CROSS-LISTENING OF JAPANESE, ENGLISH AND FRENCH SOCIAL AFFECT: ABOUT UNIVERSALS, FALSE FRIENDS AND UNKNOWN ATTITUDES

Takaaki Shochi[†], Véronique Aubergé[†], Albert Rilliard[‡]

[†] GIPSA-lab, Grenoble, France - [‡]LIMSI, Orsay, France
Takaaki.Shochi@icp.inpg.fr, Veronique.Auberge@icp.inpg.fr, Albert.Rilliard@limsi.fr

ABSTRACT

Seven affectively-neutral Japanese sentences uttered with 12 different attitudes are investigated. The listeners were 15 Japanese, 15 French and 20 American. Both groups of non-native listeners had no Japanese language skill. They were asked to choose the speaker's intended attitudes among the 12. Results showed that Japanese recognized all attitudes above chance, but with some confusion, especially for the politeness expressions. However, French and American do not recognize the two cultural politeness expressions. Especially *kyoshuku*, a type of politeness that does not occur as a conventional expression in occidental society, was incorrectly decoded by French and American as *arrogance* or *irritation*.

1. INTRODUCTION

Affects in speech are expressed according to different cognitive processing levels, from involuntarily controlled expressions of emotion to the intentionally, voluntarily controlled expressions of attitudes of the speaker. Attitudes (as "labels" or "notions") are sometimes confused with Emotions (as "labels" or "notions") since both use specialized prosodic encoding in the acoustic signal. Since the prosodic emotional expressions are carried in parallel with the prosodic linguistic structures, then we can view expressions of attitudes as an integral part of "language interaction building" (i.e. communication). Some affects such as *surprise*, are classified either as attitudes or emotions. Our position is to assume that *surprise* can be an emotion if it is involuntarily processed, and can be an attitude when it is voluntarily processed and needs to be learned. When a speaker does not produce any attitude on an utterance, it must be considered a special attitude (hereafter referred to as *declaration*): the speaker doesn't intend to express any information about his intentions. We consider that studying affects in spoken interaction is crucial for expressive speech modelling, and we need to describe precisely what the attitudes of a speaker in a language and a culture can be and how they are linguistically encoded. A cross-cultural study is a helpful method to examine this problem. Some attitudes can be expected to be universal in part, both in terms of what is expressed and how it is expressed. But because attitudes are socially constructed for and by the language, they can exist in one language but not another; moreover the prosodic realization of one specific attitude in a specific language may not be recognized (or may be ambiguous) in another language. This often constitutes problems for foreign language learners. After presenting our corpus of Japanese attitudes, we will describe how listeners either Japanese, French or American perceived them.

2. CORPUS OF JAPANESE ATTITUDES

A set of 12 Japanese attitudes were selected for their representativeness according to the literature [3], [7], [8] and Japanese language teaching methods [6]: *doubt-incredulity* (DO), *evidence* (EV), *exclamation of surprise* (SU), *authority* (AU), *irritation* (IR), *arrogance* (AR), *sincerity-politeness* (SIN), *admiration* (AD), *kyoshuku* (KYO), *simple-politeness* (PO), *declaration* (DC) and *interrogation* (IN) (see [10] for definitions). Some of these attitudes are specific or specifically important for the Japanese culture, especially those linked to the politeness strategy: *simple-politeness*, *sincerity-politeness* and *kyoshuku* vs. *arrogance*. The *sincerity-politeness* attitude appears when a speaker considered inferior is talking to someone superior in the Japanese society: the speaker expresses a serious and sincere intention by using this prosodic attitude. The *kyoshuku* attitude (there is no lexical entry to translate this in English) is a typically Japanese cultural attitude. Even if such situations occur in all cultures, the Japanese language has chosen to encode this situation as an attitude ("attitudineme"). A speaker uses *kyoshuku*

when he wants to express a conflicting opinion to an interlocutor considered socially as superior – aiming at not disturbing him but to help him, or when the speaker is desiring to get a favour from his superior. It is described by [8] as "*a mixture of suffering ashamedness and embarrassment, (which) comes from the speaker's consciousness of the fact that his/her utterance of request imposes a burden to the hearer*" (p.34).

The corpus consists of seven sentences of 1, 2, 5 or 8 moras, with either a single word or a verb-object structure as the syntactic structure. For the eight-mora utterances, the lexical stress position may be on the first, second, and third mora, or absent [4]. In order to express some attitudes like *doubt* or *surprise*, the vowel [u] may be inserted at phrase final position, and the lexical stress will be realized at the seventh mora, too. The sentences were constructed in order to have no particular affective connotations in any region of Japan. Each sentence is produced with all the attitudinal functions by a male Japanese native language teacher. A total of 84 stimuli were digitally recorded at 44.1 kHz sampling rate stereo with Sony DATTCD-D7 recorder and SHURE WBH53 headset microphone.

Table 1: Corpus of Japanese attitudes – the lexical stress is marked with a star.

mora	Utterance	Translation
1	Me	The eye
2	**Na***ra	Nara
5	**Na***rade neru	He sleeps in Nara
8	**Na***goyade nomimas	He drinks in Nagoya
8	Na**ra***shide nomimas	He drinks in Nara Town
8	Matsu**ri***de nomimas	He drinks at the party
8	Naniwade nomimas	He drinks at Naniwa

3. EXPERIMENTAL PROTOCOL

In order to cross-linguistically evaluate the recorded attitudes, a computer displayed interface was developed and used amongst several groups of listeners. The test interface proposes a translation of each label in the listener's native language, and was introduced by a long oral explanation about each attitude, giving many examples of relevant situations. No subject expressed any trouble to understand the concepts referred to by the labels. All the subjects of these experiments listened to each stimulus one time only. For each stimulus, they were asked to indicate the perceived attitude amongst the twelve. The presentation order was randomized in a different order for each subject.

Listeners from three languages were tested during the experiment: Japanese, English and French native speakers. Japanese listeners were used in effect to validate the quality of the recorded attitudes; French and Americans languages were used to represent cultures and languages distant from Japanese. In all, the subjects were 15 Japanese (11 females and 4 males, Tokyo dialect, mean age = 29.5), 15 French (10 females and 5 males, level 0 in Japanese, mean age = 25.4), and 20 American listeners (12 females and 8 males, level 0 in Japanese, mean age = 24.9). No listeners participating in this experiment reported any listening disorder.

4. RESULTS

4.1. Validation with Japanese listeners

According to a chi-square test, the distribution of answers for each attitude was above chance ($p < .05$ df:11). A possible effect of the stimuli length for the distributions of selected attitudes was tested with a chi-square test. The results show a significant effect of length between two and five-mora sentences (23.1 $p < .05$ df:11) and also between five and eight-mora ones (34.5 $p < .05$ df:11). There were no effects of the lexical stress (16.5 $p > .05$ df:33).

In order to determine which attitude listeners recognized over chance, the following criterion was used: the mean identification rate must be over twice the theoretical chance level. According to this criterion, seven attitudes (i.e. *arrogance*, *declaration*, *doubt-incredulity*, *simple-politeness*, *exclamation of surprise*, *irritation* and *interrogation*) were recognized without any particular confusion.

Authority was confused with *evidence*, (e.g., feeling confident in oneself). One possible explanation is that these two attitudes may be similar, since when imposing authority, the speaker is sure of himself.

Evidence was also confused with *arrogance*. Since *evidence* shows that the speaker is confident of himself, this expression of certainty can sometimes be perceived as disrespect to the interlocutor.

The two typical Japanese attitudes of *sincerity-politeness* and *kyoshuku* were also confused with each other. These two attitudes express essentially the humility of a speaker facing a superior person in the social hierarchy. *Sincerity-politeness* was also confused with *simple-politeness*, whereas this

confusion with *simple-politeness* was not seen for *kyoshuku*.

Concerning the attitude of *admiration*, we observed confusion with *simple-politeness*. These two attitudes are interconnected in Japanese society as can be seen by the lexical polysemy of items like *sonkee* [admiration / politeness], and *keifuku* [admiration / politeness]

Table 2: Percentage of recognition of attitudes by Japanese listeners. Good answers are in bold and significant confusions in italic bold.

Perceived attitudes	Presented attitudes											
	AD	PO	KYO	SIN	AR	AU	IR	DO	EX	QS	EV	DC
Admiration	**21,9**	4,8	1,9	3,8	0,0	0,0	0,0	1,9	1,0	0,0	0,0	0,0
Politeness	26,7	**64,8**	2,9	*17,1*	1,0	4,8	0,0	0,0	0,0	8,6	1,0	11,4
Kyoshuku	11,4	9,5	**24,8**	*27,6*	0,0	0,0	0,0	0,0	0,0	1,0	0,0	0,0
Sincerity	14,3	6,7	*26,7*	**32,4**	0,0	5,7	0,0	0,0	0,0	1,9	0,0	1,0
Arrogance	1,9	0,0	1,9	0,0	**72,4**	5,7	2,9	5,7	0,0	1,0	*21,0*	10,5
Authority	0,0	0,0	15,2	1,0	9,5	**51,4**	11,4	1,0	1,0	1,9	9,5	2,9
Irritation	1,0	0,0	12,4	1,0	5,7	4,8	**85,7**	13,3	4,8	1,0	2,9	1,0
Doubt	0,0	1,0	0,0	0,0	0,0	0,0	0,0	**56,2**	14,3	3,8	0,0	0,0
Surprise	9,5	1,9	1,0	1,9	0,0	0,0	0,0	14,3	**59,0**	1,0	4,8	0,0
Interrogation	0,0	1,0	0,0	1,9	0,0	0,0	0,0	7,6	14,3	**77,1**	2,9	0,0
Evidence	8,6	1,9	10,5	3,8	5,7	17,1	0,0	0,0	5,7	1,0	**45,7**	7,6
Declaration	4,8	8,6	2,9	9,5	5,7	10,5	0,0	0,0	0,0	1,9	12,4	**65,7**

4.2. Behavior of French listeners – level 0

According to a chi-square test, the distribution of answers for all attitudes was above chance (p<.05 df:11). By the same test, a significant effect of length was identified between the one and two-mora (32.2 p<.05 df:11), and between the five and eight-mora sentences (34.5 p<.05 df:11). It was not possible to identify any significant effect of lexical stress for French subjects (39.9 p>.05 df:33).

According to the same criterion used for Japanese listeners, the results presented in figure 1 were extracted: *authority*, *irritation* and *admiration* were perceived with no significant confusion. But, the attitude of *arrogance* showed a weak identification score by French listeners. This attitude was confused with *declaration* and *authority*. French listeners did not recognize the two attitudes of politeness particular to Japanese society, namely, *sincerity-politeness* and *kyoshuku*. Moreover, *sincerity-politeness* was confused with *simple-politeness* and *kyoshuku* which represent degrees of politeness. On the contrary, the attitude of *kyoshuku* was recognized as *irritation*, *arrogance* and *authority*. The result that these attitudes of politeness were not well-recognized could be expected, since this prosodic attitude does not exist in French society, nor does the voice quality, especially of *kyoshuku* not match any politeness expression in French. French listeners also confused *interrogation* with *declaration*. This

result shows a possibility for French people to perceive Japanese interrogation as simple declaration. They show significant reciprocal confusions between *declaration* and *evidence*, between *doubt-incredulity* and *exclamation of surprise*, and between *simple-politeness* and *sincerity-politeness*.

Figure 1: Confusion graph for French listeners (level 0): percentages outside the circles indicate the confusion rate. Percentages under the labels of attitude represent the identification rates.
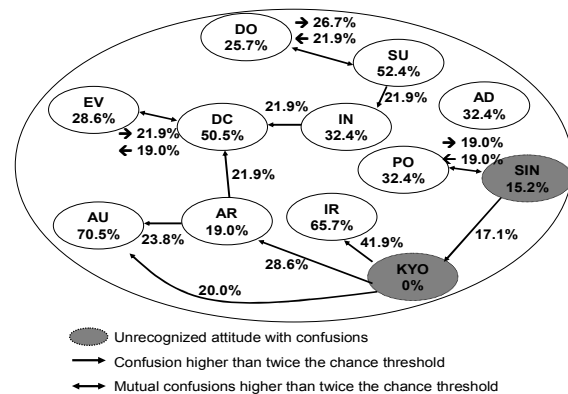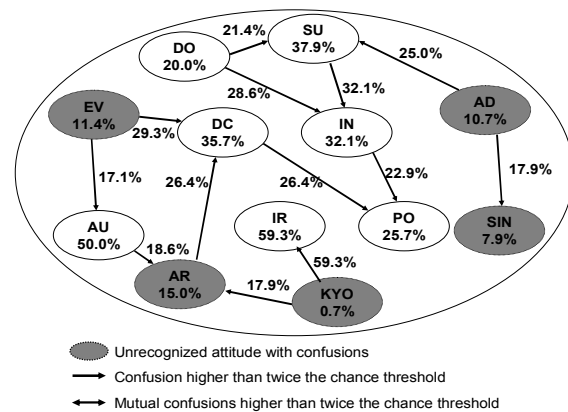

Figure 2: Confusion graph for American listeners (level 0): percentages outside the circles indicate the confusion rate. Percentages under the labels of attitude represent the identification rates.

4.3. Behavior of American listeners – level 0

The distribution of all attitudes was above chance (p<.05 df:11) by a chi-square test. A significant effect of length was identified only between the two and five-mora sentences (35.88 p<.05 df:11). No significant effect of lexical stress was identified for American subjects (40.75 p<.05 df:33).

The results showed that American listeners have a globally lower recognition score compared to French listeners. According to the confusion graph

(figure 2.), American listeners, like French listeners, did not recognize the two expressions of Japanese politeness (*kyoshuku* and *sincerity-politeness*): *kyoshuku* was recognized as *arrogance* or *irritation*. However, they did not recognize *sincerity-politeness* as an expression of politeness, whereas French listeners did. Americans recognized *arrogance* as *declaration*, and *evidence* as *authority* or *declaration*.

Unlike French listeners who recognized *admiration* well without any confusion, American listeners did not recognize this attitude but confused it with *surprise* and *sincerity-politeness*.

5. CONCLUSION

According to a chi-square test, the distribution of attitudes is above chance for main attitudes, even for French and American listeners who have no skill in Japanese. By the same statistical test, a length effect can be observed for all three groups, but in a different way (between two and five-moras, and also five and eight-moras sentences for Japanese listeners, then between one and two-mora, and also between five and eight-mora sentences for French listeners, and between two and five-mora sentences for American listeners). No effect of lexical stress was observed for any of the three groups of listeners.

Perceptual behaviour of the three groups revealed some curious perceptual differences between Japanese listeners vs. occidental listeners. The two Japanese cultural expressions of politeness (i.e. *sincerity-politeness* and *kyoshuku*) were confused within the politeness class for Japanese listeners; however, one of them *(kyoshuku)* was absolutely not recognized as a politeness expression by the occidental listeners. *Kyoshuku* is difficult to identify, even for fluent Japanese speakers as reported in [9]. This weak recognition rate for occidental subjects might come from the absence of this concept in occidental society, but may more surely come from the particular voice quality used by *kyoshuku*, a vocal fry/creak, that conveys a completely different signification (or intention) for French and American speakers. This is typically what we call a "false friend," that is, an expression whose similar acoustic characteristics lead non-native listeners to understand it according to their own language, but actually is different from what was intended by the native speaker.

In addition, only French listeners showed confusion between *interrogation* and *declaration*.

[9] compared F0 contour and intensity of these two attitudes in Japanese and French in order to find a cause of such a phenomenon but no similarity either between Japanese interrogation and declaration, nor between Japanese and French interrogation/declaration prototypes could be observed. This point must be studied with more stimuli in order to identify the causes (global contours *vs.* local stress) of ambiguities.

A perception test for Japanese listeners using a gating paradigm is under way, to ascertain how early in the utterance listeners can predict attitudinal values and to confirm that attitudinal processing involves global integration [1].

6. ACKOWLEDGEMENT

7. REFERENCES

[1] Aubergé, V., Grépillat, T., Rilliard, A. 1997. Can we perceive attitudes before the end of sentences? The gating paradigm for prosodic contours. *5th Eurospeech*, 871-874.

[2] Campbell, N. 2004. Perception of Affect in Speech – towards an Automatic Processing of Paralinguistic Information in Pro. *8th ICSLP*, Jeju, Korea, 881-884.

[3] Erickson, D., Ohashi, S., Makita, S., Kajimoto, N., Mokhtari, P. 2003. Perception of naturally-spoken expressive speech by American English and Japanese listeners. *CREST International Workshop on Expressive Speech Processing*, 31-36.

[4] Japanese accent dictionary 1960. Hirata, T. (eds), Tokyo-do shuppan (in Japanese).

[5] Maekawa, K., 1998. Phonetic and phonological characteristics of paralinguistic information in spoken Japanese. *ICSLP*, 635-638.

[6] Mizutani O., Mizutani N., 1979. *Aural Comprehension Practice in Japanese*. The Japan Times.

[7] Ofuka, E., McKeown, J.D., Waterman, M.G., Roach, P.J. 2000. Prosodic cue for rated politeness in Japanese speech. *Speech Communication*, 32, 199-217.

[8] Sadanobu, T. 2004. A natural history of Japanese pressed voice. *J. of the Phonetic Society of Japan* 8 (1), 29-44.

[9] Shochi, T., Aubergé, V., Rilliard, A., 2006. How prosodic attitudes can be false friends: Japanese vs. French social affects. *Speech Prosody*, Dresden, 692-696.