JoSS Journal of Speech Sciences

DOI: 10.20396/joss.v14i00.20379

CROSS-CULTURAL DIMENSIONS ORGANIZING PROSODIC ATTITUDES RECEPTION: A META-ANALYSIS OF FREE LABELING STUDIES

RILLIARD, Albert[1*]
DE MORAES, João Antônio[2]
ERICKSON, Donna[3]
GUERRY, Marine[4]
HÖNEMANN, Angelika[5]
LEE, Tan[6]
MA, Matthew King-Hang[7]
MIXDORFF, Hansjörg[8]
RAO, Preeti[9]
SHOCHI, Takaaki[10]

[1] Université Paris Saclay, CNRS, LISN, Orsay, France – ORCID: https://orcid.org/0000-0001-6490-2386
[2] Universidade Federal do Rio de Janeiro & CNPq, Brazil – ORCID: https://orcid.org/0000-0002-8894-166X
[3] Haskins Labs, New Haven, CT, USA – ORCID: https://orcid.org/0000-0003-2283-9432
[4] Bordeaux 3 University, France – ORCID: -
[5] Berliner Hochschule für Technik, Berlin, Germany - ORCID: -
[6] The Chinese University of Hong Kong, Hong Kong, China - ORCID: https://orcid.org/0000-0002-7089-3436
[7] The Hong Kong Polytechnic University, Hong Kong, China - ORCID: https://orcid.org/0000-0001-9194-2498
[8]Berliner Hochschule für Technik, Berlin, Germany - ORCID: https://orcid.org/0009-0002-6505-6788
[9] IIT Bombay, India - ORCID: https://orcid.org/0000-0002-2842-8450
[10] Shimonoseki City University, Japan & CLLE, France & LaBRI, France - ORCID: https://orcid.org/0000-0002-1529-385X

[*] Corresponding author: albert.rilliard@limsi.fr

Abstract: *We present a meta-analysis of results from experimental studies on attitude reception in seven languages (Brazilian Portuguese, Japanese, French, German, Cantonese, American English, Hindi). The studies involved free-labeling of perceived attitudes in audio-visual stimuli. The productions of 88 speakers from the seven languages were obtained using the same elicitation methodology, allowing to record sixteen audiovisual attitudes. These performances, rated in preceding works using a free-labeling paradigm, were grouped and analyzed to compare how the attitudinal performances spread along the main dimensions of the shared cognitive representation. A hierarchical clustering then regrouped attitudinal expressions as a function of their cognitive proximity. A large cluster solution showed the main dimension that organizes these expressions may be interpreted as the "Unpredictability" dimension proposed by Fontaine et al. (2007) for emotions, followed by the Evaluation-Pleasantness one; Activation-Arousal and Potency-Control arrived later but played a determinant role in the organization of attitudes. A fine-grained 13-cluster solution showed most attitudes were singled out by the listeners despite the variations in speakers and elicitation contexts. The analysis of each of these clusters brings insight into the cultural similarities and differences in the reception of these different attitudinal expressions. A notable result is the variation in valence attributed to the expression of Irony that underlines the potential communication problems that may be linked to interaction routines. On the other hand, Surprise was clearly identified. The existence and importance of the Unpredictability dimension and its relation to the illocutionary opposition between assertive and interrogative acts underlines the pertinence of Mello & Raso's (2011) analysis.*
Keywords: Prosodic attitudes; cross-cultural; dimensions of meaning; multidimensional analysis

1 Introduction

During verbal interactions, we use a series of social affects to provide a smooth, efficient way to exchange information. Specifically, non-lexical prosodic information is used to convey a variety of meanings that enhance the content of each speech turn. In an early tentative to study such "intonational" meaning (here opposed to the role of lexical tones), Chang (1958) classified sentences produced in an eight-hour corpus according to the sentence type (statement or question), the presence of emphasis, and a set of seven categories of attitudes or emotions that add "shades of meanings" to the literal interpretation of the sentence through their intonational variations. Uldall (1960) studied similar prosodic attitudes within a perceptual framework building on Osgood's semantic differential method (Osgood et al. 1957). Using a set of ten scales, she estimated the perceptual differences across sixteen different prosodic contours (systematic intonation variations applied to four sentences carrying different illocutionary functions). The contours were not specifically motivated and had varied interpretations depending on the sentences. A main result of this early study is the confirmation, for prosodic meanings, of the three main dimensions of meaning postulated based on lexical units by Osgood et al. (1957): evaluation, activity, and potency. These dimensions are interpretations of the first three axes of factor analyses applied to evaluation judgments from various linguistic materials. Focusing more on interactive speech productions, Fónagy & Bérard explored the prosodic change induced in one sentence when inserting it into various interaction contexts and how listeners could interpret such prosodic information (Fónagy and Bérard 1972). This leads to the "*cliché mélodique*" notion that denotes the relation between specific intonational shapes and one or a few specific interpretations by listeners and their use in given interaction contexts (Fónagy et al. 1983).

These early works on prosodic attitudes were followed by a series of works in different languages (e.g., Martins-Baltar 1977; Pierrehumbert and Hirschberg 1990; Fujisaki and Hirose 1993; de Moraes 2008; Shochi et al. 2009), mostly targeting second-language learning, describing sets of prototypical attitudes conveying predefined prosodic interactional meaning, such as producing a sentence with authority, uncertainty, politeness, etc. Scholars also classified such affective expressions into two categories: (i) expressions that affect the propositional meaning of the sentence (e.g., irony, doubt, obviousness), called propositional attitudes, and (ii) expressions directed to the addressee (like friendly, rude, seduction), called social or behavioral attitudes (Wichmann 2000; de Moraes 2011). This distinction was derived from the fact that such socio-affective prosodic meaning may be viewed as derived from emotional expressions, gradually encoded within a cultural framework in the interaction practices and within the languages, and thus recycled to denote specific meanings (Fónagy 1987). The different levels of encoding of this information have been referred to with various terms in the literature, depending notably on the authors' background. A proposal to clarify the use of these terms to address specific aspects of language has been made by Mello & Raso (2011), who attribute specific meanings to the terms modality, illocution, and attitude: "[…] we suggest a rationale that allocates modality to a semantic level in which the speaker's stance towards her locutory expression is manifested; similarly illocution belongs to a pragmatic level in which the speaker's stance towards her interlocutor is manifested, and finally attitude will be allocated to a socio-interactional conventionalized level." (Mello and Raso 2011, p. 5). We will follow this recommendation in this paper. This gradual encoding of vocal information within the language was exemplified by comparing the intonational contours of sentences produced with three illocutions (assertion, question, order) with several propositional or social attitudes, or emotions (de Moraes and Rilliard 2014, 2016). While the emotional and social expressions mostly influenced the dynamic of

prototypical contours of each illocution, propositional attitudes modified their phonological implementation (see also de Moraes 2008).

Such attitudinal meanings are thus difficult to describe in extension, and the number of distinct prosodic contours remains an open question, with possible variants or shades of meanings (Uldall 1960; Hirst 2005; Post et al. 2007), that notably depend on the language described and the descriptor's point of view (e.g., Pierrehumbert and Hirschberg 1990; de Moraes 2008; Mac et al. 2009; Lu et al. 2012). One challenge in relation to studying such types of interpersonal stances is their intercultural comparison. This is notably due to the non-equivalence of translated labels used to characterize each expression: the term "politeness" in English and its translations in other languages, such as Portuguese ("*polidez*"), French ("*politesse*") or Japanese ("丁寧", *teinei*) refers to a general 'courteous politeness' that is linked to specific behaviors (tone of voice, body position, etc.) and interaction contexts (hierarchical relationship with or distance to the addressee, for example; Spencer-Oatey 1996) that are not necessarily equivalent in different cultures. See, for example, the proposal by Wierzbicka (1996) to work on cultural scripts to avoid such problems, the discussion on cross-cultural differences for (im)politeness conceptualizations in the introduction of Shochi et al. (2023), or the prototypical polite situation for French behavior proposed by Kerbrat-Orecchioni (2005). Cross-cultural differences in terms of prosodic performances and their perceptual and cognitive organization are thus hard to study because of these many variations. To bypass these limitations, a series of recordings have been made in different languages that collect the behavior of various speakers in a given situation with a specific communication goal. This builds on Fónagy et al.'s (1972) use of one sentence in various contexts, translating these contexts into various languages, keeping the same communication goals, and distance and power relationships between the interlocutors. These datasets have been described in various works (see details later), and all have been evaluated using a free-labeling paradigm: listeners were asked to describe the communicative goal of the speaker for each presented performance (without communication context). The results of these tests are used here to conduct a meta-analysis study aiming at describing the main dimensions organizing these attitudes, which 'shades of meaning' are closer or more distant from others, and what in this organization can be found cross-culturally, as well as describing some culturally specific settings.

As stated above, the perceptual similarities of such attitudes can be spread across abstract dimensions. The most common interpretation of these dimensions of meaning goes along the three dimensions of Osgood's work on lexicon (Osgood et al. 1957, 1975) which also have a tradition in emotion psychology (e.g., Russell 1991). A notable contribution to the dimensional description of emotions is found in Fontaine et al. (2007), who showed that four dimensions are required to account for the variability of emotions. They coined these dimensions: (i) Evaluation-Pleasantness, (ii) Potency-Control, (iii) Activation-Arousal, and (iv) Unpredictability, and they are ranked here in decreasing order of importance for factors of a multidimensional analysis performed on perceptual data of emotions (Fontaine et al. 2007). It should be noted that the order of importance of such dimensions depends upon the type of material used to obtain them (e.g., in terms of acoustics, the Activation-Arousal dimension comes first; Goudbeek and Scherer 2010), although the first three are robustly found in multidimensional studies with different experimental approaches (Romney et al. 1997; Moore et al. 1999) whereas the fourth, "Unpredictability," is more elusive. It is described as separating reactions to events that are predictable, expected, or familiar from those that are unpredictable, abrupt, or novel (Fontaine et al. 2007; Scherer 2009b). A prototypical emotional reaction along this scale is Surprise, a jaw-dropping reaction to something not foreseen. An interesting point for attitudes and their use during spoken interactions is that the "Unpredictability" dimension appears to have a major role in such material, where the opposition between declarative and interrogative acts explains most of the variance (Guerry et al.

2016a; Rilliard and de Moraes 2017). A major function of prosody, in many languages, is to robustly convey the illocutionary opposition between declarative and interrogative acts (Ohala 1994; Hirst and Di Cristo 1998; Miranda et al. 2021). This relation between (a) two illocutions (declaration and interrogation), (b) the unpredictability dimension, and (c) prosody, which together function to organize how listeners conceptualize attitudes, is an important focus point here.

This paper aims to compare how sixteen illocutions or attitudes (we'll use "attitudes" only for short here) are perceived and organized in different languages: what are the main dimensions that organize them, and do they differ across languages and cultures? The paper is based on perceptual evaluations of how the communicative aims of the prosodic performances are described using vernacular languages. Participants were free to use any word; thus, for a given language, an expression is described as the set of words that all listeners attribute to it. This variation renders descriptions obviously fuzzier than the single categories used in many forced-choice experimental paradigms, but they have at least three advantages: (i) the listeners are not biased into using a given predefined list of labels, (ii) there are no translation difficulties for such labels, as each language is processed directly, and (iii) the variety of labels (and their relative frequency) used to define an expression give a nuanced view of the various facets that compose a given expression. Meanwhile, this article will not address, for obvious space restrictions, the description of the performances themselves: there is no acoustic analysis or comparison of the performances across affects and languages. Some references to existing analyses will be given when they exist, but most of this work is still to be done.

2 Methods

2.1 Original studies and corpora, perceptual datasets

This paper is based on the experimental results of several studies that all follow the same methodology developed during a collaborative project targeting a better description of prosodic attitudes across cultures. This methodology is described in Rilliard et al. (2013) for its application to English, and a summary is given hereafter to allow a full understanding of the material on which these experiments are based.

Each corpus encompasses sixteen types of attitudes (cf. Table 1) that build on preceding works; attitudes were selected to represent a set of situations where the speaker has specific relations with the interlocutor in terms of distance and hierarchy (Spencer-Oatey 1996), attitudes valence (positive, neutral or negative) and dominance (if the attitude expresses an imposition or not). The speakers perform each target attitude during an interaction with the experimenter in small, scripted dialogues that end with two target utterances (in English: "A banana" and "Mary was dancing"); recordings were made in sound-processed rooms, with the speaker audio-visually recorded. First, a prototypical situation was introduced for the speaker to train their behavior in such a case; then, the speaker and the experimenter acted out the two scripts. This leads to 32 stimuli (16 attitudes on two utterances) per speaker. A variable number of speakers was gathered in each language, with a minimum of eight different first-language speakers, spread between genders. Having several speakers was important in dealing with large individual variations in attitudinal expressivity. These audio-visual productions were later segmented to keep only the 32 target stimuli of each speaker.

For each language, the complete set of productions (the 32 stimuli by speaker, for all speakers) was then evaluated for their quality with respect to the targeted communication aim, asking listeners how well a given behavior was coherent with a given context and communication aim (for details, see Rilliard et al. 2013). This step allowed us to keep the best productions – those

that received the highest scores in terms of their adequateness to express the given prosodic attitude: the performances of two speakers for each gender (thus, four performances in total) were kept to represent each of the 16 attitudes, for the two utterances (128 stimuli: 16 attitudes * 2 utterances * 4 speakers). These stimuli were then used for further perceptual evaluations.

Table 1: The sixteen targeted attitudes, with their English labels, their abbreviation, and the prototypical context used for American English, reproduced from Rilliard et al. (2017).

English label (abbreviation): Prototypical context for American English
Admiration (ADMI): A & B are almost the same age and know each other well. Both love French cuisine, and talk about the very delicious food they ate yesterday at a famous French restaurant. The scene is at a coffee shop.
Arrogance (ARRO): both A & B are from the same university, but A is older and A's father is head of the university and A is a bit of a snob. Both know each other, but are not friends. A organized a social party, and B was not invited to the party, but A is aware of his/her presence during the party. The scene is a party room, and A says to B that only his friends are invited.
Authority (AUTH): Speaker A is a custom agent; speaker B is a traveller. B is in front of A, requesting permission to enter the country; A needs to impose his authority; the scene is at a custom counter at the airport.
Contempt (CONT): both A & B are from the same university, but A is older; both know each other, but are not friends. In fact, A really hates B. A organized a social party, and speaker B was not invited, but A is aware of his/her presence. The scene is at a party room
Doubt (DOUB): A & B are colleagues, same age. A knows that his colleague B didn't go to the baseball game yesterday, but B pretends he went to the game, and A doesn't believe it. The scene is at a coffee shop.
Irony (IRON): A & B are friends, same age; A is going to Boston to see an important baseball game, and B, who is living in Boston calls A. Unfortunately, the weather in Boston is rainy and A says it's wonderful; the scene is at an airport.
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 at a public place.
Neutral declarative (DECL): A & B are colleagues, same age; A gives information without any personal perspective; the scene is at a coffee shop.
Neutral question (QUES): A & B are colleagues, same age. A asks for information, without any personal perspective, awaiting a simple answer. The scene is at a coffee shop.
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.
Politeness (POLI): A & B are almost the same age and don't know each other well, but work together professionally. A is sitting next to B; both start social talk. The scene is at a formal party.
Seduction (SEDU): A loves B and they have an intimate relationship. A gives a compliment to B in a sexually provocative way. The scene is at a clubhouse.
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.

Surprise (SURP): A & B are friends, same age. A didn't know that B can sing well. One day, B makes A listen to his beautiful voice. The scene is at friend's home.

Uncertainty (UNCE): A & B are colleagues, same age. A saw B at the baseball game yesterday, but is not 100% sure if it was really B; the scene is at a coffee shop.

Walking on eggs (WOEG): B is chief of the section which A belongs to; 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 it is 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.

Table 2: Characteristics of the audio-visual corpora and perceptual dataset used for each language: the total number of speakers in the corpus (Speakers: total number, and female / male), the place where recordings took place (Record.), the Target utterances' translation in each language, the number of presentations during the perceptual evaluation of each target utterance in a given language (# Pres.), the total number of speakers represented in the perceptual evaluation for a given language (Spk Perc.), and the number of participants in each perceptual evaluation (# Participants: total number, and female / male)

Language	Speakers	Record.	Target utterances	# Pres.	Spk Perc.	# Participants
American English	8 5f / 3m	Tokyo	A banana Mary was dancing	4 4	8	48 37m / 11m
Japanese	20 11f / 9m	Tokyo	バナナ (Banana) マリはダンスをしていました (Mari wa dansu wo shiteimashita)	4 4	17	27 19f / 8m
French	10 6f / 4m	Bordeaux	Une Banane Marie dansait	4 4	10	27 7f / 20m
Brazilian Portuguese	21 10f / 11m	Rio de Janeiro	Uma banana Maria dançava	4 4	20	22 16f / 6m
German	20 11f / 9m	Berlin	Eine Banane Marie tanzte	6 -	15	35 5f / 30m
Cantonese	10 6f / 4m	Hong Kong	香蕉 (hoeng1 ziu1) Mary跳緊舞 (Mary tiu3 gan2 mou5)	- 6	10	30 15f / 15m
Hindi	19 10f / 9m	Mumbai	एक केला (ek kela) मैरी नाच रही थी (mairee naach rahee thee)	4 4	8	35 14f / 21m
Total	*108* *59f / 49m*	-	-	-	*88*	*224* *113f / 111m*

Seven languages are considered here: American English (Rilliard et al. 2013, 2017), Japanese (Guerry et al. 2016a), French (Guerry et al. 2015), Brazilian Portuguese (Rilliard and de Moraes 2017), German (Hönemann et al. 2014; Mixdorff et al. 2017), Cantonese (Lee et al. 2018), and Hindi (Mixdorff et al. 2023, 2024). Table 2 details the characteristics of the corpus and perceptual datasets in each language for each of these languages. There is some variation across the corpus and perceptual datasets: they don't exactly follow the exact same procedure. The notable differences are in terms of the number of speakers (in some cases, it was more difficult to recruit them), obviously in terms of the utterances produced in each language, in two cases (for

German and Cantonese), the free labeling evaluation was applied only to one utterance ("a banana" for German, "Mary was dancing" for Cantonese), and for Hindi, most answers were given in English but for one participant (whose answers were translated into English). This last aspect is linked with the fact that these two experiments tested the role of the presentation channel (audio, visual, or audio-visual) while the others only used audio-visual stimuli. This impacts the number of stimuli used for each of the sixteen attitudes: while for American English, Japanese, French, Brazilian Portuguese, and Hindi, there are eight different performances for a given attitude presented in audio-visual (two utterances and four speakers by utterance), the German and Cantonese evaluations are based on six different performances (6 different speakers for one utterance; see latter for details).

The perceptual experiment is based on free labeling evaluations. A set of stimuli (generally the audio-visual performance of either the "a banana" or "Mary was dancing" utterance performed in a context prompting a specific prosodic attitude) produced by several speakers (4 or 6 for one attitude and a given utterance; see Table 2) enacting the 16 attitudes were proposed to participants (L1 speakers of each language) who were asked to describe the expression of the speaker, that is, what they were trying to convey, using a single word. The stimuli were selected based on performance evaluations, selecting for each attitude and utterance the speakers producing the more adequate behaviors according to a panel of judges (L1 speakers of the language considered, but for Hindi, the evaluators were from different locations in India, with varying linguistic backgrounds, and all but one answered in English; all were fluent in Hindi and English). Thus, one attitude is presented several times, produced by different individuals. The number of stimuli for one attitude was eight (two utterances, four speakers by utterance) for all languages but German and Cantonese, with six presentations (one utterance but 6 speakers; see above). Within a given language, the speakers performing two different attitudes are not necessarily the same: as the best speakers were selected, they vary with each attitude and in relation to the speakers' personalities (Erickson et al. 2018). For example, in Brazilian Portuguese, 20 speakers out of a total of 21 were presented, but some were presented just in a few cases, while the best performers were viewed for most attitudes (Rilliard and de Moraes 2017). There were also variations in terms of the presentation channel: while for American English, Japanese, French, and Brazilian Portuguese, the stimuli are presented audio-visually, for German, Cantonese, and Hindi, all the stimuli are presented audio-visually, and a subset is also presented in one channel only (audio only and visual only) to evaluate the potential role of these two channels. This aspect of audio-visual speech will not be dealt with here, and only the results of audio-visual presentations are considered.

The outputs of these perceptual tests are lists of labels given to each stimulus. As the replies of participants varied vastly –some used more than a word, paraphrased, introduced typos, differed in capitalizations, etc.– a normalization procedure was applied, independently for each study. It consisted in (i) keeping the first word when more than one was used (for lists of words, while paraphrases were kept), (ii) correcting typos and using lower cases, and (iii) converting labels appearing in their noun, adjective, or verb forms into the same form (noun or adjective depending on the dominant use of participants) when it was perceived to convey the same meaning (e.g., in Brazilian Portuguese, "*afirmação*, *afirmando*, *afirmar*, *afirmativa*, *afirmativo*" was encoded as "*afirmação*"). This normalization procedure produces lists of labels used to describe the different prosodic attitudes – one list for each language, with a different number of labels. These labels attributed to the sixteen different attitudes for each language group are used as entry data for the processing presented in this paper.

2.2 Statistical analysis

The lists of words were organized in large contingency tables, one for each language, that took as rows the 32 different production contexts (sixteen attitudes on two utterances) and as columns the list of words (one column per word; the number of columns varies for each language). Each cell contains the number of times a given word was used by the participants to characterize the performances of the four (or six for German and Cantonese) speakers that produced a given context (one utterance / one attitude). For example, the word "*fröhlich*" was used by the German raters presented with the "banana" utterance twice for the expression of politeness, seven times for irony, and never for uncertainty. As stated earlier, the German and Cantonese groups were presented with one utterance only (respectively, the translations of "*a banana*" and "*Mary was dancing*"): labels are absent for the other utterance; this reduces the importance of these languages in the analysis, but shall not change the qualitative distribution of attitudes.

Next, these seven 32 rows * *x* columns matrices (where *x* equals the total number of labels used for a given language: 268 for Brazilian Portuguese, 112 for German, 508 for Japanese, 275 for French, 91 for Cantonese, 528 for American English, and 313 for Hindi) were grouped row-wise into a very large matrix *X* (32 rows * 2095 columns). The somewhat lower number of labels for German and Cantonese shall be linked with the reduced number of contexts used here, but a large variability was observed across languages. The *X* matrix contains results for the 16 attitudes performed on each of the two target utterances on the rows and all the labels produced by the participants to describe them in the seven languages. The row for the expression of Doubt based on the "*Mary was dancing*" utterance is the aggregation of labels given for the performances of several different speakers: generally, four by language, except for German participants who were not presented with this utterance, and six for Cantonese (see Table 2). The exact speakers selected for each attitude differ according to their performance evaluation (see above). The results presented here are thus somehow representative of the targeted attitudinal expression in each language and how these expressivities are conceptualized by the participants based on audio-visual performances presented out of any interaction context.

The basic idea of the statistical analysis performed on this dataset is to project the 2095-long row vectors and the 32-long column vectors on abstract dimensions that explain most of their variance. This allows studying the proximity of labels and expressions (labels used frequently for a given expression) and between expressions or between labels (expressions or labels that are used in similar contexts) for the most explanatory part of the dataset, which obviously contains considerable "noise". For example, about 40% of the labels are used only once. As we are also interested in observing the specificities of each language group, we opted for using a Multiple Factor Analysis (MFA) in the sense of Escoffier Pages (Escofier and Pagès 1994). This multidimensional analysis method takes as entries various tables (that present coherent data within one table but heterogeneous ones across tables) and regroups them into one analysis in a two-step procedure: it first applies a multidimensional analysis on each sub-table (here the table for each language) and then regroups the analyses giving the same weight to each table, using as a grouping constraint the rows, which are the common "individuals" – different observations made on the same objects. Here, these individuals are the idealized prosodic attitudes corresponding to the communication goal of the speakers in each of the 16 contexts for each of the two target utterances (the 32 rows that contain data for performances made in similar contexts). Thus, the sub-tables are regrouped into a global multidimensional analysis (Principal Component Analysis, PCA) by aligning the different models for each language according to how they organize the rows of the *X* matrix (for details, see Bécue-Bertaut and Pagès 2008). The main PCA of the analysis spread the 32 rows of the *X* matrix onto a set of abstract dimensions that are

constructed on mixtures of the dimensions coming from the analysis of the sub-tables – from which it is possible to retrieve the relative weight of the columns (i.e., the original labels) related to each principal component.

The analysis was done using R (R Core Team 2023) and the FactoMineR library (Lê et al. 2008; Husson et al. 2013) that implements these analyses. The first eleven dimensions of the global analysis were kept using an elbow criterion; these 11 dimensions explain about 60% of the total variance, leaving the data with 40% of "noise". This amount of noise corresponds to the proportion of labels used only once; it is also in line with the quantity of noise found in another work studying the cognitive structure of emotional terms, which found about 30% of unexplained variance related to individual variation, errors, etc. (Moore et al. 1999). These eleven dimensions are abstract, and although it is always possible to try interpreting them, this raises complex challenges. To have a clearer view of the relations between the 32 expressions, an agglomerative hierarchical clustering was applied (using the HCPC procedure proposed by FactoMineR with a Euclidean metric and a Ward agglomeration method) to the cloud of attitudes along these eleven dimensions, leading to the dendrogram shown in Figure 1. This dendrogram can then be split at various places to produce different numbers of clusters. Two solutions will be presented here: (i) a three-cluster solution (cutting at about 5) that allows observing the main dimensions that organize this set of expressions and (ii) a finer 13-cluster solution (cutting at about 1) for which most clusters contain only one expression (i.e., the performances of the same attitude on the two target utterances) but for two complex clusters. The motivations for presenting these two solutions are that the three-cluster solution presents the main distinctions obtained from this data, thus providing a general view of the distribution of expressions, while the fine-grained 13-cluster solution presents subtle differences. Intermediary solutions would not have fit the available space. The labels that have a significant positive association (in terms of a v-test on their frequency inside the cluster *vs*. their global frequency; for details, see Husson et al. 2017) with each cluster for each language are then listed and used as the basis for interpreting the cognitive representation of the associated attitude for each group of participants.

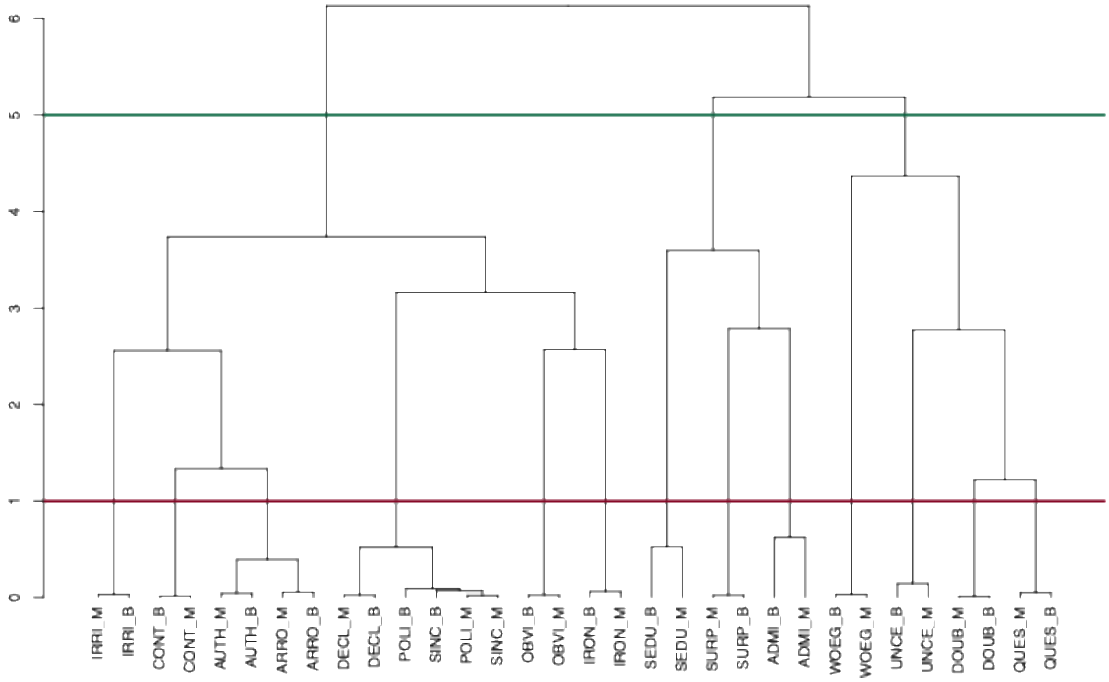

Figure 1: Dendrogram representing the distance between each attitude (abbreviations from Table 1) produced with a given utterance (B: "a banana"; M: "Mary was dancing" or translations); the vertical axis represents the agglomerative distance estimated from the MFA position of the attitudes, the colored horizontal lines represent the two sets of clusters considered here.

3 Results

A first result from the dendrogram plotted in Figure 1 is that the utterance ("Banana" or "Mary" and the contextual differences their elicitation may create) seems to play a minor role in the description of these prosodic attitudes: at the lowest level of the tree, one can observe a quasi-systematic pairing of the two utterances produced with the same attitude, but for SINC and POLI, for which the four categories are very closely grouped.

3.1 Three-cluster solution

If one is interested in the main differences across attitudes –in their organizational dimensions– this solution offers the following regroupings (reading the tree from left to right): (i) IRRI, CONT, AUTH, ARRO, DECL, POLI, SINC, OBVI, IRON, (ii) SEDU, SURP, ADMI, and (iii) WOEG, UNCE, DOUB, QUES. The labels associated with these three clusters for each language are given in Table 3 in decreasing order of association. The table is limited to the first five labels, if they exist, for space reasons and because less frequent labels are generally associated with part of the cluster only: for example, for cluster (i), the labels "*impaciência*" (impatience) associated to Irritation by Brazilian Portuguese raters and "factual" associated to cluster (4) (that contains the expressions of Declaration, Politeness and Sincerity; see below) by American English raters.

Table 3: labels positively associated with each of the three-cluster solutions for each language (Brazilian Portuguese: BP; Japanese: JP; French: FR; German: DE; Cantonese: ZH; American English: US; Hindi: HI); labels are listed in decreasing order of association

	Labels associated to each cluster, in each language (*English translation*)
(i)	***IRRI, CONT, AUTH, ARRO, DECL, POLI, SINC, OBVI, IRON***
BP	afirmação, certeza, confirmação, normalidade, resposta (*affirmation, certainty, confirmation, normality, answer*)
JP	無表情, 普通, 怒り, 横柄, ふてくされ (*expressionless, normal, angry, arrogant, sulky*)
FR	affirmation, certitude, déclaration, arrogance, confiance (*affirmation, certainty, declaration, arrogance, confidence*)
DE	feststellend, überzeugt, entschlossen (*ascertaining, convinced, determined*)
ZH	平和, 正常, 反感, 不耐煩, 不高興 (*peaceful, normal, disgust, impatient, unhappy*)
US	response, confident, assertive, explaining, sure
HI	assertive, statement, confident, information, angry
(ii)	***SEDU, SURP, ADMI***
BP	alegria, satisfação, felicidade, fascínio, contentamento (*joy, satisfaction, happiness, fascination, contentment*)
JP	喜び, 嬉しい, 驚嘆, 驚き, 感心 (*joy, glad, amazement, surprise, admiration*)
FR	heureux, perversion, envie, joie, plaisir (*happy, perversion, desire, joy, pleasure*)
DE	begeistert, sinnlich, mysteriös, naiv, erregt (*enthusiastic, sensual, mysterious, naive, excited*)
ZH	沾沾自喜, 欺騙, 滿足, 期待, 震驚 (*complacency, deception, satisfaction, anticipation, shock*)
US	joy, overjoyed, excited, ecstatic, wonder
HI	cheerful, elated, happy, admiration, excited
(iii)	***WOEG, UNCE, DOUB, QUES***
BP	dúvida, insegurança, palpite, incerteza, interrogação (*doubt, insecurity, hunch, uncertainty, question*)
JP	疑う, 疑問, 疑念, 疑い深い, 困惑 (*doubt, question, suspicion, suspicious, puzzled*)
FR	incertitude, doute, hésitation, interrogation, question (*uncertainty, doubt, hesitation, questioning, question*)
DE	unsicher, fragend, unglaubwürdig, ängstlich, zurückhaltend (*uncertain, questioning, unbelievable, fearful, reserved*)
ZH	疑惑, 猶豫, 不肯定, 懷疑, 沒有信心 (*doubt, hesitation, uncertainty, suspicion, lack of confidence*)
US	hesitant, confused, unsure, questioning, doubt
HI	doubtful, doubt, confusion, embarrassed, apprehensive

Cluster (i) is primarily opposed to cluster (iii), as (ii) and (iii) split at a later stage, and cluster (iii) keeps the features (labels) of the larger cluster while cluster (ii) is more specifically composed around SEDU SURP and ADMI. This opposition between clusters (i) and (iii) corresponds to the first dimension of the MFA that opposes the expressions of these two clusters. This dimension, or this opposition of clusters (i) vs. (iii), is reminiscent of the *Unpredictability* dimension proposed by Fontaine et al. (2007). Cluster (i) is described prominently with terms linked to the notion of assertion (like "*afirmação*, *festellend*, *response*, *assertive*") or to a neutral

expression (like "*normalidade*, *無表情*, *平和*, *information*"): it carries features linked to the assertive illocution. On the contrary, cluster (iii) is described as carrying doubt, uncertainty, or questions (with labels such as "*dúvida*, *疑う*, *incertitude*, *unsicher*, *疑惑*, *hesitant*, *doubtful*"): this fits with the notion of unpredictability and with the interrogative illocution.

Cluster (ii) is separated from the two others along at least two of the abstract dimensions of the MFA (dimensions 2 and 3). The description of this label is primarily based on labels carrying positive associations. If considering the first ones for all languages, then we have "*alegria*, *喜び*, *heureux*, *begeistert*, *沾沾自喜*, *joy*, *cheerful*." Other features are present, with some related to sexuality (e.g. "*perversion*, *sinnlich*") that are linked to Seduction (see latter), and some that carry a high arousal component (e.g., "*驚嘆*, *erregt*, *overjoyed*, *excited*"). The specificity of the cluster (ii) is related to the Evaluation-Pleasantness and Activation-Arousal dimensions, with expressions at the positive and aroused end of the dimension.

Let's note that the Potency-Control dimension may also play a role in this distribution, with the opposition between clusters (i) and (iii) having related features (e.g., "*怒り*, *entschlossen*" vs. "*insegurança*, *apprehensive*"). We do not base our analysis here on this dimension as it is not always found and is better observed within each cluster, with the opposition, for example, between Irritation and Politeness or between Doubt and Uncertainty.

3.2 Thirteen-cluster solution

The 13-cluster solution was selected on an inertia reduction criterion and a maximization of the clusters containing a single attitude, without separating the two utterances. This solution has 11 clusters composed of a single attitude, plus two complex clusters: one with the AUTH and ARRO expressions and one with the DECL, POLI, and SINC expressions. At this level, it is thus possible to study 11 single expressive situations in terms of the labels that are more commonly used to describe them and to observe some expressions that are more difficult to separate based on the proposed labels.

The first five labels associated with these thirteen clusters are reproduced in Table 4. We limit the description to five labels (when five are available) for space reasons and also because the less common labels were used only once (i.e., by a single participant), which limits their reach – even if in a few cases, some paraphrases were very accurate in capturing a complex meaning, as in the "*afirmação para pergunta idiota*" (reply to a stupid question) used for Obviousness by one Brazilian Portuguese participant. The analysis of these thirteen clusters will be done hereafter for each one. We'll try to discuss their characteristics along the four main dimensions of expressive meaning discussed in the Introduction, as they can be inferred from the labels used to describe each expression. We'll also rely on the structure of the dendrogram presented in Figure 1, which gives hints at the proximities between the expressions. For example, clusters (1) Irritation, (2) Contempt and (3) Authority / Arrogance share some features and are closer together (they split together on the dendrogram) than they are to (5) Obviousness and (6) Irony. We will also try to underline some cross-cultural differences when they are reflected in the labels used to define these expressions.

Table 4: labels positively associated with each of the 13-cluster solutions for each language (Brazilian Portuguese: BP; Japanese: JP; French: FR; German: DE; Cantonese: ZH; American English: US; Hindi: HI); labels are listed in decreasing order of association

	Labels associated to each cluster, in each language (*English translation*)
(#1)	***IRRI***
PB	agressividade, exaltação, impaciência, irritação, nervosismo
	(aggressiveness, excitement, impatience, irritation, nervousness)
JP	子供あつかい, イライラ, 苛立ち, 面倒くさい, 呆れる
	(Treating someone like a child, irritation, frustration, bother, disappointment)
FR	impatience, énervement, colère, exaspération, insistance
	(impatience, irritation, anger, exasperation, insistence)
DE	wütend, genervt, verärgert, angespannt
	(angry, annoyed, upset, tense)
ZH	嫌棄, 不重視, 煩躁, 強調, 緊張
	(Dislike, ignore, irritate, stress, nervous)
US	frustrated, reiterating, angry, irritated, impatient
HI	anger, authority, disdain, arrogance, forceful
(#2)	***CONT***
PB	nojo, desprezo, desgosto, rejeição, descaso
	(disgust, contempt, disgust, rejection, disregard)
JP	つかれ, すかし, なげやり, 面倒, 飽き
	(tired, careless, troublesome, bored)
FR	ennui, grognon, fatigue, blase, sectaire
	(boredom, grumpy, tired, jaded, sectarian)
DE	verbittert, unsympathisch, unhöflich, unfreundlich, missmutig
	(bitter, unpleasant, rude, unfriendly, sullen)
ZH	失望, 無所謂, 不滿意, 疲倦, 鬱悶
	(Disappointed, indifferent, dissatisfied, tired, depressed)
US	disappointed, sad, bored, uninterested, contempt
HI	interested, blank, sentimental, regret, irritable
(#3)	***AUTH, ARRO***
PB	seriedade, atestação, rigidez, arrogância, diretamente
	(seriousness, attestation, rigidity, arrogance, directly)
JP	威圧, 強迫, 偉そう, ふてくされ, 嫌嫌
	(intimidating, coercive, bossy, sulky, reluctant)
FR	rage, désapprobation, autoritaire, contraint, fermeté
	(rage, disapproval, authoritarian, constrained, firmness)
DE	stolz, arrogant, elitär, dramatisch, trotzig
	(proud, arrogant, elitist, dramatic, defiant)
ZH	惡, 不高興, 冷漠, 自信, 果斷
	(Evil, Unhappy, Indifferent, Confident, Decisive)
US	resolved, definite, stern, blunt, tough
HI	angry, serious, authoritative, upset, arrogant
(#4)	***DECL, POLI, SINC***
PB	afirmação, resposta, certeza, neutralidade, normalidade

	(affirmation, answer, certainty, neutrality, normality)
JP	普通, 落ち着く, 事実, 密告, 真面目
	(normal, calm, fact, inform, serious)
FR	confirmation, affirmation, acquiescement, réponse, déclaration
	(confirmation, affirmation, acquiescence, response, declaration)
DE	demonstrativ, gleichgültig, neutral, erklärend, aufrichtig
	(demonstrative, indifferent, neutral, explaining, sincere)
ZH	平靜, 陳述, 平和, 確定
	(calm, statement, peaceful, certain)
US	content, kind, factual, realization, mellow
HI	statement, neutral, confidence, information, normal
(#5)	***OBVI***
PB	simplicidade, obviedade, explicação, resistência, afronta
	(simplicity, obviousness, explanation, resistance, affront)
JP	反発, 軽視, 警告, 突っ張っている, 当然
	(rebellion, disregard, warning, stubborn, natural)
FR	évidence, logique, fatalisme, banalité, assurance
	(obviousness, logic, fatalism, banality, assurance)
DE	spöttisch, desinteressiert, offensichtlich, erschrocken, glücklich
	(sardonic, disinterested, obvious, frightened, happy)
ZH	認真, 解惑, 恍然大悟, 回答, 平常
	(Serious, solve doubts, suddenly realize, answer, normal)
US	obvious, teasing, relaxed, teaching, patient
HI	direct, obviousness, expressive, suspicious, satire
(#6)	***IRON***
PB	reforço, sarcasmo, ironia, deboche, modo como maria dançava
	(reinforcement, sarcasm, irony, mockery, the way Maria danced)
JP	笑い, けいしょう, おかしい, 嘲笑, バカにする
	(laugh, make fun of, ridicule, make fun of)
FR	désabusé, sarcastique, précieux, méchant, impertinent
	(disillusioned, sarcastic, precious, mean, impertinent)
DE	zurechtweisend, abwägend, ironisch, amüsiert, fröhlich
	(reprimanding, weighing, ironic, amused, cheerful)
ZH	重覆, 笑, 敷衍, 奸狡, 可笑
	(repetitive, laugh, perfunctory, cunning, funny)
US	humorous, funny, amused, tolerant, savoring
HI	sorrow, unbelievable, shocking, insulting, genuine
(#7)	***SEDU***
PB	sensualidade, orgulho, maldosa, paixão, sedução
	(sensuality, pride, evil, passion, seduction)
JP	感傷, 恍惚, 楽しい, 嬉しい, うれしい
	(sentimental, ecstatic, joyful, happy)
FR	sous-entendu, curiosité, amour, plaisir, séduction
	(innuendo, curiosity, love, pleasure, seduction)
DE	verfuehrerisch, seltsam, schleimig, goennerhaft, geheimnisvoll

	(seductive, strange, slimy, patronizing, mysterious)
ZH	認同, 自豪, 自大, 敬畏, 感嘆
	(Recognition, pride, vanity, awe, admiration)
US	sexual, intrigued, flirty, suggestive, sly
HI	dreamy, alluring, seductiveness, seductive, whisper
(#8)	***SURP***
PB	espanto, empolgação, surpresa, perplexidade, falsidade
	(astonishment, excitement, surprise, perplexity, falsehood)
JP	強い驚き, 驚愕, 仰天, 驚き, びっくり
	(to be very surprised, astonishment, look up the heavens, surprise, surprise)
FR	surprise, étonnement, ahurissement, intrigué, interpelé
	(surprise, astonishment, bewilderment, intrigued, challenged)
DE	dankend, überrascht, erstaunt, belehrend, angeekelt
	(grateful, surprised, amazed, instructive, disgusted)
ZH	震驚, 錯愕, 驚訝, 不相信, 奇怪
	(Shocked, stunned, surprised, unbelievable, strange)
US	surprised, shocked, interrogative, genuine, dismayed
HI	surprised, curiosity, astonished, jealousy, shock
(#9)	***ADMI***
PB	maravilhado, fascínio, encantamento, contentamento, alegria
	(amazed, fascination, enchantment, contentment, joy)
JP	興奮, 感動, 歓喜, 相手に訴えかける感じ, 憧れ
	(Excitement, emotion, delight, appeal to others, admiration)
FR	extase, soulagement, émerveillement, rêveur, envie
	(ecstasy, relief, wonder, dreamer, desire)
DE	träumerisch, spontan, schwärmend, sarkastisch, provokant
	(dreamy, spontaneous, enthusiastic, sarcastic, provocative)
ZH	興奮, 滿足, 友善, 高興, 沾沾自喜, 欺騙, 激動, 期待
	(Excited, Contented, Friendly, Happy, Smug, Deceitful, Thrilled, Anticipation)
US	love, emotional, thrilled, overjoyed, relieved
HI	amusement, excited, elated, admiration, scary
(#10)	***WOEG***
PB	constrangimento, confissão, admitindo, acanhamento, vergonha
	(embarrassment, confession, admitting, shyness, shame)
JP	自白, 怯える, がっかり, 残念, 申し訳ない
	(confess, be afraid, be disappointed, be sorry, be sorry)
FR	fautif, gêné, honte, peur, regret
	(guilty, embarrassed, ashamed, afraid, regret)
DE	überspielend, zurückhaltend, enttäuscht, unangenehm, ängstlich
	(overplaying, reserved, disappointed, unpleasant, anxious)
ZH	驚慌, 難以啓齒, 為難, 有口難言, 害羞
	(panic, difficult to speak, embarrassed, speechless, shy)
US	nervous, apologetic, reluctant, embarrassed, uneasy
HI	fearful, embarrassed, scared, fear, frightened
(#11)	***UNCE***

PB	incerteza, insegurança, pena, hipótese, habilidoso
	(uncertainty, insecurity, pity, hypothesis, skillful)
JP	不確実, 戸惑い, 躊躇, 自信がない, 不安, 迷い
	(Uncertainty, confusion, hesitation, lack of confidence, anxiety, indecision)
FR	supposition, inquiétude, doute, mystère, indéterminé
	(supposition, worry, doubt, mystery, indeterminate)
DE	verängstigt, sanft, niedergeschlagen, lügend, abgeneigt
	(frightened, gentle, dejected, lying, averse)
ZH	不肯定, 猶豫
	(Unsure, Hesitant)
US	unsure, remembering, believe, dumbfounded, assisting
HI	unsure, tiring, suspicion, remember, confused
(#12)	***DOUB***
PB	incredulidade, desconfiança, descrédito, descrença, estranhamento
	(incredulity, distrust, discredit, disbelief, strangeness)
JP	疑い, 不信, 疑念, 懐疑的, 疑い深い
	(doubt, disbelief, suspicion, skepticism)
FR	dubitatif, incompréhension, sceptique, incrédulité, impressionne
	(doubtful, incomprehension, skeptical, incredulity, impressed)
DE	unwissend, zweifelnd, verwirrt, unglaubwürdig, skeptisch
	(ignorant, doubtful, confused, unbelievable, skeptical)
ZH	不同意, 懷疑, 不相信, 疑問
	(Disagree, doubt, disbelief, question)
US	doubt, personification, negation, gross, ensuring
HI	astounded, curious, reserved, narrative, interrogate
(#13)	***QUES***
PB	pergunta, interrogação, desconhecimento, questionamento, palpite
	(question, interrogation, ignorance, inquiry, guess)
JP	軽い不思議, 質問, 疑問, 確認, 不思議
	(light wonder, question, doubt, confirmation, wonder)
FR	indécision, question, interrogation, diverti, découverte
	(indecision, question, interrogation, entertained, discovery)
DE	zynisch, fragend, hervorhebend, interessiert
	(cynical, questioning, emphasizing, interested)
ZH	指責, 反問, 疑惑, 明白, 無所謂
	(accuse, ask, doubt, understand, don't care)
US	perplexed, inquisitive, questioning, curious, peppy
HI	terrible, ques, cool, bothered, satisfaction

3.2.1 Cluster #1: Irritation

Cluster #1 contains all the performances of situations leading to an expression of Irritation with both target utterances. These expressions were described quite coherently by all seven language groups. A recurrent feature found in the labels links to the dimension of dominance, with attitudes perceived as impositions by the raters (labels "*aggressiveness*, *treating someone like a child*, *exasperation*, *upset*, *nervous*, *angry*, *forceful*"), linked to a state of stress (labels "*impatience*, *stress*") and an expressivity that shares features with the emotion of anger (labels "*nervous,*

frustration, anger, upset, angry"). It also expresses an important arousal (labels "*excitement, insistence, tense, forceful*"). Some labels also recall features of the prototypical situation, like "*reiterating, insistence, bother*".

3.2.2 Cluster #2: Contempt

Cluster #2 contains all expressions of Contempt produced with both utterances. These expressions showed variations in their description across languages. As for the Evaluation-Pleasantness scale, they are mostly described negatively, but for the Hindi dataset that received mixed labels ("*interested*" vs. "*irritable*"). This negative stance may play on different attributes, with Japanese, French, Cantonese, and American English using labels such as "*bored, boredom, tired, indifferent, sad, depressed*" (i.e., low arousal marks lack of engagement), while Brazilian Portuguese and German used more active labels signaling rejection with features linked to the emotion of disgust ("*disgust, rejection, bitter, unpleasant, rude*"). Both of these features may constitute a mark of exclusion of the addressee, which may be viewed as a dominant behavior but without active aggression patterns. Some labels used by the Hindi raters are linked to this pattern ("*blank, irritable*"), but the portrait in their case is less clear, since the remaining labels (after the first five reported in Table 4) tend to show similar patterns (we give them here: "*impatient, contentment, careless, bitchy, disgusted, disappointment, egotistic, annoyance, disgust, hesitating, sulk, hopeless, comment, stoic, ignorance, fatigue, gratifying, ashamed, irritation, sad, disappointed, declaration, arrogant*").

3.2.3 Cluster #3: Authority, Arrogance

Cluster #3 mixes all expressions of Authority and Arrogance. Let's note that the two expressions are still distinguished at a lower level (cf. Figure 1), but not considered here: if they share most of their labels, they still have specificities that some raters note. The prevalent scale for the labels is the Potency-Control one, with a series of labels referring to an active imposition on the addressee ("*directly, coercive, bossy, authoritarian, resolved, decisive, stern, blunt, authoritative*") that is related to Authority. Still on dominance, but with features of self-distinction from the addressee (and not imposition on them) is a second series of labels that are prominently used by the German group (which does not have labels in the former list) like "*proud, arrogant, elitist*" that are also found in the other languages ("*seriousness, arrogance, sulky, constrained, indifferent, confident, serious, arrogant*") and are related to Arrogance. On the Evaluation-Pleasantness scale, labels mark negative features that go along the emotion of anger ("*intimidating, rage, evil, unhappy, angry, upset*"), an emotion that carries aggressive action tendencies (Scherer 2009b). A feature in half the languages can be interpreted along the Activation-Arousal scale. It denotes the energy of the expression, marking its strength ("*rigidity, firmness, dramatic, tough*").

These first three clusters are grouped together at a higher level (cutting the tree of Figure 1 near 3): within the large group of assertive expressions, they can be seen as a subpart of them carrying features of negative valence and imposition.

3.2.4 Cluster #4: Declaration, Politeness, Sincerity

Cluster #4 regroups the expressions of Declaration, Politeness, and Sincerity. Note that at a lower level in the dendrogram, Declarations are separated from the two polite expressions. These two polite expressions are deeply mixed and are the only cases where the target utterance (which cues the scenario used for eliciting these expressions) appears to have a stronger role than the target attitude: the stimuli are first associated by utterance, then by attitude.

In terms of labels, the cluster is mostly described as "normal" ("*neutrality, normality, normal, calm, neutral, peaceful, neutral*") and informative: these expressions are typical of expected (normal) behavior in everyday unmarked encounters. Their primary goal appears to be giving reliable information to the addressee directly or answering a request ("*affirmation, answer, fact, inform, confirmation, affirmation, response, declaration, explaining, statement, factual*") – and they are typical of an assertive illocution. Some shades of expressivity are noted that are related to polite expressions ("*serious, demonstrative, sincere, peaceful, certain, kind, mellow*"). These labels carry a positive undertone for these expressions along the Evaluation-Pleasantness scale. Polite attitudes here are seen as a (positive) coloring of the underlying assertive illocutions; they could have been associated with, e.g., interrogations, and their global localization may have been very different.

3.2.5 Cluster #5: Obviousness

Cluster #5 is based on the expressions of Obviousness. The communication goal is relatively well-spotted, with five in seven languages using "obvious" as a possible label (all but Japanese and Cantonese). Nonetheless, the expression shows notable variations in its description. Languages that use the "obvious" radical also use terms like "*simplicity, explanation, logic, banality, disinterested, teaching, patient, direct*" to convey that the information is neither new nor complex. Three of them (German, American English, Hindi) also use lexemes such as "*sardonic, happy, teasing, relaxed, satire*" that introduce a notion of humoristic criticism. An opposite view (by Brazilian Portuguese and Japanese) is based on more negative aspects, such as "*resistance, affront, rebellion, disregard, stubborn,*" keeping the criticism, not the humor. The French and Cantonese are more neutral on valence, reinforcing the factual aspect of the performances ("*assurance, serious, answer, normal*").

3.2.6 Cluster #6: Irony

Cluster #6, which is grouped with cluster #5 at a higher level, is based on the performances of Irony. The inversion of meaning characteristic of the ironic process is mentioned through the use of labels such as "*irony, ironic, unbelievable*," and it is often characterized by features of humor (see the label "*laugh*") – but on an Evaluation-Pleasantness scale, this humor may be perceived as positive ("*amused, cheerful, cunning, funny, humorous, tolerant, savoring*" for German, Cantonese, and American English) or negative ("*sarcasm, mockery, make fun of, ridicule, sarcastic, mean, impertinent, sorrow, shocking, insulting*" for Brazilian Portuguese, Japanese, French, and Hindi).

3.2.7 Cluster #7: Seduction

Cluster #7 belongs at a higher level to the group of active and positive attitudes (see part 3.1) together with the next two clusters, #8 and #9. If the prototypical situation has an explicit sexually oriented objective, this aspect of the performance does not appear in the labels given by Japanese and Cantonese raters that focus on more general positive expressions like "*ecstatic, joyful, happy, recognition, pride, awe, admiration*"; such labels marking a positive interest are also found in other languages ("*pride, curiosity, mysterious, strange, intrigued, dreamy*"). Nonetheless, all the other language groups note explicitly the sexually oriented aspect of the performances, using "*seduction*" or "*seductive*" for four of them and a series of sexually related labels, but that can bear a positive ("*sensuality, passion, innuendo, love, pleasure, flirty, suggestive, alluring*") or

negative ("*evil, slimy, patronizing*") valence, plus some ambivalent ones ("*sexual, sly*"). Japanese and Cantonese also used a label more related to Seduction ("*sensuality*") or its tentative ("*vanity*").

3.2.8 Cluster #8: Surprise

Cluster #8 is based on expressions of Surprise. The fact it falls within the "positive" expression is interesting, as in other analyses, it tends to participate in the "unpredictability" group; let's note this is the case here if one considers only two clusters, but this discrepancy underlines the positive aspect of the performances that elicit Surprise in these corpora, where the prototypical situations are rather positive indeed (cf. Table 1 and Rilliard et al. 2013). A first observation is the perfect match of the expression and its name in each language: "surprise" is used by all groups. Other common labels are emphasized synonyms: "*astonishment, to be very surprised, bewilderment, amazed, shocked, stunned, dismayed, astonished, shock*". This set of labels reinforces the strong arousal linked with the expression. Another set of labels ("*perplexity, intrigued, challenged, instructive, unbelievable, strange, interrogative, curiosity*") matches with the unpredictability dimension (see above) and the notion of novelty linked to the expression. Finally, the other labels tend to mark negatively ("*falsehood, look up the heavens, disgusted, strange, jealousy*") valued performances, apart from a few positives ("*grateful, genuine*"). With the many terms directly linked to surprise cited above not having a clear valence, it is difficult to attribute a clear position to these expressions on the Evaluation-Pleasantness scale: it is rather its strong arousal that shall explain its regrouping within the cluster (ii) described above.

3.2.9 Cluster #9: Admiration

Cluster #9 contains the expressions of Admiration. These expressions are described mostly by two sets of labels. The largest one is based on positively valued terms ("*fascination, enchantment, contentment, joy, delight, appeal to others, admiration, ecstasy, relief, wonder, dreamer, desire, dreamy, contented, friendly, happy, smug, love, thrilled, overjoyed, relieved, amusement, elated*") and the second of terms linked to high arousal ("*amazed, excitement, emotion, ecstasy, relief enthusiastic, excited, emotional, thrilled, overjoyed, elated*"); note some labels are in both lists). A few negative labels are also used ("*sarcastic, provocative, scary*"), that may be linked to the high level of arousal of the expressions, which may be difficult to interpret outside of their interaction contexts.

3.2.10 Cluster #10: WOEG

Cluster #10 is part of the larger group carrying "unpredictability" features, together with clusters #11, #12, and #13. It is based on the performances of "walking-on-eggs", a situation derived from the Japanese concept of *kyoshuku* (恐縮), which has, according to Shochi et al. (2023), the literal meaning of "*shrink fear*" and constitutes a politeness strategy used to address complex situations with a superior in order to show respect and "suffering" when imposing something on them. It is described 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*" (Sadanobu 2004, p. 34). The prototypical situation used to elicit the expression derives from such descriptions. It is first notable that, as a politeness strategy, it was not grouped within, or close to, cluster #4, which contains the other two politeness strategies (Politeness and Sincerity, cf. Shochi et al., 2023, for a discussion). The labels are coherent across language groups with a large set that fits the semantics of shame and confession ("*embarrassment, confession, admitting, shame, confess, be sorry, guilty, embarrassed, ashamed, regret, unpleasant, difficult to speak, apologetic, uneasy*") carrying a negative valence, and a group expressing submissive behavior along the Potency-Control dimension with labels such as "*shyness, be afraid, reserved,*

anxious, panic, shy, nervous, fearful, scared, dear, frightened". These two sets of features notably respect the definition of the *kyoshuku* expression, which contains aspects of fear and apology.

3.2.11 Cluster #11: Uncertainty

Cluster #11 contains the expressions of Uncertainty. The communication goal is captured adequately with the use of terms like "*uncertainty, supposition, unsure*," in line with features of hesitations typical of the unpredictability dimension, with labels such as "*hypothesis, confusion, hesitation, indecision, doubt, mystery, indeterminate, averse, hesitant, believe, dumbfounded*". Another set of features may be viewed as negative on the Evaluation-Pleasantness scale and submissive on the Potency-Control one: "*insecurity, pity, lack of confidence, anxiety, worry, frightened, dejected, tiring.*"

3.2.12 Cluster #12: Doubt

Cluster #12, which contains the Doubt expressions, is quasi-exclusively described by labels expressing the rejection of or questioning a preceding assertion: "*incredulity, disbelief, strangeness, distrust, discredit, doubt, suspicion, skepticism, doubtful, skeptical, unbelievable, disagree, negation, reserved, interrogate*". Questioning a preceding assertion is an act of imposition and, thus, is dominant on the Potency-Control scale and expresses a negative valence.

3.2.13 Cluster #13: Question

The last cluster, #13, contains the expressions of Question. It is well described, with most terms linked to the interrogative illocution ("*question, interrogation, ignorance, inquiry, guess, confirmation, indecision, discovery, questioning, interested, ask, understand, inquisitive, curious, ques*"). Some labels show the proximity with the expressions of surprise and doubt ("*wonder, doubt, perplexed*")

4 Discussion

We presented the results of two clustering solutions for the distribution of attitudinal expressions obtained from their description using a free labeling paradigm in seven languages. This method (the Multiple Factor Analysis) explains about 60% of the total variance in labels – a rather large quantity, but in line with results obtained through other methods for similar purposes – comparing the distribution of perceived "meaning" linked to affect (for example Moore et al. 1999). The paradigm in itself is a source of fuzziness, and the stimuli used in these experiments are not written words as in Moore et al. (1999) but performances collected from 88 different speakers from seven different languages and labeled by more than 200 individuals. It is thus expected to find some unexplained variation. This variation is also a source of information if we compare it with results obtained from identification paradigms that propose a given list of predefined answers from which the participants have to pick the best possible match (e.g., Aubergé et al. 1997; Shochi et al. 2009b; Gu et al. 2011). With such an approach, the translated labels may mask potential cultural differences because they are interpreted differently by individuals of different cultural backgrounds (Wierzbicka 1986, 1992, 2004). The lists of labels we obtained here give a more nuanced portrait of the different performances that allow observing cultural differences and convergences, and are unbiased by experimenter-induced choices.

The first important result, which validates our collection and evaluation of prosodic attitudes, is that the sixteen expressions are regrouped together, regardless of the target utterance that was used ("a banana" or "Mary was dancing"). This finding that the performances produced on different utterances through different dialogues have almost no effect on the classification

obtained from open labeling produced for performances of different performers and of different languages strongly underlines the important role that audiovisual prosody (Swerts and Krahmer 2005) plays in interspeaker interactions for conveying attitudinal expressions.

A second important result is the presence and importance of the "unpredictability" dimension in organizing these expressions. Different from Fontaine et al. (2007), for which dataset it is the fourth dimension in importance, this dimension explains the main split in our data (see Figure 1); the fact this distinction fits the difference between assertive and interrogative illocutions is indicative of the importance of this dimension for the interactive and linguistic nature of attitudes (Mello and Raso 2011). One may raise the question of how to interpret the abstract dimension that splits the dendrogram into two large groups: Fontaine et al. (2007) found it prototypically related to the emotion of Surprise (or "appraisals of novelty and unpredictability", Fontaine et al. 2007: 1051). In our case, it corresponds to the opposition of labels like "affirmation, certainty, declaration, affirmation, certainty, declaration" to "uncertainty, doubt, hesitation, confused, question, unsure" (see Table 3, clusters (i) and (iii)) – thus, an opposition between "assertive" and "interrogative" groups. Although these labels themselves lack features of novelty and unpredictability, they can be viewed as reactions to events that are predictable or not: thus, the assurance of the assertive group (the speaker knows how to behave), while with the "interrogative" group, speakers are trying to find an adequate strategy to respond to something unexpected. This is typical of the expression of Doubt: what was just said is unexpected, and the speaker thinks it is erroneous. These expressions could be viewed as behaviors responding to interactions that are less predictable. A reminder that we work here with attitudinal expressions, not emotions; an important difference is the integration of attitudes within linguistic interactions, which require longer cognitive processing (Levinson 2016) than the appraisals linked to, for instance, the emotion of Surprise in Scherer's model (Scherer 2009a). The attitude of Surprise is a linguistically formulated expression in response to an unpredictable event. As for the relationship between the emotional feature and the prototypical, a machine learning experiment applying emotion recognition models to the Hindi attitudinal dataset found that Surprise is the most accurately classified category (Khatri et al. 2025). It is interesting, however, that the attitude of Surprise (prototypical example of the dimension in Fontaine et al., 2017, work) is not prominent in this "unpredictable" cluster (even if it is part of it in a two-cluster solution); this may be linked to its positive evaluation (which is linked to the specific elicitation dialogues), that brought it closer to the Seduction / Admiration set along the Evaluation-Pleasantness dimension. It is also interesting that this dimension was not described by previous works based on isolated words (Osgood et al. 1957, 1975; Romney et al. 1996; Moore et al. 1999): the discursive nature of illocutions is observed at a higher level that allows for dialogic interactions (de Moraes 2011; Bossaglia et al. 2020). We hope our contribution here shall reinforce the perceived importance of prosodic cues (as defined by Swerts and Krahmer, 2005) in the community for these levels of linguistic meaning. The other three dimensions (Evaluation-Pleasantness, Potency-Control, Activation-Arousal) are also present in the data, but at lower levels, as has been detailed during the analysis of the 13-cluster solution: these dimensions are fundamental in organizing the different expressions, but they are highly dependent of each attitude's specificities. Let's summarize the main findings.

The clusters #1 (Irritation), #2 (Contempt), and #3 (Authority, Arrogance) are regrouped together at a higher level (cutting the tree of Figure 1 near 3) and belong to the even larger group of assertive expressions (cluster (i)). They can be seen as a subpart of the set of assertions, carrying negative valence and imposition features. Within these negative impositions, some more subtle differences have been described that differentiate, for example, an imposition by someone having

an authoritative role (e.g., a policeman) from an imposition between hierarchically comparable individuals (as in Irritation). The acoustic profiles of both expressions are different (with much more vocal effort in the latter case; see Erickson et al., 2024) and their acceptance by the addressee may also differ in relation to the expectation of these behaviors. We see here that the attitudinal performances are deeply intertwined within each interaction's social and pragmatic features. In the case of contempt, its performances seem to signal impoliteness in the sense of Culpeper et al. (2003), with different strategies being employed by the performers and described by the raters – strategies that target either the positive or the negative face of the interlocutor by impeding their social inclusion, or by actively rejecting them from the social group. On the contrary, the expressions regrouped within cluster #3 (Authority and Arrogance) also express a social superiority with respect to the hearer, but without denying them their social inclusion: they express an imposition but not necessarily in an impolite (i.e., face-threatening) way.

Cluster #4 (Declaration, Politeness, Sincerity) is the more complex one, regrouping three types of expressions (Declaration, Politeness, and Sincerity), and is often described as "normal". One may view polite expressions as the expected behavior during social interactions with unknown individuals, thus their "normality": they are expected. A courteous behavior would be unmarked in this definition of normality, and it may not be very different from the "neutral" assertion. As a rather positive set of expressions, cluster #4 is opposed to the preceding group (clusters #1 to #3), which is negatively valued. It also differs regarding dominance, being neutral on the Potency-Control dimension. If the first dimension was "unpredictability", the others also have an important role. Let's note that the Frequency Code (Ohala 1994) postulates that polite expressions tend to be performed with a raised pitch to mark the absence of aggressivity by displaying a more submissive tone of voice (and not a low dominant voice). If a higher pitch is indeed observed in the productions linked to this experiment for Politeness and Sincerity (Rilliard and de Moraes 2017; Erickson et al. 2018, 2024), features of "submission" were not found through the labels. The expressions of cluster #4 are perceived as "*positive, kind, sincere, peaceful, mellow, calm*" but do not contain explicitly submissive labels (note peaceful and calm demonstrations may fit Ohala's view) – if there is no imposition pattern either.

Clusters #5 (Obviousness) and #6 (Irony) are grouped in the same sub-group within the larger set of assertions (cluster (i)). In this sub-group, an important feature is linked to the expression of humor, and the potential interpretations of such a humoristic strategy within an interaction may vary cross-culturally, as it may receive opposed interpretations along the Evaluation-Pleasantness scale. The fact to signal the obviousness of an assertion may be used at face value for Obviousness or interpreted at a second degree in the case of Irony. Its social significance seems to be culture-dependent, with positive or negative valence attributed to the performances. This cultural discrepancy in the valence attribution to an attitude that is otherwise correctly identified is an important aspect of the complexity of attitudinal expressions and their interactional use in foreign language contexts – thus, the importance of their teaching in foreign language classes (Shochi et al. 2009a, 2010, 2016; Guerry et al. 2016b). Such discrepancies may be related to the different communication ethos across cultures (Kerbrat-Orecchioni 2022).

Within the cluster (ii), the expression of Seduction (cluster #7) was grouped in the set of positive performances. Meanwhile, it also received a series of negative ratings ("*evil, slimy, patronizing*"). This may be related to individual views on the appropriateness of such behavior by a given individual with respect to oneself, which may explain part of the variation in the labels. There seem to be language groups rating Seduction explicitly and positively (French, Hindi), while others (Japanese, Cantonese) gave positive ratings but not explicit ones, and others gave mixed ratings (Brazilian Portuguese, German, American English) with variation in the number of

positive or negative labels. A French lover may be viewed as a patronizing bastard on the other side of the Rhine (we do not present cross-cultural comparisons here, though; see e.g., Shochi et al. 2009b; Mixdorff et al. 2017, 2020; Rilliard et al. 2017).

Surprise (regrouped in cluster #8) is accurately named by all language groups, and it is the only category that is named so accurately. This reinforces the prototypicality of this expression, which has already been noted to receive high identification scores (de Moraes et al. 2010) and is well-recognized cross-culturally (Shochi et al. 2010). This may be related to the corresponding emotion (surprise), a reaction to an unexpected stimulus. In Scherer's model (2009a), the predictions for the first evaluation check (novelty) in the case of novel stimuli define a physiological reaction for facial expressions that fits the prototypical performance of the attitude ("[…] pupillary dilatation, local muscle tonus changes; brows and lids up […]" Scherer, 2009a: 3464). In the social expression counterpart, speech is not interrupted as in the emotional reaction but demonstrates vocal source characteristics coherent with the physiological settings (high energy and fundamental frequency, notably; cf. Goudbeek and Scherer 2010). Cluster #9 (Admiration) is typically positively valued and with a high rating of arousal. Since there is not much variation across language groups in its perceptual evaluation, it may be interesting to look more closely at the performances to see which aspects are coherent and which are divergent across languages.

The group corresponding to cluster (iii), which may be related to the "unpredictable" side of the unpredictability dimension, comprises three low-level clusters. Walking-on-egg (cluster #10) was grouped within this large cluster and not together with the other two politeness strategies within cluster #4 (Declaration, Politeness, Sincerity), which is part of the "assertive" group (cluster (i) – the "predictable" side of the dimension). Cluster #4 contains positively valued assertions. This separation across politeness strategies is in line with multidimensional studies of the perception of politeness and impoliteness strategies in Japanese that observed close perceptual proximity between neutral declaration, politeness, and sincerity (as if they were the normal, expected behavior for unmarked encounters), opposed to impolite expressions along a polite–impolite dimension where "politeness" is considered as a mark of courtesy (marking an in-group recognition). In such a scheme, the expression of *kyoshuku* (from which derive the WOEG situation, see part 3.2.10) does not fit this dimension of courtesy politeness but is organized by participants' rating the cognitive distance between such expressions on another orthogonal dimension (Rilliard et al. 2014; Shochi et al. 2023). The interpretation that was made of such a semantic structure of prosodic use for (im)politeness strategies is, using the framework of politeness theory proposed by Brown and Levinson (1987), that Politeness and Sincerity are based on linguistic strategies of positive politeness by demonstrating respectively courtesy and sincerity toward the addressee, while *kyoshuku* is a strategy of negative politeness trying to redress a Face Threatening Act by showing suffering and indicating reluctance in performing such an attitude (see also Hill et al. 1986; Ide 2002).

The other expressions in the cluster (iii) are Uncertainty (#11), Doubt (#12), and Question (#13). Cluster #11 shares a negative valence on the Evaluation-Pleasantness scale and a submissive rating on the Potency-Control one with cluster #10, but is chiefly characterized by hesitations. On the other side of this cluster (iii) are two types of interrogations: one (Doubt) that interrogates the truth value of a preceding assertion, and a second one (Question) that genuinely asks for information. The expression of Doubt was rated negatively on the Evaluation-Pleasantness scale, as clusters #10 and #11, but as a dominant behavior (contradicting the addressee), which separates it from the two previous expressions: it is an imposition on the interlocutor. Questions were more neutral on these two scales.

5 Conclusions

The paper regrouped the experimental data of attitudinal expressions collected in studies of seven languages and compared the perception of these affects in terms of their communication goals, as perceived by listeners of each of these languages. Based on multidimensional methods, this work seeks to observe the distribution of different prosodic attitudes and compare their perceptual proximity. This allows, following the proposal of Romney and colleagues (Romney et al. 1996; Moore et al. 1999), to draw a representation of cultural representations in terms of such expressions used in interpersonal spoken exchanges. We do not choose to represent the spread of these attitudes on "maps" (as Romney did), mainly because it is best represented with a high dimensionality that does not fit a 2D plane; we preferred a hierarchical representation using the dendrogram of Figure 1. We do not try either to represent the variation: each leaf of the dendrogram is one expression that regroups in a somehow hidden manner the performances of several speakers from different linguistic backgrounds. This was done because the labels were given in different languages, and their statistical comparison through translation is far from trivial (it raises difficulties with the tendency to use adjectives or nouns in different languages, for example, which would be represented as different categories even if having very close meanings).

Given these limitations, we focused on working with the lists of labels associated with each cluster of performances. This was done because these lists are rich in information about the multifaceted nature of our expressivity, which is produced with shades of meaning by each speaker and interpreted with even more variations by the receivers. This depends notably on each individual's personality (Erickson et al. 2018) and habit of playing or eliciting a character (Sadanobu 2012, 2015) with varying vocal characteristics. A major result that goes through all these sources of variation is the robustness of the within-attitude classification obtained: all performances on the two target utterances are grouped together, a fact that was not guaranteed. This shows there are specific prosodic patterns that can be reliably perceived and described and that allow a neat structuration of our attitudinal behaviors. Another important finding is related to the obtained spread of attitudes – that follows the dimensions long advocated for "meaning" in general (Osgood et al. 1957) and, more specifically, for affective meaning (Osgood et al. 1975), with a major role for the "fourth" dimension of unpredictability (Fontaine et al. 2007). The important role of this dimension and its motivation by our data (which has basically a dialogical function) shows that the pragmatic notion of illocution certainly has a major communicative role in relation to these dimensions of interactive meaning, as advocated by Mello and Raso (2011).

REFERENCES

1. Aubergé V, Grepillat T, Rilliard A. Can we perceive attitudes before the end of sentences? the gating paradigm for prosodic contours. In: 5th European Conference on Speech Communication and Technology (Eurospeech 1997). Rhodes, Greece: ISCA; 1997. p. 871–4.
2. Bécue-Bertaut M, Pagès J. Multiple factor analysis and clustering of a mixture of quantitative, categorical and frequency data. Comput Stat Data Anal. 2008 Feb;52(6):3255–68.
3. Bossaglia G, Mello H, Raso T. Chapter 7. Illocution as a unit of reference for spontaneous speech: An account of insubordinated adverbial clauses in Brazilian Portuguese. In: Izre'el S, Mello H, Panunzi A, Raso T, editors. Studies in Corpus Linguistics [Internet]. Amsterdam: John Benjamins Publishing Company; 2020. p. 221–56. Available from: https://benjamins.com/catalog/scl.94.07bos
4. Brown P, Levinson SC. Politeness: some universals in language usage. Cambridge [Cambridgeshire] ; New York: Cambridge University Press; 1987. (Studies in interactional

sociolinguistics).
5. Chang NCT. Tones and Intonation in the Chengtu Dialect (Szechuan, China). Phonetica. 1958 Jan 1;2(1–2):59–85.
6. Culpeper J, Bousfield D, Wichmann A. Impoliteness revisited: with special reference to dynamic and prosodic aspects. J Pragmat. 2003 Oct;35(10–11):1545–79.
7. Erickson D, Rilliard A, de Moraes J, Shochi T. Personality Judgments Based on Speaker's Social Affective Expressions. In: Fang Q, Dang J, Perrier P, Wei J, Wang L, Yan N, editors. Studies on Speech Production [Internet]. Cham: Springer International Publishing; 2018. p. 3–13. (Lecture Notes in Computer Science; vol. 10733). Available from: http://link.springer.com/10.1007/978-3-030-00126-1_1
8. Erickson D, Rilliard A, Thurgood E, Moraes J, Shochi T. Acoustic and perceptual profiles of American English social affective expressions: a case study. J Speech Sci. 2024 Dec 18;13:e024004.
9. Escofier B, Pagès J. Multiple factor analysis (AFMULT package). Comput Stat Data Anal. 1994 Aug;18(1):121–40.
10. Fónagy I. Vocal Expression of Emotions and Attitudes. Quad Studi Semiot. 1987;47/48:65–85.
11. Fónagy I, Bérard E. «Il est huit heures»: contribution à l'analyse sémantique de la vive voix. Phonetica. 1972;26(3):157–92.
12. Fónagy I, Bérard E, Fónagy J. Clichés mélodiques. Folia Linguist. 1983;17(1–4):153–86.
13. Fontaine JRJ, Scherer KR, Roesch EB, Ellsworth PC. The World of Emotions is not Two-Dimensional. Psychol Sci. 2007 Dec;18(12):1050–7.
14. Fujisaki H, Hirose K. Analysis and perception of intonation expressing paralinguistic information in spoken Japanese. In: ESCA Workshop on Prosody. Lund, Sweden; 1993. p. 254–7.
15. Goudbeek M, Scherer K. Beyond arousal: Valence and potency/control cues in the vocal expression of emotion. J Acoust Soc Am. 2010;128(3):1322–36.
16. Gu W, Zhang T, Fujisaki H. Prosodic analysis and perception of Mandarin utterances conveying attitudes. In: Interspeech 2011 [Internet]. ISCA; 2011. p. 1069–72. Available from: https://www.isca-archive.org/interspeech_2011/gu11_interspeech.html
17. Guerry M, Rilliard A, Erickson D, Shochi T. Perception of prosodic social affects in Japanese: A free-labeling study. In: Speech Prosody 2016 [Internet]. ISCA; 2016a. p. 811–5. Available from: https://www.isca-speech.org/archive/speechprosody_2016/guerry16_speechprosody.html
18. Guerry M, Shochi T, Rilliard A, Erickson D. Perception of prosodic social affects in French: a free-labeling study. In: 18th International Congress of Phonetic Sciences [Internet]. Glasgow, UK; 2015. p. 4p. Available from: https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/proceedings.html
19. Guerry M, Shochi T, Rilliard A, Erickson D. Perception of Prosodic Social Affects in Japanese by American Learners of Japanese. In: International Symposium on Applied Phonetics (ISAPh 2016) [Internet]. ISCA; 2016b. p. 84–8. Available from: https://www.isca-speech.org/archive/isaph_2016/guerry16_isaph.html
20. Hill B, Ide S, Ikuta S, Kawasaki A, Ogino T. Universals of linguistic politeness: Quantitative evidence from Japanese and American English. J Pragmat. 1986 Jun;10(3):347–71.
21. Hirst D, Di Cristo A. A survey of intonation systems. In: Hirst D, Di Cristo A, editors. Intonation systems: a survey of twenty languages. Cambridge, U.K.: Cambridge University Press; 1998. p. 1–44.
22. Hirst DJ. Form and function in the representation of speech prosody. Speech Commun. 2005;46(3/4):334–47.
23. Hönemann A, Mixdorff H, Rilliard A. Social Attitudes - Recordings and Evaluation of an audio-visual Corpus in German. In: 7th Forum Acusticum 2014, Krakow. 2014.
24. Husson F, Josse J, Le S, Mazet J. FactoMineR: Multivariate Exploratory Data Analysis and Data Mining with R [Internet]. 2013. Available from: http://CRAN.R-project.org/package=FactoMineR
25. Husson F, Lê S, Pagès J. Exploratory multivariate analysis by example using R. Second

edition. Boca Raton: CRC Press; 2017.
26. Ide S. The Speaker's Viewpoint and Indexicality in a High Context Culture. In: Kataoka K, Ide S, editors. Bunka, Intaakushon, Gengo [Culture, Interaction, and Language]. Tokyo: Hitsuji Shobo; 2002. p. 3–20.
27. Kerbrat-Orecchioni C. Politeness in France: How to buy bread politely. In: Hickey L, Stewart M, editors. Politeness in Europe. Clevedon, UK: Multilingual Matters; 2005. p. 29–44.
28. Kerbrat-Orecchioni C. A multilevel approach in the study of talk-in-interaction. Pragmat Q Publ Int Pragmat Assoc IPrA. 2022 Jul 6;7(1):1–20.
29. Khatri P, Mixdorff H, Rao P, Rilliard A. Recognition of audio-visual attitudes. In: Grawunder S, editor. Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2025 [Internet]. TUDpress, Dresden; 2025. p. 19–26. Available from: https://www.essv.de/pdf/2025_19_26.pdf
30. Lê S, Josse J, Husson F. **FactoMineR** : An *R* Package for Multivariate Analysis. J Stat Softw [Internet]. 2008;25(1). Available from: http://www.jstatsoft.org/v25/i01/
31. Lee T, Ma KHM, Rilliard A, Mixdorff H, Hönemann A. Free Labeling of Audio-visual Attitudinal Expressions in Cantonese. In: Speech Prosody 2018 [Internet]. ISCA; 2018. p. 483–7. Available from: https://www.isca-speech.org/archive/speechprosody_2018/lee18_speechprosody.html
32. Levinson SC. Turn-taking in Human Communication – Origins and Implications for Language Processing. Trends Cogn Sci. 2016 Jan;20(1):6–14.
33. Lu Y, Aubergé V, Rilliard A. Do you hear my attitude? prosodic perception of social affects in Mandarin. In: Speech Prosody 2012. ISCA; 2012. p. 685–8.
34. Mac DK, Aubergé V, Rilliard A, Castelli E. Audio-visual prosody of social attitudes in vietnamese: building and evaluating a tones balanced corpus. In: Interspeech 2009. ISCA; 2009. p. 2263–6.
35. Martins-Baltar M. De l'énoncé à l'énonciation: une approche des fonctions intonatives. Paris: Didier; 1977.
36. Mello H, Raso T. Illocution, modality, attitude: different names for different categories. In: Mello H, Panunzi A, Raso T, editors. Pragmatics and Prosody: Illocution, Modality, Attitude, Information Patterning and Speech Annotation [Internet]. 1st ed. Firenze: Firenze University Press; 2011. p. 1–18. (Strumenti per la didattica e la ricerca; vol. 120). Available from: https://books.fupress.com/isbn/9788866550846
37. Miranda L, Swerts M, de Moraes JA, Rilliard A. The Role of the Auditory and Visual Modalities in the Perceptual Identification of Brazilian Portuguese Statements and Echo Questions. Lang Speech. 2021 Mar;64(1):3–23.
38. Mixdorff H, Ghosh D, Rilliard A, Hönemann A. Perception of Audio-visual Expressions in German and Cantonese by Native Speakers of Hindi. In: Speech Prosody 2020 [Internet]. ISCA; 2020. p. 31–5. Available from: https://www.isca-speech.org/archive/speechprosody_2020/mixdorff20_speechprosody.html
39. Mixdorff H, Hönemann A, Rilliard A, Lee T, Ma MKH. Audio-visual expressions of attitude: How many different attitudes can perceivers decode? Speech Commun. 2017 Dec;95:114–26.
40. Mixdorff H, Nayan N, Rilliard A, Rao P, Ghosh D. Developing a corpus of audio-visual attitudinal expressions in Hindi. In: Proceedings of the 20th ICPhS [Internet]. Prague; 2023. p. 4150–4. Available from: https://guarant.cz/icphs2023/631.pdf
41. Mixdorff H, Rilliard A, Nayan N. Perceptual Evaluation of Attitudinal Expressions. In: 5th International Symposium on Applied Phonetics (ISAPh 2024) [Internet]. ISCA; 2024. p. 60–4. Available from: https://www.isca-archive.org/isaph_2024/mixdorff24_isaph.html
42. Moore CC, Romney AK, Hsia TL, Rusch CD. The Universality of the Semantic Structure of Emotion Terms: Methods for the Study of Inter- and Intra-Cultural Variability. Am Anthropol. 1999 Sep;101(3):529–46.
43. de Moraes JA. The pitch accents in Brazilian portuguese: analysis by synthesis. In: Speech Prosody 2008. ISCA; 2008. p. 389–97.
44. de Moraes JA. From a prosodic point of view: remarks on attitudinal meaning. In: Mello H, Panunzi A, Raso T, editors. Pragmatics and Prosody: Illocution, Modality, Attitude, Information

Patterning and Speech Annotation [Internet]. 1st ed. Firenze: Firenze University Press; 2011. p. 19–37. (Strumenti per la didattica e la ricerca; vol. 120). Available from: https://books.fupress.com/isbn/9788866550846
45. de Moraes JA, Rilliard A. Illocution, attitudes and prosody: A multimodal analysis. In: Raso T, Mello H, editors. Studies in Corpus Linguistics [Internet]. Amsterdam: John Benjamins Publishing Company; 2014. p. 233–70. Available from: https://benjamins.com/catalog/scl.61.09mor
46. de Moraes JA, Rilliard A. Prosody and Emotion in Brazilian Portuguese. In: Armstrong ME, Henriksen N, del Mar Vanrell M, editors. Issues in Hispanic and Lusophone Linguistics [Internet]. Amsterdam: John Benjamins Publishing Company; 2016. p. 135–52. Available from: https://benjamins.com/catalog/ihll.6.07mor
47. de Moraes JA, Rilliard A, Mota BADO, Shochi T. Multimodal perception and production of attitudinal meaning in Brazilian Portuguese. In: Speech Prosody 2010. Chicago, IL, USA: ISCA; 2010. p. 1–4.
48. Ohala JJ. The frequency code underlies the sound-symbolic use of voice pitch. In: Hinton L, Nichols J, Ohala JJ, editors. Sound Symbolism [Internet]. 1st ed. Cambridge University Press; 1994. p. 325–47. Available from: https://www.cambridge.org/core/product/identifier/CBO9780511751806A036/type/book_part
49. Osgood CE, May WH, Miron MS. Cross-cultural universals of affective meaning. University of Illinois Press; 1975.
50. Osgood CE, Suci GJ, Tannenbaum PH. The measurement of meaning. Urbana: University of Illinois Press; 1957.
51. Pierrehumbert J, Hirschberg J. The Meaning of Intonational Contours in the Interpretation of Discourse. In: Cohen PR, Morgan J, editors. Intentions in communication. Cambridge, Mass.: MIT Press; 1990. p. 271–311. (System development foundation benchmark series).
52. Post B, d'Imperio M, Gussenhoven C. Fine phonetic detail and intonational meaning. In Saarbruecken, Germany; 2007. p. 191–6.
53. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2023. Available from: https://www.R-project.org/
54. Rilliard A, Erickson D, de Moraes JA, Shochi T. Cross-cultural perception of some Japanese politeness and impoliteness expressions. In: Baider F, Cislaru G, editors. Linguistic Approaches to Emotions in Context [Internet]. Amsterdam: John Benjamins Publishing Company; 2014. p. 251–76. (Pragmatics & Beyond New Series; vol. 241). Available from: https://benjamins.com/catalog/pbns.241.15ril
55. Rilliard A, Erickson D, de Moraes JA, Shochi T. Perception of expressive prosodic speech acts performed in USA English by L1 and L2 speakers. J Speech Sci. 2017 Nov 1;6(1):27–45.
56. Rilliard A, Erickson D, Shochi T, de Moraes JA. Social face to face communication — American English attitudinal prosody. In: Interspeech 2013 [Internet]. ISCA; 2013. p. 1648–52. Available from: https://www.isca-speech.org/archive/interspeech_2013/rilliard13_interspeech.html
57. Rilliard A, de Moraes JA. Social affective variations in Brazilian Portuguese: a perceptual and acoustic analysis. Rev Estud Ling. 2017 Jun 13;25(3):1043–74.
58. Romney AK, Boyd JP, Moore CC, Batchelder WH, Brazill TJ. Culture as shared cognitive representations. Proc Natl Acad Sci. 1996 May 14;93(10):4699–705.
59. Romney AK, Moore CC, Rusch CD. Cultural universals: Measuring the semantic structure of emotion terms in English and Japanese. Proc Natl Acad Sci. 1997 May 13;94(10):5489–94.
60. Russell JA. Culture and the categorization of emotions. Psychol Bull. 1991;110(3):426.
61. Sadanobu T. A natural history of Japanese pressed voice. J Phon Soc Jpn. 2004;8(1):29–44.
62. Sadanobu T. An Unofficial Guide for Japanese Characters [Internet]. 2012. Available from: http://dictionary.sanseido-publ.co.jp/wp/author/sadanobu-e/
63. Sadanobu T. "Characters" in Japanese Communication and Language: An Overview. Acta Linguist Asiat. 2015 Dec 29;5(2):9–28.
64. Scherer KR. Emotions are emergent processes: they require a dynamic computational architecture. Philos Trans R Soc B Biol Sci. 2009a Dec 12;364(1535):3459–74.

65. Scherer KR. The dynamic architecture of emotion: Evidence for the component process model. Cogn Emot. 2009b Nov;23(7):1307–51.
66. Shochi T, Brousse A, Guerry M, Erickson D, Rilliard A. Learning effect of social affective prosody in Japanese by French learners. In: Speech Prosody 2016 [Internet]. ISCA; 2016. p. 970–3. Available from: https://www.isca-speech.org/archive/speechprosody_2016/shochi16_speechprosody.html
67. Shochi T, Gagnié G, Rilliard A, Erickson D, Aubergé V. Perception of prosodic French social affects for Japanese learners of French language. Trans Tech Comm Psychol Physiol Acoust Acoust Soc Jpn. 2009a;39.
68. Shochi T, Gagnié G, Rilliard A, Erickson D, Aubergé V. Learning effect of prosodic social affects for Japanese learners of French language. In: Speech Prosody 2010. ISCA; 2010. p. 1–4.
69. Shochi T, Rilliard A, Aubergé V, Erickson D. Intercultural perception of English, French and Japanese social affective prosody. In: Hancil S, editor. The role of prosody in affective speech. Bern, Switzerland: Peter Lang AG; 2009b. p. 31–60. (Linguistic Insights).
70. Shochi T, Rilliard A, Erickson D. Chapter 8. Perceptual changes between adults and children for multimodal im/politeness in Japanese. In: Jucker AH, Hübscher I, Brown L, editors. Multimodal Im/politeness: Signed, spoken, written [Internet]. Amsterdam: John Benjamins Publishing Company; 2023. p. 213–49. (Pragmatics & Beyond New Series; vol. 333). Available from: https://benjamins.com/catalog/pbns.333.08sho
71. Spencer-Oatey H. Reconsidering power and distance. J Pragmat. 1996 Jul;26(1):1–24.
72. Swerts M, Krahmer E. Audiovisual prosody and feeling of knowing. J Mem Lang. 2005 Jul;53(1):81–94.
73. Uldall E. Attitudinal meanings conveyed by intonation contours. Lang Speech. 1960;3(4):223–34.
74. Wichmann A. The attitudinal effects of prosody, and how they relate to emotion. In: ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion. 2000. p. 143–8.
75. Wierzbicka A. Human emotions: universal or culture-specific? Am Anthropol. 1986;88(3):584–94.
76. Wierzbicka A. Defining emotion concepts. Cogn Sci. 1992;16(4):539–81.
77. Wierzbicka A. Japanese cultural scripts: Cultural psychology and “cultural grammar.” Ethos. 1996;24(3):527–55.
78. Wierzbicka A. “Happiness” in cross-linguistic & cross-cultural perspective. Dædalus. 2004;133(2):34–43.