson D., Shochi T, &, de Moraes J. 2014. US English attitudinal prosody performances in L1 and L2 speakers. 7th International Conference on Speech Prosody, Dublin, Ireland, 895–899.
[18] 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.
[19] Butler, E. A., Lee, T. L. and Gross, J. J. 2007. Emotion regulation and culture: Are the social consequences of emotion suppression culture-specific?. Emotion, 7: 30–48.
[20] Shochi, T., Aubergé, V., & Rilliard, A. 2006. How prosodic attitudes can be false friends : Japanese vs. french social affects. Proc. Speech Prosody 3, pages 692–696, Dresden, Germany.