Classification of Auditory-Visual Attitudes in German

Angelika Hönemann[1], *Hansjörg Mixdorff*[2], *Albert Rilliard*[3]

[1] University of Bielefeld, CITEC, Germany, [2] Beuth University Berlin, Germany,
LIMSI-CNRS, Orsay, France [3]

ahoenemann@techfak.uni-bielefeld.de, mixdorff@bht-berlin.de, Albert.Rilliard@limsi.fr

Abstract

This paper presents results from an auditory-visual recognition experiment employing short utterances of German produced with varying attitudinal expressions. It is based on 16 different kinds of social and/or propositional attitudes which place speakers in various social interactions with a partner of inferior, equal or superior status, and having a communication aim with a positive, neutral or negative, valence. Data from ten German subjects were classified by native perceivers regarding the attitude portrayed. Participants were given five choices: The intended attitude, two closely related attitudes, and two randomly chosen ones. Higher recognition scores were obtained in audio-visual presentations (45%), over 36% with audio-only stimuli. The best recognized attitudes were doubt, (neutral) statement, surprise and irritation which all yielded audio-visual recognition scores over 50%. Lowest recognition scores were obtained for irony, 'walking-on-eggs' and politeness. A hierarchical clustering based on correspondence analysis showed that groupings of stimuli in one cluster are consistent with their original labels - these consistent stimuli yield better recognition scores. Conversely, clusters with heterogeneous populations simply aggregate bad performances.

Index Terms: A/V perception, social attitudes

1. Introduction

Human communication almost always has a social goal. Above and beyond pure linguistics, information about e.g. the mental state, emotions, mood or attitudes of the collocutors is exchanged during the dialog. The affective state is influenced, for instance, by the situation or roles of the dialog partners. Mutual understanding of the social intention between communication partners should not be difficult as long as they belong to the same language or culture. In contrast, interaction between partners from different cultures sometimes leads to misinterpretations of social expressions: it has been shown that the verbal and non-verbal expressions depend, to some extent, on the culture in which we grow up. A study by Shochi et al. investigated twelve social attitudes (e.g. surprise, irritation, command-authority) for prosodic effects in British English, French and Japanese [1]. It found similarities across these languages, but also some culture-specific uses of prosodic parameters, typically in Japanese-specific expressions of politeness. Interestingly, these confusions observed in an audio-only condition are disambiguated when presenting non-Japanese listeners with the audio-visual performances: if cultural differences exist in the conceptual interpretations, the prosodic performances are interpreted without noticeable differences between Japanese and non-Japanese listeners, especially in multimodal presentations [2]. It has been proposed that visual information may give accurate contextual information to decode specific prosodic changes [3,4]. Intercultural comparison of linguistic and paralinguistic effects has enjoyed growing attention as the knowledge about how verbal and non-verbal social affect are expressed in different languages is paramount for mutual understanding between cultures.

A main obstacle to the ecological study of social affect lies in the need to record such data with reasonably high quality while maintaining a certain level of spontaneity. To this effect and to facilitate the speaker's task, [5] proposes to place target sentences in affectively loaded texts; similarly, [6] recorded attitudinally-neutral sentences embedded into dialogues that prepare the speaker to perform an adequate expression for the target sentence. An important issue here is the adequate labeling of attitudes elicited as the associated terminology will vary between languages [7].

The current work is based on the framework developed by [8] in which attitudes are characterized by their situational descriptions, taking into account between whom, where, and with what aim they occur. A difference from [6] is that recordings also concern the visual channel, as facial gestures are known to be a vital part of attitudinal expressions [3]. In the following section this approach will be discussed in more detail. Based on this protocol, two instances of 16 different attitudes were elicited from a total of 20 native speakers of German. In a recent study we had native German subjects rate the credibility of the expressions portrayed by the first ten of the speakers [9]. In the current paper we present a perception experiment in which German native listeners were asked to classify these attitudes, presented either in auditory-visual or audio-only modality. Results for the various expressions and for each modality are then analyzed and discussed. The aim of this study is to establish a representation of the perceptual similarities among these various expressions, find main expressive dimensions, and relate them to the two modalities used to express them.

2. Speech Data Elicitation

Sixteen attitudes such as arrogance, politeness, doubt or irritation (see Table 1 for a complete list and abbreviations henceforth used in this paper) were elicited through short dialogs which ended in the target sentences 'Eine Banane' (engl. *a banana*) or 'Marie tanzte' (engl. *Marie was dancing*). Preceding the target dialog, a test dialog was performed in order to prepare the speakers and help them immerse themselves in the context of the attitude. These dialogs were

designed according to different social situations differing in social and linguistic aspects such as the type of speech act (propositional/social) [10], hierarchical distance, social distance or valence of speech act (cf. [11] for details). Among these situations, some correspond to culture-specific concepts. For example, the situation coined WOEG reproduce what could correspond to *kyoshuku* in Japanese – a concept that has no direct translation in English or in German [1]. Conversely, the situation coined SEDU did not correspond to any prototypical behavior among Japanese male speakers. Presenting these concepts through the same situations and dialogs across cultures, one can compare the expressive strategies of speakers of various origins, and the perception of these strategies cross-culturally. We focus here on German.

All 20 native German subjects (11 female, 9 male) participating had academic background, were asked to produce the sixteen attitudes twice and paid for their time. Ages ranged from 20 to 60 with a median of 31.5 years.

3. Design of the Recognition Experiment

Since our corpus contains 32 utterances by each of the ten speakers we had to limit the number of stimuli to be presented in the recognition experiment. Results from [9] suggested that the sentence was irrelevant for the performance ratings therefore we selected either "eine Banane" or "Marie tanzte" from each speakers. We created two sets of 160 stimuli each, one containing audio-visual stimuli (AV) by speakers 1-5 and audio-only (AU) from speakers 6-10, the other set the reverse.

Attitude		**associated with:**	
admiration	ADMI	seductiveness	politeness
arrogance	ARRO	contempt	authority
authority	AUTH	arrogance	irritation
contempt	CONT	arrogance	irritation
(neutral) statement	DECL	politeness	irony
doubt	DOUB	uncertainty	surprise
irony	IRON	doubt	obviousness
irritation	IRRI	authority	contempt
obviousness	OBVI	contempt	irony
politeness	POLI	sincerity	walking-on-eggs
(neutral) question	QUES	doubt	uncertainty
seductiveness	SEDU	admiration	politeness
Sincerity	SINC	politeness	admiration
surprise	SURP	doubt	(neutral) question
uncertainty	UNCE	walking-on-eggs	doubt
walking-on-eggs	WOEG	uncertainty	sincerity

Table 1: List of attitudes (presented in German) and their closest associates.

In order to reduce the search space for the participants' answers, we did not offer all 16 choices of attitudes, but reduced the number to five, including the intended attitude, two closely related ones (see Table 1), and two randomly chosen attitudes from the 13 other ones. However, contradicting attitudes such as question and statement were excluded from the random choices.

Participants were 21 German speaking students of Beuth University (16m, 5f, age: 18-37).

4. Results

We yielded in total 1680 answers from the participants for each modality which are pooled in a matrix and expressed: (1) either as proportion of good answers (a good answer being the choice of the intended attitude), (2) as a dispersion matrix, with the presented stimuli in the rows, and the 16 possible labels in the columns. The intersection of rows and columns presents the count of how often the 21 subjects answered using one label for a given stimulus.

4.1 Proportion of good answers

The proportions of good answers were analyzed with the method of deviance analysis [12]. The analysis took the proportion of good and wrong answers for each stimulus as the dependent variable, and the attitude, the modality and the speaker as fixed factors. This complete model is then simplified [12]. The model that best explains the data is composed of the following factors: attitude, modality and speaker and two-way interactions between them – each with significant effects on the results.

The factors that explain most of the deviance are the modality of presentation and the targeted attitude. The mean recognition rate of the attitudes presented in audio-only (37%) is lower than that of audio-visual presentations (45%). Attitudes receiving the highest recognition scores, above 50% (whatever the modality and the speaker) are (in decreasing order) DECL, SURP, DOUB and IRRI; those receiving the lowest scores (below 20%) are WOEG and POLI. Speakers also show varying performance, with recognition levels between 28 and up to 48%. This suggests an important variability in the way attitudes are performed and recognized.

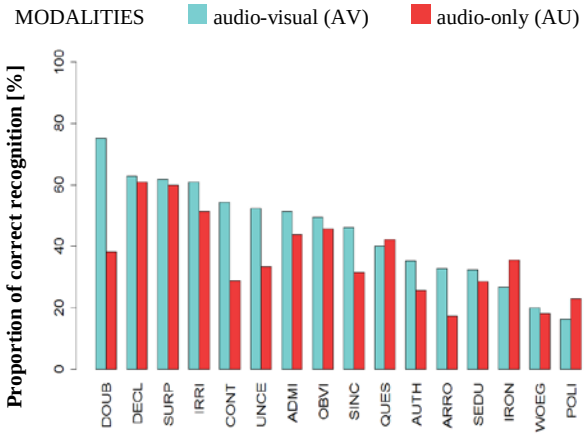

Figure 1: Proportion of recognition (%) of 16 attitudes in the modality audio-only (AU) and audio-visual (AV)

Figure 1 shows recognition rates for the 16 attitudes presented in audio-only and audio-visual condition. Complex interactions between modality and attitudes can be observed. For DECL, SURP and QUES, audio-only and audio-visual

presentations receive similar recognition scores: there is thus little supplementary information brought by the visual modality – which is not the case for most of the other attitudes where the A/V presentations show higher scores than the audio-only ones. On the contrary, QUES, POLI and IRON yielded better performances in audio-only presentation. Hence in these cases visual information somehow contradicts the audio.

Figure 2 displays the percentage of recognition for the ten speakers in each modality. Only two speakers (S07, S10) were recognized better when presented audio-only - even if only slightly. This indicates that strategic choices of the speakers may be involved when a given speech act is conveyed through the acoustic and/or visual channel.

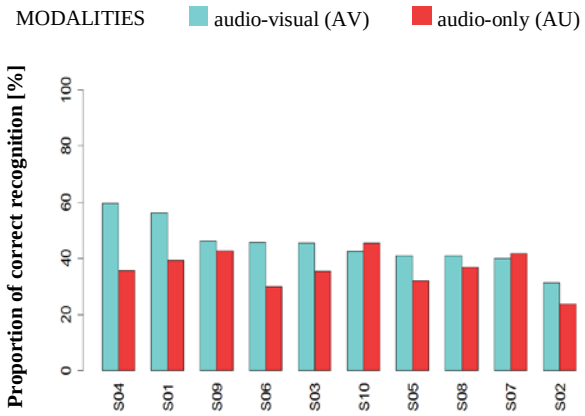

Figure 2: Proportion of correct recognition (%) for the ten speakers in the audio-only (AU) and audio-visual (AV) modality.

The attitudes CONT and DOUB show the largest auditory-visual gain. Figure 3 displays examples of these two expressions performed by two of the best-rated performers and best-recognized speakers S01 and S04.

Figure 3: Examples expressions CONT (above) and DOUB (below) performed by speakers S01 (left) and S04 (right).

4.2 Correspondence Analysis and Clustering

We applied correspondence analysis (CA) to find the main relations between the presented stimuli and the labels chosen by the participants. Prior to the analysis the raw scores in the contingency table have to be normalized. The main problem is that labels were presented an unequal number of times for each stimulus. For a given stimulus, the intended attitude and the two associated ones (see Table 1) were provided as choices each time; the other labels were presented a random number of times, i.e. in practice less frequently. To account for this difference, the raw number of answers given by listeners to a label is corrected according to the individual number of presentations in the following way:

$$NRS_{ans}(s_i, l_j) = \frac{N_{ans}(s_i, l_j)}{5 * N_{label}(s_i, l_j)} \quad (1)$$

where $\mathrm{NRS_{ans}}(\mathrm{s_i}, \mathrm{l_j})$ is the normalized recognition score (NRS) of label $\mathrm{l_j}$ to be used as an answer for stimulus $\mathrm{s_i}$; $\mathrm{N_{ans}}(\mathrm{s_i}, \mathrm{l_j})$ is the actual number of answers using label $\mathrm{l_j}$ for stimulus $\mathrm{s_i}$. $\mathrm{N_{label}}(\mathrm{s_i}, \mathrm{l_j})$ is the actual number of times label $\mathrm{l_j}$ was presented to subjects for stimulus $\mathrm{s_i}$. $\mathrm{N_{label}}(\mathrm{s_i}, \mathrm{l_j})$ is multiplied by 5 to account for the fact that on each presentation, five labels were offered as choices to the subjects.

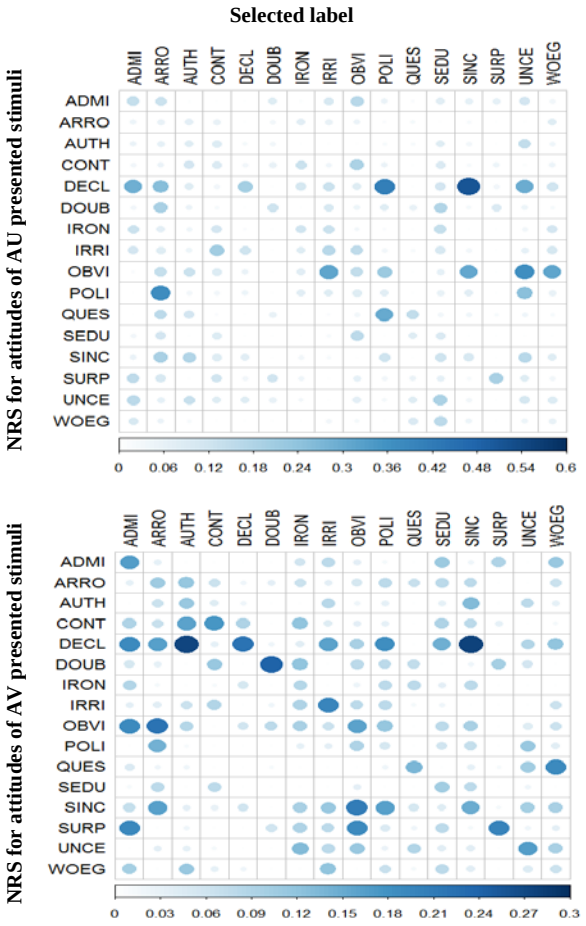

Figure 4: Distributions of normalized recognition scores (NRS): audio-visual (bottom), audio-only (top).

Figure 4 shows the distribution of the calculated NRS of listeners according to the presented labels. The columns of the graph are the presented labels and the rows are the presented stimuli in the different modalities (AU/AV). The size of circles together with the brightness of color show the calculated value. The higher the NRS the darker and larger the circle.

Cluster 1: N=31	**Label: IRRI, N_{AU}=5, N_{AV}=6, reco: 72%**			
IRRI,N=11	AUTH,N=5	CONT,N=3	WOEG,N=3	OBVI,N=2
2.511	1.227	1.761	1.682	2.109
Cluster 2: N=22	**Label: ARRO, N_{AU}=1, N_{AV}=2, reco: 27%**			
CONT,N=8	AUTH,N=4	ARRO,N=3	DECL,N=2	IRON,N=2
2.276	1.725	1.709	1.804	2.054
Cluster 3: N=85	**Label: DECL, N_{AU}=7, N_{AV}=8, reco: 67%**			
DECL,N=15	SINC,N=13	POLI,N=12	WOEG,N=7	ARRO,N=6
1.975	1.595	1.653	1.409	1.454
Cluster 4: N=54	**Label: ---**			
QUES,N=9	UNCE,N=7	WOEG,N=6	SEDU,N=4	AUTH,N=4
1.835	1.963	1.662	1.452	1.624
Cluster 5: N=23	**Label: IRON, N_{AU}=5, N_{AV}=5, reco: 48%**			
IRON,N=10	ADMI,N=2	IRRI,N=2	SEDU,N=2	QUES,N=1
1.988	2.025	1.841	2.259	2.333
Cluster 6: N=18	**Label: SEDU, N_{AU}=5, N_{AV}=4, reco: 70%**			
SEDU,N=6	OBVI,N=3	ADMI,N=2	ARRO,N=2	CONT,N=2
4.141	2.584	2.025	2.357	2.592
Cluster 7: N=15	**Label: QUES, N_{AU}=5, N_{AV}=4, reco: 59%**			
QUES,N=9	WOEG,N=2	UNCE,N=1	AUTH,N=1	POLI,N=1
2.686	2.228	1.711	1.990	2.184
Cluster 8: N=20	**Label: ADMI, N_{AU}=4, N_{AV}=4, reco: 48%**			
ADMI,N=8	SURP,N=3	ARRO,N=2	SEDU,N=2	DOUB,N=1
2.477	1.887	1.560	3.606	2.313
Cluster 9: N=52	**Label: SURP, N_{AU}=13, N_{AV}=19, reco: 62%**			
SURP,N=17	DOUB,N=17	OBVI,N=4	CONT,N=3	POLI,N=2
2.722	2.359	1.734	1.823	1.624

Table 2: Overview of the nine clusters, total number of stimuli, associated label, numbers of audio-only and audio-visual stimuli, percentage of recognition score, first five stimuli with highest number, average distance to other clusters.

Table 2 lists the nine clusters with the total number of stimuli and number of the first five attitudes contained in each cluster. Additionally the average distance to the centers of other clusters for each of the attitudes contained in the cluster is shown. These distances provide information about the dissimilarity of the attitude with respect to all others. In order to quantify the differences between the information received via the audio-only modality and the audio-visual one, we compared the shapes of the clouds of points created by the positions of the 16 stimuli of each speaker, in each modality, in the multi-dimensional space of the correspondence analysis. This approach is based on the work of [13] who propose to compare the dispersion of clouds of words according to their dispersion in the n-dimensional space obtained at the output of a correspondence analysis [14]. It involves measuring the Euclidean distance between each pair of stimuli produced by the same speaker, in one modality, and stacking these distances in a vector. Each vector thus represents the distribution of the stimuli in the space obtained from the perception results – thus representing the "perceptual space" in which stimuli are spread. The study of correlations between the vectors of the speakers in different modalities allows measuring the similarities and differences between speakers, or between modalities. Applying this measure to the stimuli spread onto the correspondence analysis space, one can observe that 39% of the observed variance is common to all the stimuli, whatever their modality of presentation, and constitute what Romney coined the knowledge (in our case the knowledge of expressive behavior) shared, whatever the modality or the speaker. The part of the observed variance in the perception that can be attributed to modality amounts to 11%, while the part of variance that is specific to individual speakers is 30%. The remaining 20% are to be linked to chance, error and noise in the experiment.

4.3 Performance Scores and Recognition Rates

Table 3 shows the average performance scores (audio-visual/audio-only) for the 16 attitudes, measured on a scale from 1 to 9, for each attitude yielded in our previous study [9].

	AV	**AU**			**AV**	**AU**
DOUB	8.143	7.556		SINC	6.581	6.375
IRRI	7.468	7.506		AUTH	6.483	6.400
SURP	7.450	-		ADMI	6.433	6.213
DECL	7.428	-		UNCE	6.135	5.778
OBVI	6.960	-		IRON	6.030	5.400
QUES	6.862	6.875		POLI	5.905	5.420
CONT	6.785	5.996		SEDU	5.858	4.250
ARRO	6.605	5.017		WOEG	5.669	5.944

Table 3: Mean performance scores for the AV and AU modalities obtained in [9]. AU stimuli for SURP, DECL and OBVI were not presented, hence results are unavailable.

Mean performance scores of attitudes (sorted by the audio-visual scores) are ordered in a similar fashion as the recognition rates (see Figure 1). Attitudes such as DOUB, IRRI, SURP and DECL received the highest audio-visual scores, and POLI, WOEG and SEDU the lowest ones. Differences between the two perception tests were found with respect to the audio-only stimuli. In contrast to the recognition task, audio-only stimuli only show marginally higher performance scores for IRRI and WOEG.

The NRS values for individual stimuli and their performance scores are strongly correlated (Pearson's $r=0.566$, $p < 0.01$). This indicates that well performed stimuli are also recognized more easily.

5. Discussion

The experiment shows the importance of visual information for the correct recognition of social expressions. As can be seen in Figure 4 (top), attitudes presented in audio-only condition were often misclassified. When stimuli of type CONT are presented they only reach an NRS of 0.092. Labels ARRO (NRS=0.533), AUTH (0.110) and OBVI (0.191) appear to be more plausible choices. CONT, ARRO, AUTH and OBVI are attitudes conveying mostly negative intentions and very similar in the form of their expression. Similar misclassifications occur when the subject is presented stimuli of type DOUB (NRS= 0.122). ARRO (0.193), AUTH (0.051) and OBVI (0.046) attract many votes.

These cases show the importance of visual cues providing helpful information. Cues such as slightly closed eyes with a blank or threatening gaze, and a wry mouth seem to indicate dislike towards the interlocutor, whereas strongly furrowed eyebrows and forward head movements convey skepticism regarding the message (see examples in Figure 3).

Figure 4 (bottom) also shows that DECL presented audio-visually was often misinterpreted by the subject. Confusion occurred with AUTH and SINC. In contrast, however, SINC and SURP were often confused with OBVI, and the latter often with ARRO. SINC and AUTH are attitudes which are more likely to lack emotional expressiveness therefore confusion with DECL seems plausible.

Out of the sixteen situations presented to the listeners, nine labels are used consistently to describe the expressive behaviors of the ten speakers. Interestingly (but predictably), the labels describing situations not conventionalized in German culture (e.g. the WOEG and SINC, typical of the Japanese culture [1]) are not used in a coherent manner by listeners, and did not form a cluster of their own. This is especially true for the WOEG expression, which receives a particularly low recognition score. This low score may be related to the inexistence of the concept in German culture, but also to the lack of a coherent expressive behavior in speakers.

Other labels such as IRON, SEDU and POLI also receive low recognition scores. The first two appear in the set of labels obtained from the clustering: they are thus concepts that are sufficiently clearly distinguished from the others for the subject to label a few stimuli as typical of the behavior they would associate with the concepts. These two labels form small clusters with few stimuli, and particularly not the most neutral stimuli – which are mostly regrouped either under the label DECL (cluster 3) or in the indiscriminate cluster 4.

The cluster labeled as SEDU contains six stimuli of that attitude. It also groups seven stimuli expressing an imposition (potentially negative: ARRO, CONT, OBVI), plus three positive expressions (ADMI and SINC). Such performances included in the SEDU cluster seem to convey an imposition pattern that is not found e.g. in US-English speakers (mixed with irony) [15]. The taboo nature of such an expression may explain partly the low recognition scores, as it is a demanding expression to perform for untrained speakers. About half the stimuli in the cluster labeled IRON express irony, while other stimuli did not show a particular pattern. This could be interpreted as if half the productions of irony (10 out of the 20 presented to listeners) where mostly recognized as irony, as they do not show a particular confusion pattern – which is a good result, considering the fact that irony may be expressed via a contrast between two aspects of the message [16] (here the context and the expression, as the lexical content is constrained). As the context is not given to the listeners, they should have resorted to a specific ironic marking in the stimuli, most possibly prosodic as audio-visual recognition scores are lower than the audio-only ones [17].

The case of POLI, a concept obviously existing in the German culture, but a label not used to depict a consistent behavior in this experiment, may be understood in the light of [18] work on "German politeness". She described politeness in German speakers as the norm of neutral behavior – thus as an unmarked way of speaking, moreover emphasizing the tendency of German speaker to prefer directness, as compared to English speakers, resorting more often to indirect speech acts as a polite behavior. There could be thus few marks of politeness produced by speakers, and fewer expectations of such marks on the part of the listeners. At least, most of the existing marks seem to be carried by the audio modality, as the recognition scores in audio-visual are even lower. Other labels that form a cluster with a small set of stimuli (about 20) are ARRO, QUES and ADMI.

The expression of question is one of the basic functions of prosody – it is thus not surprising to find a cluster with that label. It is the smallest cluster – the one with the smallest number of confusions (in terms of misclassified stimuli), together with the SURP/DOUB cluster. This shows the prosodic function performing interrogative speech acts consistently produced by speakers, and well retrieved by listeners.

The cluster labeled as IRRI contains mostly stimuli of irritation, plus five of authority. This fits the situation, where a speaker has to impose her/his will over the interlocutor. Expressions in this cluster clearly express a dominance trait, but lack the negative trait that was found in cluster 2 (ARRO). Note that at a higher level in the clustering, the two clusters mix together, and are mostly labeled IRRI. It may be that the dominance trait comes first in all these expressions.

Finally, the cluster grouping most stimuli (85) was labeled as DECL. It contains most of the neutral declarative stimuli, but also many others (this neutral expression was certainly used as a default answer) – in decreasing order of frequency: SINC, POLI, UNCE, WOEG, ARRO, AUTH, OBVI, IRON, IRRI, ADMI, SEDU and CONT. This list contains only expressions of assertions. Thus, no interrogative expressions were mixed with declarative ones: this reinforces the importance of the main modal distinction between assertion and question classically conveyed by prosody. The four expressions mostly mixed with DECL include the three types of politeness defined in this corpus (POLI, SINC, WOEG) – thus reinforcing House's description of German politeness focusing on directness [18]. The fourth expression is uncertainty, that receives rather high recognition scores, but does not form a cluster of its own. Whether there is a specific prosodic performance for uncertainty remains an open question.

Cluster 4 that regroups expressions without a particular association to one label is most certainly a group of the poorest performances.

6. Conclusion

We presented at study on the recognition of auditory-visual attitudes. As was shown modality explains about 10% of the changes observed in the distribution of attitudes in their perceptual space. Thus multimodal presentation does not change the main dimensions of expressivity; it is mostly used to adapt the fine-grained associations between concepts and the prosodic performances.

The comparison of results of this study with results of our previous perception experiment shows similar effects. Attitudes presented audio-visually which were recognized well (hence yielded high recognition scores) were also perceived as well-performed. It must be stated that the use of predefined "associated attitudes" in the labels offered to the subjects introduced a strong bias. Future work will involve

refined paradigms for evaluating the perception of attitudes letting subjects choose their own terminology.

7. Acknowledgements

This work was funded through a French *digiteo* grant for Mixdorff for a research stay at LIMSI. Recordings of attitudes were funded through German DLR research grant no. 01DN13007 ARG.

8. References

[1] Shochi. T.. Rilliard. A.. Aubergé. V. & Erickson. D.. "Intercultural perception of English. French and Japanese social affective prosody". in S. Hancil (ed.). The Role of Prosody in Affective Speech. Linguistic Insights 97. Bern: Peter Lang. AG. Bern. 31-59. 2009.

[2] Rilliard, A., Erickson, D., Moraes, J. A. & Shochi, T. "Cross-Cultural Perception of some Japanese Politeness and Impoliteness Expressions". In Baider, F. & Cislaru, G. (Eds.), Linguistic approaches to emotion in context. Amsterdam: John Benjamins, 251-276. 2014.

[3] Swerts, M. and Krahmer, E., "Audiovisual prosody and feeling of knowing", Journal of Memory and Language 53(1): 81-94, 2005.

[4] Nadeu, M. & Prieto, P. "Pitch range, gestural information, and perceived politeness in Catalan". Journal of Pragmatics, 43(3): 841–854, 2011.

[5] Grichkovtsova. I.. Morel. M.. & Lacheret. A.. "The role of voice quality and prosodic contour in affective speech perception", Speech Communication, 54(3):414-429, 2012.

[6] Gu, W., Zhang, T. & Fujisaki. H., "Prosodic analysis and perception of Mandarin utterances conveying attitudes", Proceedings of Interspeech, Firenze, Italy. 1069-1072, 2011.

[7] Wierzbicka. A.. "Defining emotion concepts", Cognitive Science 16: 539–581, 1992.

[8] Rilliard, A., Erickson, D., Shochi, T., de Moraes, J.A., "Social face to face communication - American English attitudinal prosody", INTERSPEECH 2013. 1648-1652.

[9] Hönemann, A., Mixdorff, H., Rilliard, A. "Social attitudes - recordings and evaluation of an audio-visual corpus in German", Forum Acusticum 2014, Krakow, Poland.

[10] de Moraes, J. A. "The pitch accents in Brazilian Portuguese: analysis by synthesis". Proc. of Speech Prosody 2008, Campinas, Brazil, 389-397, 2008.

[11] Spencer-Oatey, H. "Reconsidering power and distance". Journal of Pragmatics, 26:1–24, 1996.

[12] Crawley, M.J., 2013. The R Book, Second ed. John Wiley & Sons, West Sussex, United Kingdom.

[13] Romney, A.K., Moore, C.C., Batchelder, W.H. & Hsia, T.-L. "Statistical methods for characterizing similarities and differences between semantic structures". Proceedings of the National Academy of Sciences 97(1): 518-523, 2000.

[14] Lê, S., Josse, J. & Husson, F. "FactoMineR: An R Package for Multivariate Analysis". Journal of Statistical Software. 25(1): 1-18, 2008.

[15] Rilliard, A., Erickson, D., de Moraes, J.A. & Shochi, T. "Perception of Expressive Prosodic Speech Acts Performed in USA English by L1 and L2 Speakers". Journal of Speech Sciences, (submitted).

[16] Bryant G.A. "Verbal irony in the wild". Pragmatics & Cognition, 19(2):291-309, 2011.

[17] Niebuhr, O. "A little more ironic"–Voice quality and segmental reduction differences between sarcastic and neutral utterances. In proceedings Speech Prosody, Dublin, 608-612, 2014.

[18] House, J. "Politeness in Germany: politeness in Germany?". In Hickey, L., & Stewart, M. (Eds.). Politeness in Europe. Multilingual Matters (Vol. 127), 13-28, 2005.