Social affect production and perception across languages and cultures – the role of prosody

Albert Rilliard

Doctor in Cognitive Sciences. LIMSI-CNRS, Orsay, France.
Albert.Rilliard@limsi.fr

Donna Erickson

Doctor in Linguistics. Showa University of Music, Kawasaki, Japan; Sofia University, Tokyo, Japan.
ericksondonna2000@gmail.com

João Antônio de Moraes

Doctor in Linguistics. Laboratório de Fonética Acústica, FL/UFRJ/CNPq, Brazil.
jamoraes3@gmail.com

Takaaki Shochi

Doctor in Linguistics. CLLE-ERSSàB, UMR 5263, Bordeaux, France.
Takaaki.Shochi@u-bordeaux3.fr

Resumo: Que recursos prosódicos são usados para codificar atitudes, ou "afetos sociais"? Como se comporta essa codificação em relação a outros códigos, como por exemplo, a acentuação? Existem codificações universais e outras específicas a determinadas culturas? Os falantes produzem atitudes similares em sua língua materna e no contexto de uma segunda língua? Essas questões serão abordadas através da revisão de alguns estudos. Medidas acústicas e experimentos perceptivos são apresentados para respaldar conclusões sobre as diferenças que se observam entre atitudes sociais e proposicionais, bem como para melhor compreender as restrições acentuais na manifestação prosódica das atitudes e para evidenciar semelhanças e diferenças entre falantes de diferentes origens culturais.

Palavras-chave: atitudes prosódicas, códigos vocais, percepção intercultural

Abstract: Which prosodic variations are used to encode attitudes or social affects? How is this prosodic code arranged in competition with other codes (e.g. accentuation)? Are there universal and culture-specific encodings? Did speakers produce attitudes in their first language similar to how they produced them in a second language context? These questions will be addressed through reviewing a few studies. Acoustic measurements and perceptual experiments are presented to support conclusions on some differences between social and propositional attitudes, to observe accentual constraints on the prosodic expression of attitudes, and to show similarities and differences between speakers of different cultural origins.

Keywords: prosodic attitudes; vocal codes; cross-cultural perception

Introduction

What WICHMANN (2000) coined "attitudes" are the prosodic expressions of affects that follow socially conventionalized codes. Such prosodic variations allow the speakers to finely tune a targeted speech act – so it may address the linguistic content of the utterance (e.g. an expression of doubt) or the addressee (e.g. an expression of politeness; about this distinction, cf. also MORAES, 2008; GU et al., 2011). The codes used for prosodic attitudes may partly be derived of phylogenetically evolved constraints, such as the proposed *frequency code* (OHALA, 1984), or other sound-symbolism relations as detailed by LÉON (1993) or GUSSENHOVEN (2004). Prosodic codes may also be based on arbitrary signs encoded in a given language for a specific purpose; such a conventionalization of prosodic expressions was felt language-specific enough to deserve to be taught in foreign language classes (cf. MARTINS-BALTARD, 1977 – see also LÉON, 1993:82ff for varying conventionalization of speech styles across cultures).

Between these two extremes – a universal *frequency code* and language-specific prosodic attitudes – how is prosody organized to express the intended meaning to the interlocutor? And how do listeners of foreign linguistic and cultural background decode (or express) such prosodic attitudes? By reviewing a few studies dealing with prosodic attitudes in cross-cultural contexts, we will try to analyze some shared strategies as well as some language-specific behavior.

A set of Brazilian Portuguese (BP) attitudes, as presented in Moraes and colleagues (MORAES et al. 2010, 2011) will allow us to describe the multimodal nature of such expressions, and will serve to emphasize the distinction between propositional and social expressions. This distinction shows the importance of language in the structure of prosodic changes – and it stresses its language-specificity. Such language specific attitudes can be found

in Japanese social affects (cf. SHOCHI, RILLIARD et al. 2009), which present culturally specific expressions based on a typical voice quality. An analysis of their perception will be useful to introduce cross-cultural perception studies and findings as well as the problems they raised. Mainly, concepts of attitudes may bias the perceptual evaluation of prosodic performances. Finally, a recording paradigm will be presented through some results obtained on U.S. English (RILLIARD et al., 2013). This paradigm is dedicated to the comparison of cross-cultural strategies for prosodic attitudes in a given set of communicative situations allowing putting speakers in comparable recording situations, whatever their linguistic origin, and thus controlling for potential conceptual differences between attitudinal labels. Results for first language and second language speakers are discussed.

1. Propositional vs. social attitudes: importance of the modality

MORAES (2008) separates attitudes addressing the propositional content of a sentence (by expressing doubt, irony, obviousness, etc.), and attitudes addressing the social relation that the speaker wants to maintain (or not) with the listener (by expressing politeness, contempt, charm, etc.). Such prosodic attitudes may have an important role in the efficiency of the speech act targeted by the speaker. Moraes moreover stresses that the illocutionary category of sentences (assertion, yes-no questions, directive, etc.) plays a concurrent role at the prosodic level – some attitudes being incompatible with a given illocution. This mostly affects the propositional attitudes: one can hardly imagine an interrogative sentence spoken with obviousness, whereas it can be polite or not.

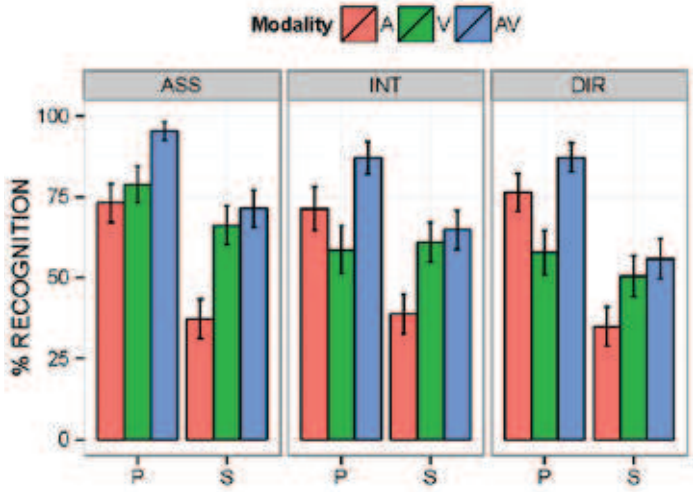

Fig. 1: mean percentage of recognition obtained by each group of attitudes (propositional, P, or social, S) for each category of illocution (assertive, ASS, interrogative (yes-no questions), INT, or directive, DIR) in their three modalities of presentation (audio, A, visual, V, audio-visual, AV).

To study the multimodal expression of such prosodic affects, Moraes and colleagues (MORAES et al., 2010) have audio-visually recorded a set of expressions acted by both a female and a male first language BP speakers. Attitudes are performed on the same set of sentences, so only attitudinal prosody may vary, while the position of accented syllables is controlled. Sentences of different lengths are used, with a systematic variation of the lexical stress position for the last word of the sentence: e.g. the oxytonic "*Roberta vai dançar*" and paroxytonic "*Roberta dançava*" sentences are used for assertive and interrogative (cf. figures 2 and 3) illocutions while the oxytonic "*Destranca o gavetão*" and paroxytonic "*Destranca a gaveta*" sentences are used for directive illocutions, for all the attitudes.

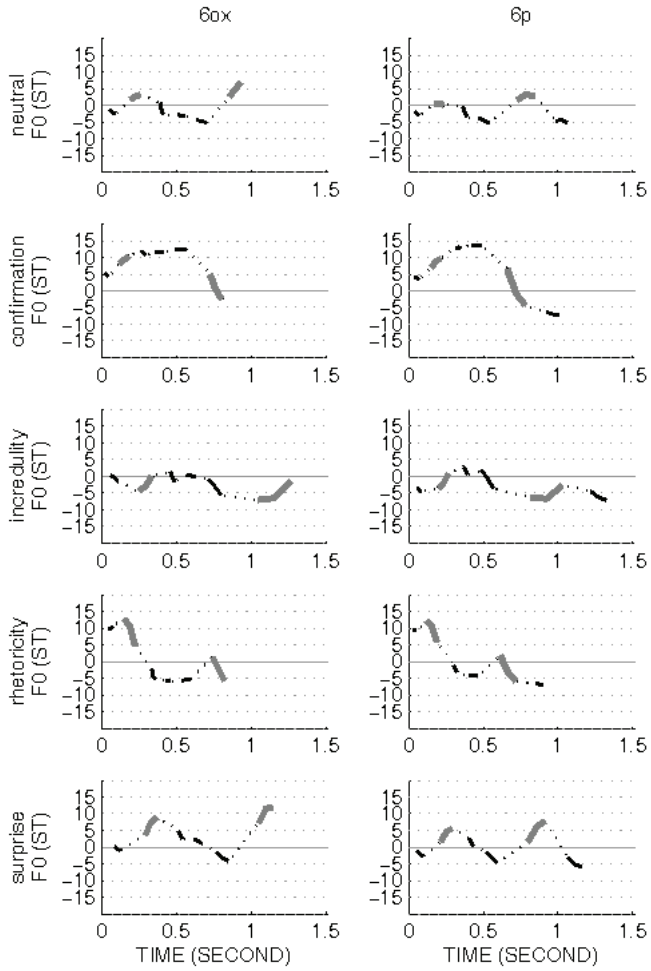

Fig.2: mean F0 contours for the three repetitions of each propositional attitude in interrogative (yes-no questions), six-syllable, oxytonic (6ox) or paroxytonic (6p) sentences as spoken by a male speaker.

Given these two sets of attitudes, one of propositional and one of social attitudes, crossed with three sentence's illocutionary category (assertive, yes-no question, directive), the performance of the two speakers was evaluated in three independent perception tests (one for each illocution category, detailed procedures and

results are presented in MORAES et al. 2010; MORAES et al. 2011; MORAES & RILLIARD, to appear). Subjects had to recognize the intended attitudes, presented either in audio only, in video only, or in a bimodal condition, using a forced-choice paradigm.

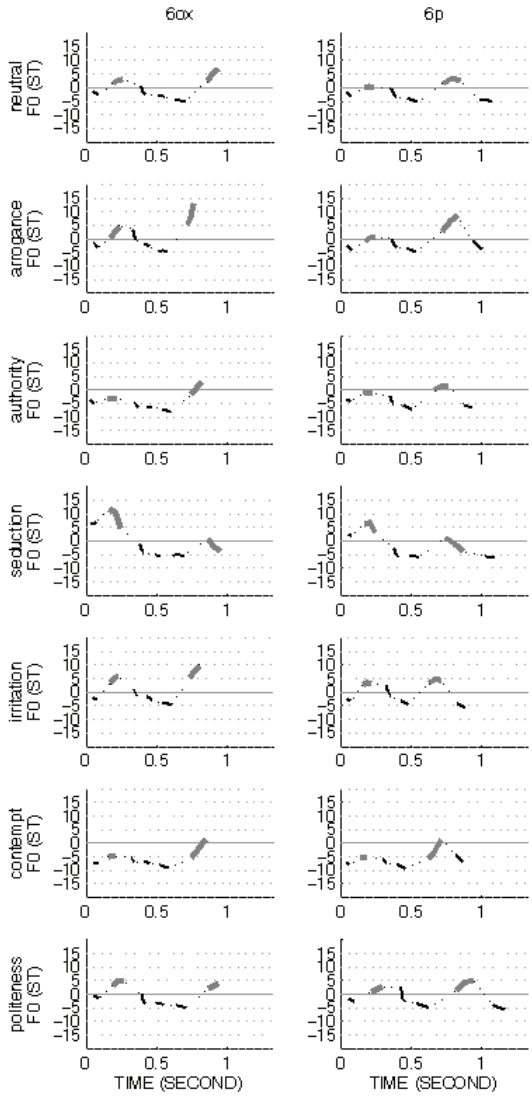

Fig. 3: mean F0 contours for the three repetitions of each social attitude in interrogative (yes-no questions), six-syllable, oxytonic (6ox) or paroxytonic (6p) sentences as spoken by a male speaker.

A main result learned from these experiments is the dominant role of the visual modality for the expression of social attitudes, while the audio modality is dominant in most cases for the propositional attitudes: in each case, the dominant modality allows better recognition scores than the non-dominant one, and/or shows the same pattern of confusions than the one observed in the bimodal condition. Percentages of recognition for each type of illocution are presented in figure 1.

This result is interpreted as an indication of the linguistic difference between both groups of attitudes. As propositional attitudes directly address the linguistic content, they are deeply embedded in the linguistic construction of the sentence, and thus more involved in the construction of its spoken form. Conversely, social attitudes help the speaker to manage the social intercourse during face-to-face communication; thus, they do not change the semantics of the sentence (note they can change the meaning of the interaction), and are less deeply involved in its linguistic construction.

What we mean by a "prosodic attitude more or less deeply involved in the linguistic construction of a sentence" is illustrated in figures 2 & 3. On these figures, the mean F0 values, as performed by the male BP speaker is represented (expressed in semitones with reference to the speaker's mean F0), for each vowel (stressed vowels are depicted in grey) of the oxytonic "*Roberta vai dançar*" (Roberta is going to dance) and paroxytonic "*Roberta dançava*" (Roberta danced) sentences. The F0 contours are plotted for the propositional attitudes in interrogative illocutions (expression of confirmation, incredulous, rhetoric, and surprised questions) in figure 2, and for the social attitudes in interrogative illocutions (expression of arrogant, authoritative, seductive, irritated, contemptuous, or polite questions) in figure 3, plus for each figure the neutral interrogative sentence

for reference. The prosody of this "neutral" sentence expresses a yes/no question addressed to the interlocutor ("*Roberta dançava?*"). This genuine, information-seeking interrogative illocution is then modified by each attitude in order to convey the specific speech act targeted by the speaker (e.g. a rhetoric question).

What is striking on these two figures is that (1) most social attitudes conserve basically the shape of the neutral sentence, with the two rising (or raised) pitch accents on the accented syllables (note the difference between the oxytonic and paroxytonic sentences), but with a register shift (typically for the expression of contempt) and/or a varied range (typically for arrogance). The expression of seduction shows the most distinct pattern among the social attitudes. Conversely, (2) the propositional attitudes depict distinct shapes that clearly depart from the neutral sentence's pattern. Confirmation shows a high plateau followed by a steep slope on the last accented syllable; incredulity is performed with a low register and a downward slope throughout the sentences, with a main rise on the first accented syllable and a delayed rise and a lengthening of the last accented syllable; rhetoric questions begins with an initial rise and show a downward slope on the last accented syllable (cf. MORAES, 2008 for details). Only surprise depicts the neutral interrogative pattern, but with a wider frequency range. Note that these different pitch contour patterns always respect the accentual structure of the sentences: oxytonic and paroxytonic sentences carry comparable overall pitch contours for the same attitudes, if one looks at their respectively accented final or penultimate syllable.

It seems clear from these experiments that propositional attitudes express a linguistic meaning that changes the reception of the sentence thanks to a modified prosodic contour, while social attitudes express

a paralinguistic meaning that changes the meaning of the interaction, and mostly modifies the range and register of the intonation contour. The modification of pitch register observed in this corpus follows the predictions of OHALA's frequency code (1984): for example, authority and contempt, two expressions where the speaker takes a dominant situation over her/his interlocutor, show a low mean pitch. Note that complex voice quality settings may also be involved in such kinds of expressions. The prosodic settings that follow a code such as the frequency code are proposed as universal – and thus should be recognized by listeners of any linguistic background. But for other prosodic changes, such as the intonation contours associated with the propositional attitudes in BP, perhaps they will not be understood by second language speakers who have had no experience with that language / culture.

2. Culturally specific attitudes and cross-cultural perception

2.1. Japanese-specific prosodic attitudes

If sound-symbolic changes in prosody, such as the one described by Ohala's frequency code, are proposed as universal, does that mean that attitudinal variations in prosody other than those related to universal codes are necessarily culture-specific? In other words, what kind of prosodic cues may be shared across cultures, or be culture-specific? Possible answers to this question will be looked for within studies based on some Japanese attitudes. Formal descriptions of Japanese politeness insist on the importance of the hierarchical relations between speakers (IDE 1982), and of the social context where a dialogue takes place (cf. the notion of discernment in HILL et al. 1986). A polite – or impolite – prosodic display can be viewed as examples of "*What A shows A thinks of B*" (HAUGH & HINZE, 2003:1584).

Work on such Japanese attitudinal prosody carried on by SHOCHI, RILLIARD et al. (2009) have described different prosodic expressions of politeness, encoded in this language and corresponding to expressive behaviors that are expected from the speaker in given social settings. Among these expressions, the most typically Japanese one is the expression of *kyoshuku*, a term that has no lexical equivalent in English. It has been described by SADANOBU (2004:34) as "*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*". This expression of politeness is dependent of the social context and of the illocutionary aim of the speaker. Conventionalized behaviors have been set up in the Japanese culture for such a communication situation. SHOCHI, ERICKSON et al. (2009) and RILLIARD et al. (2012) have shown that Japanese children acquire these codes from infancy until the age of about 10 years old. This expression of *kyoshuku* uses a harsh and tense voice quality or a breathy voice. SHOCHI, RILLIARD et al. (2009) have shown that this specific voice quality is misunderstood by U.S. English and French listeners. These listeners recognize the use of such a voice quality as typical of an irritated attitude. Meanwhile, later tests based on audio-visual cues show that a *kyoshuku* voice quality is not perceived as polite by foreign (in this case French) listeners if presented through an audio-only modality, but conversely that its audio-visual performance is perceived by the same subjects as the most polite expression amongst the presented ones (SHOCHI, KAMIYAMA et al., to appear). The suffering face, as well as the gentle bow, displayed by the speaker gives strong indices to decode the specific voice quality as clearly not irritated. This voice is thus interpreted in a different way in the monomodal and bimodal conditions. This finding enhances the results described in the

preceding section for BP: the audio-visual presentation of prosodic attitudes gives to listeners better clues to decode the prosodic expression. As understanding a prosodic stimulus outside any communication context during a perception test is a difficult task; visual cues can serve as a context to decode prosodic attitudes (cf. NADEU & PRIETO 2011).

2.2. Cross-cultural reception of prosodic attitudes

A main criticism of SHOCHI, RILLIARD et al. (2009)'s cross-cultural tests is linked with the problems raised by the translation of attitudinal labels, the description of which is generally based on concepts that may differ across languages. This problem was described by WIERZBICKA (1985, 1992) for studies on emotion, and is certainly as acute for attitudes. To bypass the limitations raised by the necessary translation of attitudinal labels, another paradigm was used, based on ROMNEY et al. (1997, 2000). In a pair comparison paradigm, subjects from four different linguistic backgrounds (Japanese [JP], Brazilian Portuguese [BP], U.S. English [US], French [FR] – non-Japanese subjects don't know the Japanese language) had to judge the distance between pairs of prosodic attitudes, presented either in audio only, in visual only or in audio-visual – but without any labels, just their performance. In order to evaluate if the subjects, when giving their pair rating distance judgments, were effectively judging the targeted meaning of the prosodic attitude, not only the form of the performance, a semantic condition was added. To avoid the translation problem, the solution proposed in WIERZBICKA (2005) to use Natural Semantic Metalanguage scripts (NSM, GODDARD, 2002) to describe the meaning of a concept was tested.

An NSM script tackles the problem of definition and of translation by using *semantic primitives* universally used across languages (cf. WIERZBICKA 1996, 2005 for definition and example of use). Pairs of such NSM scripts were thus presented to subjects (after the audio-visual stimuli), who had to judge the difference between the two concepts expressed by the scripts. The scripts were translated in the first languages of subjects by using the translations of the universal semantic primitives (GODDARD, 2002) found in the NSM literature. Five attitudes, presenting different expressions of politeness or impoliteness in Japanese were used: three expressions of politeness (courtesy politeness [PO], sincerity politeness [SIN], *kyoshuku* [KYO]), a politely neutral declarative sentence [DC], and an impolite expression of arrogance [AR]. The details of the procedure and of the results analysis are presented in RILLIARD et al. (2014).

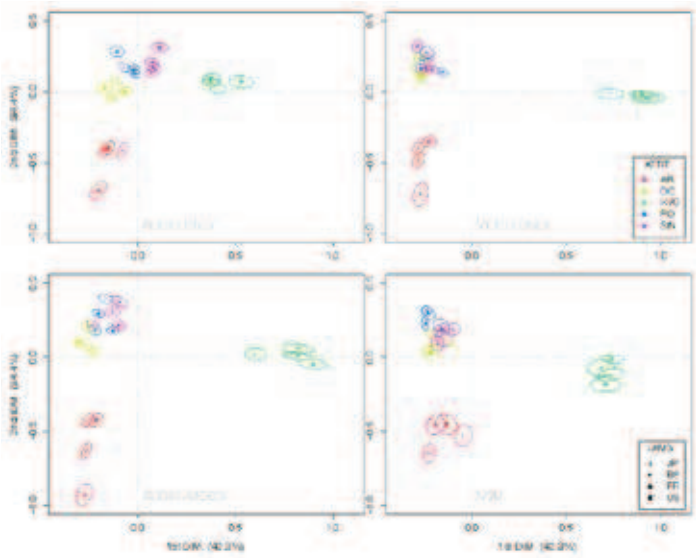

Fig. 4: Separate plots of the positions of attitudes on the first two dimensions of the CA, for each modality of presentation, grouped by languages (see text). The ellipses around each attitude represent the 97.5% confidence regions.

After compiling the distances obtained between each pair of attitudes by listeners of each linguistic group and in each of the four conditions of presentation, a correspondence analysis (CA) allows to plot the distribution of these attitudes in what ROMNEY et al (1997) coined their "semantic structure" (cf. figure 4). The further two points are on the plot, the more different the expressions are for the subjects. The analysis of these results shows that about 60% of the observed distribution is common to all linguistic groups and all modalities of presentation. This result stresses the overall comparability of information carried by the audio, visual and NSM modalities – and the even greater similarity between groups of different linguistic origins. Only 1% of the variance can be assigned to the language origin of subjects, while 8% is linked with the modality of presentation.

By comparing the shapes of the groups of attitudes obtained in each modality, we can have an idea of the cognitive differences that subjects perceived between stimuli. From this comparison, it is clear that audio information shows the answers grouped on the smallest area (thus the perceptually closest ones) – but with 5 distinct clusters (one distinct set of points for each attitude, grouping answer by all language groups): whatever the language of subjects, they separate clearly each attitude, but put them not that far apart, compared to the visual and audio-visual modalities. Visual modality makes mostly three main clusters, but placing each at one extremity of the plot. Finally, the NSM presentation shows more important differences perceived by listeners between the proposed concepts – but also stresses the variation in the interpretations made by each language group. NSM is the modality where variation linked to linguistic origin is the more important. Thus, the presentation of concepts of attitudes gives rise to more

variability across cultures than the presentation of the (audio or visual) expression of these concepts. In other words, the "*how*" of the proposed speech acts is decoded in a more similar fashion across cultures, while the "*what*" shows more variability. It is somehow as if prosody expresses only part of the complete speech act – a part that is (in the cases of these five attitudes) perceived similarly by first language and as well as by foreign listeners; however, the complete speech act is interpreted from contextual cues that are linked to more culture-specific concepts. This conceptual part of the speech act tends to show more cross-cultural variability.

Thus, another question may be raised: to what extent does the performance (their *how*) of speakers from varying linguistic backgrounds differ in expressing the same prosodic attitude (the *what*)? This question is difficult to answer on data such as those presented so far, as the task assigned to speakers was dependent on the concepts underlying each attitude – concepts with a varying interpretation across cultures and languages.

3. Speech acts in prototypical situations: cross-cultural strategies

3.1. Collection of comparable attitudes across cultures

In the preceding section, we saw that the interpretation of a given attitudinal performance may vary with the cultural origin of the listener (even if this effect is small for the proposed stimuli). A prosodic performance is also linked to its context of occurrence – the interpretation of which is constrained by culture. To compare prosodic performances across languages, one would need to neutralize the effect of the context. But how can one obtain prosodic performances from speakers of varying cultural origins in comparable communication contexts?

To that aim, a controlled recording paradigm has been set up – that is fully detailed in RILLIARD et al. (2013). To avoid the use of attitudinal labels – the conceptual descriptions of such communication contexts that have been shown to be language-dependent, communication situations have been defined, that explicitly state the communicative aim of the speaker, as well as the social relation between the speaker and the listener. These social relations typically focus on the hierarchical difference that may exist between the speaker and the listener. Then, two target sentences ("*a banana*" and "*Mary was dancing*") were selected, and translated in the three languages recorded up to now (U.S. English, Japanese, French). Small dialogues have been written, staging the two speakers (target speaker is A, her/his interlocutor is B) in a small interaction that ends with speaker A saying one of the target sentence, with a communication goal that fits a given attitudinal expression. Sixteen situations, corresponding to sixteen attitudes have thus been selected, for which a prototypical situation has been defined, in order to explain it to the speakers. Here are the descriptions of some of these communication situations between A and B:

- *Irritation* (IRRI): A & B are almost the same age and know each other. A is sitting next to B. Suddenly, B starts to smoke, and A is very angry; he wants him/her to stop, expressing his irritation toward speaker B. The scene is a public place.
- *Obviousness* (OBVI): A & B are colleagues, same age; everyone knows B doesn't like French movies, but A asks B if he likes French movies or not; the scene is at a coffee shop.
- *Seduction* (SEDU): A likes B a lot and they have an intimate relationship. A gives a compliment to B in a sexually provocative way. The scene is

at a club house (note: "seduction" might also be called "flirtatious").

- *Sincerity* (SINC): B is chief of the section which A belongs to; B is older than A. The chief (B) wants A to take on a big project; A is pleased to be asked to do this, and expresses his enthusiasm, honesty, and sincerity for this task. The scene is at B's office.
- "*Walking on eggs*" (WOEG, note: sometimes referred to as "Walking on eggshells"): B is chief of the section to which A belongs; B is older than A. The chief (B) wants A to do a task which is a lot of work, and it seems to A impossible to do this, so A tries to reject this request by trying to make sure her/his boss (B) doesn't get angry for refusing. The scene is at B's office.

Sixteen situations have been inspired from the attitudes already recorded in different languages, but are not necessarily conventionalized expressions of each language. Hence, *seduction* is not a conventionalized expression in Japanese, and maybe not in French either under such a definition, which could be coined a "Hollywood seduction". Conversely, the *walking on eggs* situation has been defined to somewhat cover the situation of the typically Japanese *kyoshuku*, that is not conventionalized in U.S. English nor in French.

Speakers from these three languages have been recruited, mostly among university students. Each speaker has been selected so to have a first language (L1), and a second language (L2) with a good proficiency level (a criterion which has been defined as having had a one-year stay studying in the country speaking that language). Performances for each speaker have been recorded in their L2 and L1 languages (in that order, so the first language may have less influence on the second

language). At least 3 female and 3 male speakers have been recruited so far for each language pair.

3.2. Performance comparison for L1 and L2 speakers of US English

The recorded audio-visual performances of first language speakers of U.S. English and of L2 speakers of English, whose first language is Japanese, have been evaluated by a set of first language U.S. English listeners. Listeners rate each audio-visual performance on a 1 to 9 scale, knowing the communication goal of the speaker. The resulting analyses (detailed in RILLIARD et al. 2013 and SHOCHI, ERICKSON et al. 2013) show significant effects of the speaker and of the targeted attitudinal expression on the evaluated performance. Speakers perform differently in a given situation, certainly according to their own individuality and personal history. Some attitudes are more easy or difficult to perform, certainly because of varying factors. Surprise received the highest score for both speakers group, but factors such as the unnatural communication settings and sentences may be a main explanation for the low performances observed for irony, or expressions linked to taboo may partly explain the inter speaker variation observed for seduction. Another factor had a significant effect on the results: the cultural origin of the speaker. Higher mean performance levels are observed for L1 speakers over L2 speakers. This result was expected; meanwhile, the detailed analysis (figure 5) shows that this difference between L1 and L2 is significant in only five attitudes out of the 16 – and moreover, for two attitudes, L2 speakers perform better than L1 speakers.

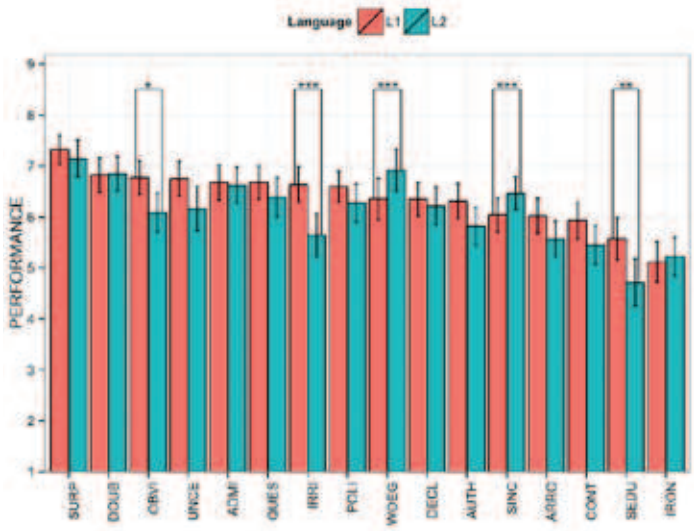

Fig. 5: mean rated performances for each attitude, as performed by L1 and L2 speakers (asterisks indicate significant differences for an attitude performed by the two groups, error bars indicate the confidence intervals at 99%).

Note that these scores reflect the average performance of all speakers of the L1 or L2 groups – not the individual performances. Of course, some speakers do produce more efficient expressions than others – and this varies among the expressions. We do not enter the individual performances here – which are reflected in the error bars of figure 5: in most cases, the intra-cultural-group variance is more important than the inter-cultural-group variance. An analysis of the subject-specific variance requires an in-depth analysis of their performance and personality. Here we focus on the comparison of the main tendencies observed between these two groups. Figure 6 presents the distribution of performance measures for all speakers in both groups (L1 and L2), for the five attitudes who show a difference, plus the expression of surprise, that received the best scores for both groups. The distribution observed for surprise shows

very similar tendencies for both groups, with about one third of the answers at the top of the performance scale.

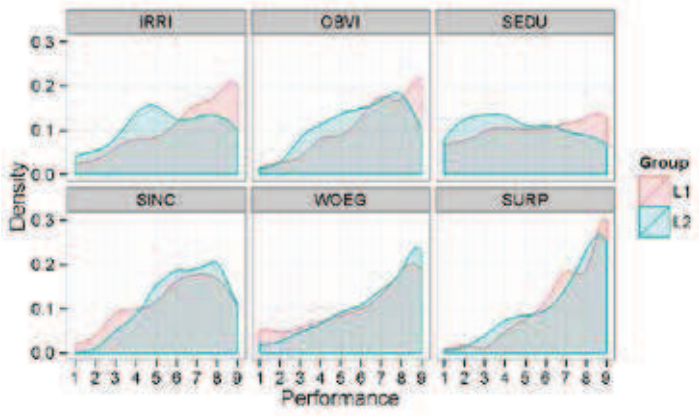

Fig. 6: density of performance answers for the 5 attitudes that show a difference between the L1 and L2 groups of speakers, plus surprise.

For irritation, obviousness, and seduction, L1 speakers outperform L2 ones. For irritation and seduction, the distributions show that few L2 speakers have high performance ratings, most ratings being regrouped on the bottom of the rating scale - contrary to L1 ratings; for obviousness, the shape of L2's distribution resembles the L1's distribution, but with a peak rating 1 point below L1's peak and fewer "9" judgments. For sincerity and WOEG, L2 receives higher mean performances than L1. Both groups show more similar distribution shapes for these two attitudes than for the preceding three. The difference lies mainly in a higher level of high performance ratings (between 6 to 8 for sincerity, 8 or 9 for WOEG) for L2 subjects.

These observations may be interpreted as if Japanese speakers of English show difficulties in eliciting seduction and irritation in a manner convincing to US English listeners. For seduction, this may be linked with the unconventionalized status of this expression in the Japanese

culture; for irritation, it may be linked to a varying strategy to express it in both the U.S. and Japanese cultures.

The higher performances obtained by Japanese speakers over U.S. English speakers for sincerity and WOEG, on the other hand, can be linked with a lack of conventionalization of these two expressions in U.S. culture, while it is an integral part of Japanese culture, observed over centuries (SADANOBU, 2004; SHOCHI, RILLIARD et al., 2009).

Conclusions

The first part of the paper shows how prosody is used by speakers to change the form of speech to tune the meaning induced by the lexicon. If important changes are induced on the prosodic shapes, these changes always respect linguistic constraints, so the message continues to be receivable by the listener. Thus, attitudinal prosody respects the place of accents (the stressed syllables), but not necessarily their intonational shape. Stressed syllables are the preferred place for intonational variability, and typically for the expression of propositional attitudes. Propositional attitudes generally show wide prosodic changes on these syllables. Attitudes are also expressed through the multimodality of speech, and audio and visual modalities are used respectively to address preferably the linguistic message or the social relation between interlocutors.

These expressive changes follow codes. The codes may be universals of animal communication, like Ohala's *frequency code* or other sound symbolism codes, or may depend on language-specific conventionalizations built up in a given culture. The second part of the paper has shown that a set of attitudes related to politeness expressions, including culturally specific expressions such as Japanese *kyoshuku*, is perceptually organized by

listeners from various linguistic and cultural origins with very few differences that can be linked to their cultural origin. Meanwhile, the perception of *kyoshuku*, in its audio modality alone, can be misinterpreted - and such misinterpretations are likely to originate in a lack of contextual cues to decode the vocal form of *kyoshuku*. Let non Japanese-speaking subjects see the face of the speaker, and they will no longer misinterpret *kyoshuku's* peculiar voice quality. The speaker's face disambiguates prosodic variations unclear to foreign listeners: the facial information functions as a context to decode prosodic information (cf. NADEU & PRIETO, 2011). The suffering face has the same iconic meaning as the rough voice of the speaker in conveying embarrassment, not irritation. Japanese speakers do not need as many contextual cues, as they already know the prosodic code.

Such culture-specific codes are acquired by first language subjects throughout infancy. This long training allows them to master the specificity of the targeted expression. This may be the cue to interpret the results of the third section. Cross-cultural recordings show that in some situations of communication, speakers of different cultural origins show varying performance levels. For most prosodic attitudes, L1 and L2 speakers have comparable performances; with a small facilitation effect observed for first language speakers. But when it comes to culturally specific attitudes, two shifts of behavior are observed. On the one hand, L2 speakers have greater difficulties in achieving well-performed expressions not conventionalized in their culture, compared to those attitudes that are not culturally specifics. On the other hand, L2 speakers show higher performance levels for expressing certain attitudes, which are encoded in their L2 culture, than do L1speakers for whom these attitudes are not well-encoded in the L1 culture. Most first-language speakers do not show a high performance level for such

"foreign" expressions, while first language listeners welcome the strategy developed by the L2 speakers.

But this observation raises questions: in the case of WOEG, do Japanese speakers use the same prosodic strategy in U.S. English as in Japanese - where WOEG corresponds to the *kyoshuku* expression? In other words, have we observed a prosodic transfer from the Japanese *kyoshuku* into this WOEG situation for a dialogue carried out in U.S. English? And if a prosodic transfer is plausible in the case of WOEG, is this a general situation? Is it possible to transfer any culturally specific code, built-up in a given language, into another language? We may have another example of this by observing how the expression of seduction is performed in the Japanese language by first language Japanese speakers and by L2 U.S. English speakers. Finally, the expression of irritation that exists in both U.S. English and Japanese, gives us a conflicting case. Why do Japanese speakers receive such low scores for this attitude? Do they try to use their first language attitudinal strategy? And thus, this foreign code could not have been accepted as valid by first language listeners, compared to their own code. These questions about prosodic encoding of attitudinal meaning in languages and cultures have still to be investigated.

Acknowledgements

This work was partly supported by the ANR PADE project, and by the Japanese Grant-in Aid (A) 23242023. We are deeply grateful to the speakers we recorded for this work, as well as to the listeners who participated in the experiments. Our special thanks to Mariko Kondo, Sylvain Detey, and Kelly Kirk for helping us in this research. We are deeply indebted towards the two anonymous reviewers whose comments helped us improve the final version of this paper.

References

GODDARD, C. The search for the shared semantic core of all languages. In Goddard, C.; Wierzbicka, A. (Ed.). *Meaning and Universal Grammar - Theory and Empirical Findings*. Volume I. Amsterdam: John Benjamins, 2002, p. 5-40.

GU, W.; ZHANG, T.; FUJISAKI, H. Prosodic Analysis and Perception of Mandarin Utterances Conveying Attitudes. *Proceedings of Interspeech*, Firenze, 2011, p. 1069-1072.

GUSSENHOVEN, C. *The phonology of tone and intonation*. Cambridge: Cambridge University Press, 2004.

HAUGH, M.; HINZE, C.A Metalinguistic approach to deconstructing the concepts of 'face' and 'politeness' in Chinese, English and Japanese terms. *Journal of Pragmatics*, v. 35, p. 1581–1611, 2003.

HILL, B.; IDE, S.; IKUTA, S.;KAWASAKI, A.; OGINO, T. Universals of linguistic politeness - Quantitative Evidence from Japanese and American English. *Journal of Pragmatics*, v. 10, p. 347-371, 1986.

IDE, S. Japanese sociolinguistics politeness and women's language. *Lingua*, v. 57, p. 357-385, 1982.

LÉON, P. *Précis de phonostylistique* – parole et expressivité. Paris: Nathan, 1993.

MARTINS-BALTAR M. *De l'énoncé à l'énonciation: une approche des fonctions intonatives*. Paris: Didier, 1977.

MORAES, J. A. The pitch accents in Brazilian Portuguese: Analysis by synthesis. *Proceedings of Speech Prosody 2008*, Campinas, 2008, p. 389–397.

MORAES, J. A.; RILLIARD, A.; MOTA, B.; SHOCHI, T. Multimodal perception and production of attitudinal meaning in Brazilian Portuguese. *Proceedings of Speech Prosody 2010*, Chicago, 2010, paper 340.

MORAES, J. A.; RILLIARD, A.; ERICKSON, D.; SHOCHI, T. Perception of attitudinal meaning in interrogative sentences of Brazilian Portuguese. *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)*, Hong Kong, China, 2011, p. 1430-1433.

MORAES, J. A.; RILLIARD, A. Illocution, Attitudes and Prosody. In Raso T. et al., *Spoken Corpora and Linguistic Studies*, Amsterdam: John Benjamins, to appear.

NADEU, M.; PRIETO, P. Pitch range, gestural information, and perceived politeness in Catalan. *Journal of Pragmatics*, v. 43, n. 3, p. 841-854, 2011.

OHALA, J.J. An ethological perspective on common cross-language utilization of F0 of voice. *Phonetica*, v. 41, p. 1-16, 1984.

RILLIARD, A.; SHOCHI, T.; ERICKSON, D.; MORAES, J.A. Developmental perception of polite & impolite non-verbal behaviours in Japanese. In Mello, H.; Pettorino, M.; Raso, T. (Ed.). *Proceedings of GSCP International Conference : Speech and Corpora*. Firenze: Firenze University Press, 2012, p. 167-171.

RILLIARD, A.; ERICKSON, D.; SHOCHI, T.; MORAES, J. A. Social face to face communication – American English attitudinal prosody. *Proceedings of Interspeech 2013*. Lyon, 2013.

RILLIARD, A.; ERICKSON, D.; MORAES, J. A.; SHOCHI, T. Cross-Cultural Perception of some Japanese Expressions of Politeness and Impoliteness. In Baider, F.; Cislariu, G. (Ed.) *Linguistic approaches to emotions in context*. Amsterdam: John Benjamins, 2014, p. 251-276.

ROMNEY, A.K.; MOORE, C.C.; RUSCH, C.D. Cultural universals: Measuring the semantic structure of emotion terms in English and Japanese. *Proceedings of the National Academy of Sciences*, v. 94, p. 5489-5494, 1997.

ROMNEY, A.K.; MOORE, C.C.; BATCHELDER, W.H.; HSIA, T.-L. Statistical methods for characterizing similarities and differences between semantic structures. Proceedings of the National Academy of Sciences, v. 97, n. 1, p. 518-523, 2000.

SADANOBU, T. A natural history of Japanese pressed voice. *Journal of the Phonetic Society of Japan*, v. 8, n. 1, p. 29-44, 2004.

SHOCHI, T.; RILLIARD, A.; AUBERGÉ, V.; ERICKSON, D. Intercultural perception of English, French and Japanese social affective prosody. In Hancil, S. (Ed.). *The role of prosody in affective speech*. Bern: Peter Lang, 2009, p. 31-59.

SHOCHI, T.; ERICKSON, D.; SEKIYAMA, K.; RILLIARD, A.; AUBERGÉ, V. Japanese children's acquisition of prosodic politeness expressions. *Proceedings of Interspeech 2009*, Brighton, UK, 2009, p. 1743-1746.

SHOCHI, T.; KAMIYAMA, T.; RILLIARD, A.; AUBERGÉ, V. Effet d'apprentissage des expressions prosodiques et gestuelles de politesse en japonais chez des apprenants français. *Japon Pluriels*, v.9, to appear.

SHOCHI, T.; ERICKSON, D.; RILLIARD, A.; MORAES, J. A. Prosodic differences between L1 and L2 performance in social face to face American English context. *Proceedings of Workshop on Affective Social Speech Signals*, Grenoble, 2013.

WICHMANN, A. The attitudinal effects of prosody, and how they relate to emotion. *Proceedings of ISCA Workshop on Speech and Emotion*, Newcastle, UK, 2000, p. 143–148.

WIERZBICKA, A. A semantic metalanguage for a cross-cultural comparison of speech acts and speech genres. *Language in Society*, v. 14, n. 4, p. 491-513, 1985.

WIERZBICKA, A. Defining Emotion Concepts. *Cognitive Science*, v. 16, p. 539-581,1992.

WIERZBICKA, A. Japanese Cultural Scripts: Cultural Psychology and "Cultural Grammar". *Ethos*, v.24, n. 3, p. 527-555,1996.

WIERZBICKA, A. Empirical Universals of Language as a Basis for the Study of Other Human Universals and as a Tool for Exploring Cross-Cultural Differences. *Ethos*, v. 33, n. 2, p. 256–291, 2005.

[Recebido em 30 de junho de 2013
e aceito para publicação em 22 de novembro de 2013].