

<https://doi.org/10.1016/j.intcom.2006.04.003>

Access to this work was provided by the University of Maryland, Baltimore County (UMBC) ScholarWorks@UMBC digital repository on the Maryland Shared Open Access (MD-SOAR) platform.

Please provide feedback

Please support the ScholarWorks@UMBC repository by emailing scholarworks-group@umbc.edu and telling us what having access to this work means to you and why it's important to you. Thank you.

Cross-Cultural Differences in Recognizing Affect from Body Posture

Andrea Kleinsmith^a, P. Ravindra De Silva^a,
Nadia Bianchi-Berthouze^{a,b,*}

^a*Database Systems Laboratory, University of Aizu, Aizu Wakamatsu 965-8580, Japan*

^b*UCL Interaction Centre, University College London, London, WC1E 7DP, UK*

Abstract

Conveyance and recognition of human emotion and affective expression is influenced by many factors, including culture. Within the user modeling field, it has become increasingly necessary to understand the role affect can play in personalizing interactive interfaces using embodied animated agents. However, little research within the computer science field aims at understanding cultural differences within this vein. Therefore, we conducted a study to evaluate if differences exist in the way various cultures perceive emotion from body posture. We used static posture images of affectively expressive avatars to conduct recognition experiments with subjects from three cultures. After analyzing the subjects' judgments using multivariate analysis, we grounded the identified differences into a set of low-level posture features. We then used Mixture Discriminant Analysis (MDA) and an unsupervised Expectation Maximization (EM) model to build separate cultural models for affective posture recognition. Our results could prove useful to aide designers in creating more effective affective avatars.

Key words: Affective communication, affective body postures, embodied animated agents, intercultural differences, emotion nuances

1 Introduction

As Picard (1998) points out, the manner in which humans convey emotion or affective messages in general is affected by many factors, such as age, gender, posture, culture, context, and so on. According to Mehrabian and Friar

* Corresponding author: Tel.: +44 (0) 20 7679 5216; fax: +44 (0) 20 7679 5295
Email address: n.berthouze@ucl.ac.uk (Nadia Bianchi-Berthouze).

(1969), changes in a person's affective state (used as a general term for discussing mood, emotion, and feeling) are reflected by changes in body posture. Understanding these factors has become increasingly important within many areas of computer science. More than ever, it is necessary to create interactive interfaces that are capable of communicating with the user through an affective channel. While we acknowledge the importance of other modalities, such as face and voice, research in these fields is quite extensive and continues to be a major focus. However, whole body postures are also shown to be quite important for conveying emotion (indeed, 55% of nonverbal communication is said to be expressed through body language (Mehrabian and Friar, 1969)), and remains a novel area of research within the area of computer science.

The role of posture in affect recognition, and the importance of emotion in the development and support of intelligent and social behavior has been accepted and researched within several fields including psychology, neurology, and biology. According to several researchers, central nervous system structures are responsible for the perception of emotion, and furthermore, its placement in these structures may be specific to both emotion and modality (Keltner et al., 2003), (Scherer et al., 2003).

There is also evidence to support that the way in which emotions are expressed and controlled (Mehrabian and Friar, 1969), as well as the interpretation of emotion (Keltner et al., 2003) is clearly shaped by culture. In our work we consider Matsumoto's definition of culture which defines it as "a shared system of socially transmitted behavior that describes, defines, and guides people's ways of life" (Matsumoto, 2005). Many researchers have used cross-cultural emotion recognition studies to validate evidence in favor of emotion universality (Elfenbein et al., 2002). For some emotions, there is cross-cultural support for the universality of many modes of nonverbal behavior, including face, voice, and body expressions, as well as changes in a person's physiology (Mesquita, 2003). However, the majority of the research on emotion universality has concentrated on the recognition of facial expressions using still photographs (Ekman, 1994), (Russell, 1994), (Elfenbein et al., 2002). Few researchers are examining the cross-cultural differences of emotion recognition in whole body posture (Kudoh and Matsumoto, 1985). In the realm of facial expression, much research has been conducted comparing emotion differences between Japan and the United States. In a study by Friesen (1972), cross-cultural differences were compared in the facial expressions of Japanese and American subjects while viewing both neutral films, and films intended to cause stress. As a support to the universality of facial expressions, it was shown that both groups expressed almost exactly the same facial expressions when watching the films alone. In a study by Scherer et al. (1988), it was reported that the Japanese use fewer hand, arm, and whole body gestures than Americans when in emotional situations.

Why do we need models of culture for embodied agents? These models are necessary to develop animated agents capable of affective communication on a more individual, personalized basis. By understanding the differences between cultures, avatars can be made to relate more to individual cultures by individualizing the models. For example, an avatar that is very expressive may convey emotions that are considered to be too intense for Japanese observers and just right for U.S. observers. By starting with a basic model, adaptation can help to personalize each model according to different variables, such as culture, without having to manually create a new model for every culture.

While researchers do acknowledge the importance of adding cultural dimensions to embodied agents (Lamolle et al., 2005), (Maya et al., 2004), (Krenn et al., 2004), few have actually implemented them. In order to accurately represent the emotional state of the user, many of her aspects must be identified, including culture. As many cultures differ in their ways to express and perceive emotion, these aspects also must be identified in order to design and implement avatars that express the appropriate emotional reactions and intensity levels of those reactions (termed cultural display rules by Ekman et al. (1972)) in a more natural and plausible way to which users can relate.

Such culture models could be integrated into avatar animation software such as Demeanour (Ballin et al., 2004) and EMOTE (Chi et al., 2000). The Demeanour software maps from high-level emotional states to low-level animation of expressive body language. In a similar vein, EMOTE is an animation system for 3D characters that uses the Effort and Shape components of Laban Movement Analysis (von Laban, 1988) to improve character body movement.

Believability is an important factor for avatar design. Becker et al. (2005) have created an empathic embodied agent that expresses affect while playing a card game with a human user. They hypothesize that the believability of the character can be increased through the addition of negative emotion expression by the character. Other systems try to increase believability by reacting to the affective state of the user. Prendinger et al. (2004) have created an animated interface agent that escorts the user during a virtual job interview. The agent obtains affective information from the user and displays empathic behavior accordingly.

We believe that believability can be further increased by the addition of cultural models. Embodied museum agents are gaining much attention, and several such systems currently exist (Lim et al., 2005), (Kopp et al., 2005). Museums are a particularly appropriate arena in which to have an agent that embodies culture, as tourists from many different cultures are often the visitors in these settings. The importance of culture models in eLearning systems has been evidenced by research in the UK by Dunn and Marinetti (2002) that has found high dropout rates for eLearning due to “*culturally insensitive con-*

tent” as described by user feedback. As these systems replace humans with avatar agents, it is important that their nonverbal behaviors are as natural and culturally sensitive as possible in order not to make the user/student uncomfortable. Even human teachers are taught what kind of body language conveys the appropriate messages for specific situations (Neill and Caswell, 1993).

We conducted this study to evaluate the possible cultural differences between Japan, Sri Lanka, and the U.S. in the perception of emotion through whole body postures. If, in fact, differences do exist, our aim is twofold. One, we will attempt to qualify these differences through an examination of how the cultures vary in recognizing emotion and its intensity. Two, we will ground these differences into a set of low-level posture features by identifying which features are used by all three cultures, and which features play a different role in each of the three models. In our studies, we use images of a non-culturally specific 3D avatar created from original motion capture data instead of human photos as an attempt to eliminate bias. By not using human photos of emotion, culture, gender and context elements are removed, and so many of the social and cultural constraints that exist in human-human communication are eliminated. Without these constraints, we assert that people do not feel as if they have to relate to the avatar, and instead they can focus more on the important aspect of the avatar, which is the message or information being conveyed.

Many considerations went into choosing Japanese, Sri Lankans, and Caucasian Americans as the three cultures to study. One, we chose two Asian cultures (Japanese and Sri Lankan), as the way each of these cultures interacts socially, on a nonverbal level, is quite different. Sri Lankans are considered to have a more Latin way of interacting (Lewis, 1999). Two, each of the three cultures seems to be representative of a larger group of cultures that are distinct, yet also share some similarities according to cultural dimensions (Hofstede, 2003).

The remainder of this paper is organized as follows. Section 2 gives a description of the posture data collection method and our tri-cultural (Japanese, Sri Lankan, and American) posture recognition experiment. In Section 3 we examine the cultural differences, and how the different cultures rated emotion and intensity. Section 4 describes how we grounded these differences into a set of low-level postural features. Our conclusions are presented in Section 5.

2 Method

2.1 Posture Data Collection

As a first step, we used a motion capture system to collect 3D affective postures of 13 human subjects (called actors hereafter). As the study took place in Japan, the majority (11) of the actors were Japanese, one was from Sri Lanka and one was from the U.S. In this context we define a posture as any *stance involving the hands and/or body that can convey emotions or feelings*.

Directions were given in English to the two non-Japanese actors and in Japanese and English to the Japanese actors. English was used for the Sri Lankan actor (and for the observers, discussed below in describing the posture evaluation experiment) because English is the predominant language used in Sri Lanka’s educational system. Dressed in the same suit with 32 markers attached to various joints and body segments, each actor was asked to perform an in-place posture expressing *anger*, *fear*, *happiness*, and *sadness*. The actors were directed to perform the emotion postures in their own way, as no constraints were placed on them. As a starting point, these emotions were chosen on the basis that they are included in the set of universal emotions defined by Ekman and Friesen (1975).

Each affective posture was captured by eight cameras and represented by contiguous frames describing the position of the 32 markers in the 3D space. In total, we captured 108 affective postures. We then used the original motion capture data to build affectively expressive avatars, shown in Figure 1, by selecting the frame (i.e., a static posture) that the actor evaluated as being the most expressive instant of the posture.

We used computer avatars instead of human photos in order to create a non-gender, non-culturally specific “humanoid” in an attempt to eliminate bias. Each avatar always has the same body. Furthermore, by using the avatar, the subjects are not affected by facial expressions. The reason for removing the face is to understand how posture alone is able to convey the desired impression.

Use of the face could have confounded the observers’ evaluations because we would not be able to discern which information was used in judging the expression. Bodily information and facial expressions have different roles, and the two may affect people differently. In fact, a neuroscience study by Meeren et al. (2005) shows that bodily information does bias the perception of the affective state conveyed by the face, particularly when the posture information and the facial information are incongruent.



Fig. 1. Examples of the 3D affectively expressive avatars for each emotion category. (a) Angry (b) Fear (c) Happy (d) Sad.

2.2 Posture Evaluation Experiment

We used an eight word forced-choice experimental design to evaluate potential cultural differences in emotion perception through whole body posture. This method was purposely chosen as it has been widely used in studies aimed at assessing cross-cultural agreements in the expression of emotion (Keltner and Ekman, 2003).

The participants (called observers hereafter) consisted of 25 Japanese, 25 Sri Lankans, and 20 Caucasian Americans. The American participants were born and raised in the U.S. with a heritage dating back at least two generations. Ancestry of the Americans was non-Asian and non-Hispanic in all but one case (the mother was born in Mexico but has been living in the U.S. for more than 35 years). The Japanese and Sri Lankan observers were all within a similar age range and came from a similar educational background, educational level, and career. The age range amongst the Americans was much greater, as was the career status, while the educational level was similar to that of the other cultures.

We used 108 affective posture images as the data set (refer to Figures 1, 4, and 5 for examples). The experiment was conducted online as a series of webpages. Experiment instructions were given in English for the non-Japanese observers, and in both English and Japanese for the Japanese observers. Again, as discussed above, English was used for the Sri Lankan observers as well as the Americans because it is the main language used in their educational system. For everyone, the main webpage was presented in both English and Japanese. The postures were presented in a randomized order, differing for each participant. For each page (one posture per page), subjects were asked to (1) rate the intensity of the emotion, defined by a value between 1 (lowest) and

5 (highest) to indicate how emotional is the posture, and (2) choose an emotion label to represent the posture displayed based on an eight word list comprised of pairs of labels indicating two nuances of the same emotion translated in both English and Japanese: anger (*angry, upset*), (*okotta, nayanda*), fear (*fearful, surprised*), (*osore, odoroki*), happiness (*happy, joy*), (*tanoshii, yorokobi*), and sadness (*sad, depressed*), (*kanashii, yuutsuna*). These words were presented on each evaluation page in both English and Japanese romaji (a system for transliterating Japanese words into the roman alphabet).

A survey was conducted in both English and Japanese to establish that, indeed, each pair of adjectives represented two levels of intensity and that it was the same for both languages. The results verified that the two levels of intensity for *fear/surprise* and *angry/upset* represent different levels of valence, and the two levels of intensity for *happy/joy* and *sad/depressed* represent burst/duration nuances. To further rule out differences due to semantic issues, when differences do occur, a two word forced-choice experiment was conducted to validate those differences. First we will report on the results of the eight word forced-choice experiment, and the two word forced-choice experiment is presented at the end of Section 3.

2.3 Concordance Between Actors and Observers Across Three Cultures

The results for each of the three cultures are represented in Table 1. The rows indicate the frequency of use by the subject of a culture for each emotion label to classify the set of postures corresponding to the emotion intended by the actors. Overall, a high correlation is observed between the observer-selected pairs of labels (i.e., *happy/joy*) and the actors' intended emotions, with the *sad/depressed* categories showing the overall highest agreement for all three cultures with an average of 63.4%. In fact, the emotion lexicons for depression-type words were similar in these three cultures as described in a study by Brandt and Boucher (1986). The emotion category with the lowest agreement was different for each culture. We hypothesize that this may be due to each culture placing greater importance on different features for distinguishing between postures within those emotion categories.

We then applied Cohen's Kappa coefficients (Cohen, 1960) to Table 1 considering each culture separately. Cohen's Kappa coefficients is a statistical technique used to measure the strength of agreement between raters, actors and observers in our case. The Cohen's Kappa values obtained for each of the three cultures are 0.42 for the Japanese, 0.40 for the Sri Lankans, and 0.43 for the Americans. According to the interpretation of the Cohen's Kappa coefficients obtained by Altman (1991), there is a moderate agreement between actors and observers. To further evaluate the agreement between the three

Table 1

The observers' ratings from the three cultures for each of the eight emotions are shown here. This table depicts the frequency of use (percentage) of the four emotions categories (separated out as two intensities of the same emotion). (sur. = surprised and dep. = depressed)

Actors	Japanese Observers				Sri Lankan Observers				American Observers			
Emotion Labels	Angry (angry upset)	Fear (fear sur.)	Happy (happy joy)	Sad (dep. sad)	Angry (angry upset)	Fear (fear sur.)	Happy (happy joy)	Sad (dep. sad)	Angry (angry upset)	Fear (fear sur.)	Happy (happy joy)	Sad (dep. sad)
Angry	41.7 14.4	4.4 9.2	10.4 7.9	7.3 4.4	22.9 25	5.5 10.7	10.8 7.7	8.5 7.1	22.2 22.3	4.9 9.2	19.2 8.4	3.9 4.3
Fear	15.5 4.4	26.1 28.7	9.4 10.9	2.3 2.6	18.7 10	30.4 13.4	8.8 9	4.5 3.3	17.3 1.2	23.2 26.3	10.7 5.5	1 1.8
Happy	13.6 5.7	1.4 20.7	22 32.4	1 3.2	8.2 7.9	4.5 11.9	29.3 30.2	4.4 2.3	11.6 10.2	1.8 15.3	29 26.2	1.5 1.7
Sad	5.4 23.9	2.8 3.7	2.2 1.4	34.6 25.6	5.4 22.6	1.8 1.5	1 1	29.5 35.4	6.6 14.1	2.2 4.3	2.6 1.8	35.8 29.2

cultures, each posture was associated with the most frequent emotion category for each culture. The most frequent label was the same across the three cultures for 71 of the 108 postures (i.e., 66% agreement). Again we estimated Cohen's Kappa coefficients for agreements between actors and observers (according to the emotion label). While in this case agreement levels increased for each culture (0.54 for the Japanese, 0.56 for the Sri Lankans, and 0.54 for the Americans), they still indicate that agreement between the groups, actors and observers, remains at a moderate level.

Table 2

Testing the differences in correct recognition. JA=Japanese, SL=Sri Lankans, US=Americans.

Emotion	Univariate Test	Post hoc p-value	Cultural diffs
Angry	Significant differences p-value=0.037	JA-SL=0.335 JA-US=0.029 SL-US=0.784	JA > US, p-value=0.041
Fear	Significant differences p-value=0.002	JA-SL=0.001 JA-US=0.132 SL-US=0.043	JA > SL, p-value=0.02 SL > US, p-value=0.047
Happy	No significant diff. p-value=0.510	No significant diff.	No significant diff.
Sad	No significant diff. p-value=0.284	No significant diff.	No significant diff.

Next, we investigated both the differences in the rate of correct recognition¹ between the cultures within each emotion category and the differences in correct recognition for each posture, individually. Thus, the postures were divided into four categories according to the actors' labels. There are 23 postures for angry, 40 postures for fear, 20 postures for happy, and 25 postures for sad.

¹ For the remainder of the paper, the statements "correct recognition" or "correct classification" refer to postures being classified according to the emotion label defined by the actors

For each emotion (angry, fear, happy, and sad), according to the actors' labels, we chose the observers' frequency of use for that label, even if it was not the label most frequently used by the observers'. The repeated measurements method was applied to each emotion category separately. We used a post hoc test with Bonferroni corrections for discovering which differences occurred between which cultures. A paired t-test was used for testing recognition rates of each posture. Table 2 shows the statistical results of the differences in correct recognition between each of the three cultures. When we evaluated all of the emotion categories, happy and sad showed no significant differences in correct recognition by the observers belonging to the three different cultures. In fact, the p-values for these two categories are well above a significant level of 0.05 (see column 1 in Table 2).

The post hoc test results showed that Japanese observers performed slightly better than the other cultures in recognizing fear and anger. We hypothesize that this could be due to the fact that the majority of the actors were Japanese. Our results could lend evidence to support the idea of emotional "dialects" as described by Elfenbein and Ambady (2002). They consider the idea that emotional expression is a universal language, and that different dialects of that universal language exist across cultures.

Another interesting point is that according to our data, the Japanese and Sri Lankan observers recognized negative emotions as well as, or even better than the U.S. observers. This result contradicts the findings of (Matsumoto, 1989, 1992). As an absolutist (i.e. he believes in the existence of unchangeable characteristics for either the groups expressing emotion or the groups perceiving emotion), Matsumoto argues that as a collectivistic group (meaning that individuals within this type of culture typically have stronger ties to the rest of the community as a whole (Hofstede, 2003)), the Japanese are less likely to perceive and express negative emotions, in order to preserve the social order.

Also contradictory to Matsumoto's findings is a study by Elfenbein et al. (2002) in which they could not identify any tendencies by Japanese, American, and Indian observers of facial expressions to more often select negative emotions. Their results found that instead, Americans attributed positive emotions more to Japanese facial expression photos, possibly due to a preconceived idea that the Americans may have about the Japanese as a culture. Elfenbein's results suggest that a relational view of cross-cultural differences in emotion communication is more likely. This view focuses on the relationship between the expresser's culture and the perceiver's culture, indicating that the relationship affects the observers' judgments. As we use a non-human, non-cultural avatar in our study, we believe that cultural issues and social constraints are, at the very least, lessened, if not removed completely. As a future study, it would be interesting to examine how the idea of the avatar possessing its own kind of culture also affects an observer's interpretation of emotion

expression.

To look more closely at specific classification differences, by evaluating the postures (according to emotion) for each culture, one interesting finding was seen in how the cultures seem to recognize angry postures. Japanese observers tended to show higher frequencies for angry postures in which the hands are placed on the hips, while Sri Lankan observers showed higher frequencies for postures with a more animated body. This may help to explain Scherer et al. (1988)’s study discussed in the introduction which reported that in general, the Japanese tend to use more subdued gestures in emotional situations.

Further results showed that the Japanese tended to assign the surprised label to postures more often than the fear label. These results are similar to the previously mentioned cross-cultural study by Elfenbein et al. (2002). Their results showed that the Japanese tended not to assign “afraid” to expressions, and instead more often assigned “surprised”. One explanation for this could be related to the cultural dimension of uncertainty avoidance (Hofstede, 2003). The idea here is that people brought up in cultures that rank high in this dimension will feel more uncomfortable in uncertain situations, and attempt to avoid them as much as possible. The consequence of this may be that these individuals will be less able to recognize these types of expressions, such as fear, since they have not been observed very often (Schimmack, 1996). Another explanation is that surprise, more than fear, seems to be an expression displayed by the Japanese in very uncomfortable situations with a wide variety of postures.

3 Intensity Ratings and Cultural Differences

3.1 Method

The second aim was to assess the cultural differences in the evaluation of emotional intensity on the 71 common postures. Toward this goal, we analyzed the data using two approaches. First, we aimed to assess the intensity according to the intensity value (i.e., a number from 1(lowest) to 5(highest)) associated by the observers of the three cultures to each posture during the posture evaluation experiment presented in Section 2. Second, we attempted to assess how each culture used the pair of labels within each emotion category. For both of these examinations we used the repeated measurement method (Field, 2000) and either a multivariate test or a univariate test² to identify the existence

² We used Mauchly’s test for checking the assumption of sphericity. The *sphericity* assumption is an assumption about the structure of the covariance matrix in a re-

of statistically significant differences between the cultures. This method was applied to each emotion case separately. There are 21 postures for *angry*, 19 for *fear*, 15 for *happy*, and 16 for *sad*. Subsequently, to better qualify which differences occurred between which cultures, for each of the emotion categories, we used a post hoc test with the Bonferroni correction.

3.2 Discussion

Table 3 shows the results of the analysis of the intensity rating. Overall, we can see that only in the case of the fear category is there no significant difference between the intensity ratings of the observers belonging to the three different cultures. In fact, its p-value is well above a significance level of 0.05. Our results are similar to brain studies in which the fear emotion seems to be triggered more at the sensorial level than at the cognitive one (de Gelder et al., 2003). Other research by Kleinsmith et al. (2005) indicates that motion features may be necessary in order to increase recognition rates for body expressions of fear. Table 4 shows the results of the analysis on label usage. The overall results point out that in the case of angry and fear postures, there is no significant difference in the use of the pair of labels between the three cultures.

Table 3

Testing the differences in intensity ratings. JA=Japanese, SL=Sri Lankans, US=Americans.

Emotion	Multivariate test	Post hoc p-value	Cultural diffs
Angry	Significant differences p-value=0.032	JA-SL=0.896 JA-US=0.756 SL-US=0.032	US > SL p-value=0.043
Fear	No significant diff. p-value=0.176	No significant diff.	No significant diff.
Happy	Significant differences p-value=0.021	JA-SL=0.021 JA-US=0.089 SL-US=0.742	JA > SL p-value=0.034
Sad	Significant differences p-value=0.041	JA-SL=0.022 JA-US=0.976 SL-US=0.0876	JA > SL p-value=0.012

For both analyses, when significant differences were detected by the multivariate test (p-value below 0.05), a post hoc test was conducted to better qualify these differences. The results of this test are represented in the second column of each table. This column indicates in which pairs of cultures a significant

peated measures design. Following Field (2000) (pp. 337-338), we used a univariate test (Greenhouse-Geisser) to treat cases in which the sphericity was statistically satisfied, while we used a multivariate test (Wilks' Lambda) in the cases of significant violation of the sphericity assumption.

Table 4

Testing the differences in the use of the eight emotion labels.

Emotion	Univariate test	Post hoc p-value	Cultural diffs
Angry	No significant diff. p-value=0.432	No significant differences	No significant differences
Fear	No Significant diff. p-value=0.078	No significant diff. differences	No significant differences
Happy	Significant diff. p-value=0.022	JA-SL=0.032 JA-US=0.073 SL-US=0.987	JA used the joy label more than SL
Sad	Significant diff. p-value=0.033	JA-SL=0.001 JA-US=0.323 SL-US=0.625	JA used the depressed label more than SL

difference exists (p-value below 0.05). For these pairs, we applied a post hoc test to assess if one culture used a higher intensity rating than the other. The results, shown in the last column of Table 3, point out that the Japanese are more likely than the Sri Lankans to associate a higher intensity to both happy and sad postures. In the case of angry, it is the Americans who are more likely than the Sri Lankans to assign a higher intensity to these postures. Refer to Figure 2 for examples of observer-defined higher intensity postures.

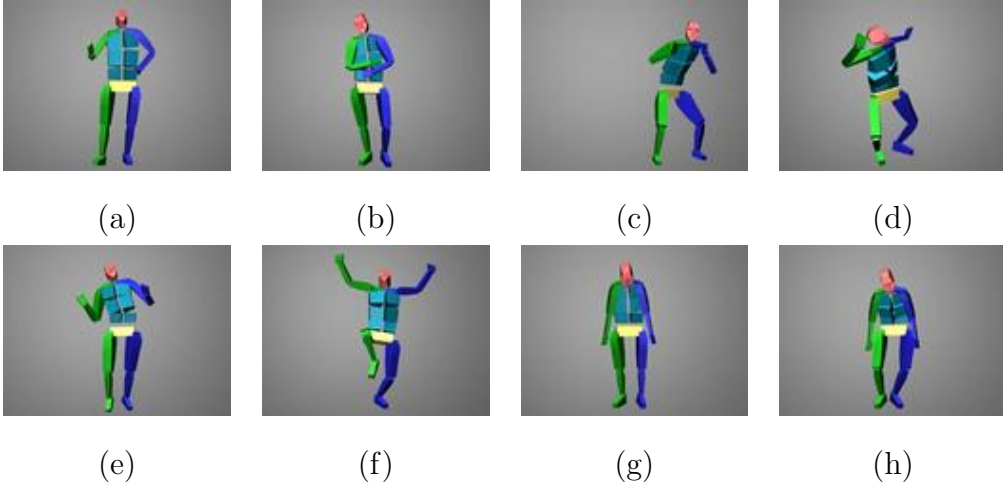


Fig. 2. These figures show the avatars for which the three cultures associated higher intensities. (a) Angry (b) Upset (c) Fear (d) Surprise (e) Happy (f) Joy (g) Sad (h) Depressed

We then compared these culture pairs in their use of the emotion labels. The results, shown in the last column of Table 4, show that the Japanese tended to use the emotion terms joy and depressed more often than the Sri Lankans.

To rule out the possibility that these previous differences were due to semantic issues, we conducted a survey to examine possible semantic differences between the emotion words of the two cultures. With the help of a linguist, the Japanese emotion words for *happy/joy* and *sad/depressed* were translated into Sinhalese

(a language of Sri Lanka), and a set of sentences was defined to evaluate if there are differences in how the two cultures use the two pairs of labels. (Please refer to Table 5 for the Japanese and Sri Lankan translations and sentence examples). There were six sentences for the *happy/joy* labels, and five sentences for the *sad/depressed* labels. 15 Japanese, 12 Sri Lankans, and 12 Americans participated. Subjects were asked to fill in the blank with their idea of the appropriate emotion word (from the pairs of labels) depicted by the sentence.

Table 5

Sentence examples used in the survey to understand the semantic meaning within the Japanese and the Sri Lankan cultures.

(English)	Japanese	Sri Lankan	Sentence examples
(Happy)	Tanoshii	Sathutu	<i>I'm ----- to be here with my family.</i>
(Joy)	Yorokobi	Pritimat	<i>She looks so ----- for receiving the scholarship.</i>
(Sad)	Kanashii	Duak wenawa	<i>She was so ----- when he heard the bad news.</i>
(Depressed)	Yuutsuna	Dukata pat wenawa	<i>She was so ----- because she cannot find a job.</i>

The results of the survey show that in all cases except for two of the *happy/joy* sentences, for both the Japanese and the Sri Lankans there is quite a clear and similar pattern for *happy/joy* and *sad/depressed* word usage (i.e., for what types of situations), meaning that the emotion words for each of the two cultures show strong and similar semantics. However, the pattern was not so clear for the Americans. In particular, most of them never used the label joy. In fact, in English these terms appear to have lost their distinction (called semantic weakening) according to Wierzbicka (1999). For example, the term happy is now used with a very broad range of meanings, including joy, while in the past joy and happy were used in different cases.

Because of the semantic similarity observed between the Japanese and Sri Lankans, we repeated the avatar evaluation experiment in a two word forced-choice design using both Japanese and Sri Lankan languages. Subjects from the two cultures (20 Japanese and 20 Sri Lankans) viewed single posture webpages for which they were asked to choose one of two labels, *happy/joy* in one session, and *sad/depressed* in another session. To test for reliability, each posture was shown three times. We then applied the repeated measurement method to each emotion category and each culture separately. The results, shown in Table 6, indicate that the Japanese use the labels within each of the word pairs differently. By this we mean that the Japanese appear to use the joy label more than the happy label (p-value of 0.037) and the depressed label more than the sad label (p-value of 0.041). Therefore, the overall results may imply that on average, while the two cultures tend to assign the same label to the same postures, the Japanese tend to assign a more intensive (burst of emotion) label than the Sri Lankans.

Table 6

Testing the differences in correct recognition according to the frequency values for happy and sad emotions.

Culture	Emotion	Univariate Test	Emotion label diffs
Japanese	Tanoshii (<i>Happy</i>) Yorokobi (<i>Joy</i>)	Significant differences p-value=0.025	Joy > Happy p-value=0.037
Sri Lankan	Sathutu (<i>Happy</i>) Pritimat (<i>Joy</i>)	No significant diff. p-value=0.523	No significant diff.
Japanese	Kanashii (<i>Sad</i>) Yuutsuna (<i>Depressed</i>)	Significant differences p-value=0.026	Depressed > Sad p-value=0.041
Sri Lankan	Duak wenawa (<i>Sad</i>) Dukata pat wenawa (<i>Depressed</i>)	No significant diff. p-value=0.931	No significant diff.

4 Grounding Cultural Differences in Posture Features

The final goal of the study was to ground the perceptual differences between the three cultures into a set of features describing the affective postures, and to determine if and how cultures consider nuances within the four main categories. Toward this objective, we used a set of 24 low-level postural features to create a numerical description of our 108 postures. As proposed in Bianchi-Berthouze and Kleinsmith (2003), direction and volume of the body were described by projecting each marker on the three orthogonal planes and measuring the lateral, frontal and vertical extension of the body. The postures were initially rotated to simulate a frontal view. Each computed feature (listed in Table 7) was normalized according to the body structure of the actor, i.e., according to the maximal extension of her/his body (shown in Figure 3). Based on this set of features, we applied the non-linear Mixture Discriminant Analysis (MDA) (Hastie and Tibshirabi, 1996) modeling technique to create separate cultural models for discriminating between affective postures, and an unsupervised clustering algorithm, Expectation Maximization (EM), for deciding the number of sub-classes (nuances) for each emotion category. Using the MDA dimensions, it is possible to map the postures onto a multidimensional discriminant space defined by the axes that maximize the separation between groups and minimize the variance within groups.

For each culture, an MDA model was created using the 108 postures. Specifically, for each posture we used the vectors describing its postural features and the most frequent emotion category (four categories) that was assigned by the observer of that culture to the avatar associated with that posture.

In the MDA model for each culture, the four emotion groups of postures are quite well separated within the two dimensional space created by the first two discriminant dimensions. In fact, for each of the three models, the classification performances are very high. The Japanese model correctly classifies (with respect to the Japanese observers) 90% of the postures. The Sri Lankan model correctly classifies (according to the Sri Lankan observers) 88% of the

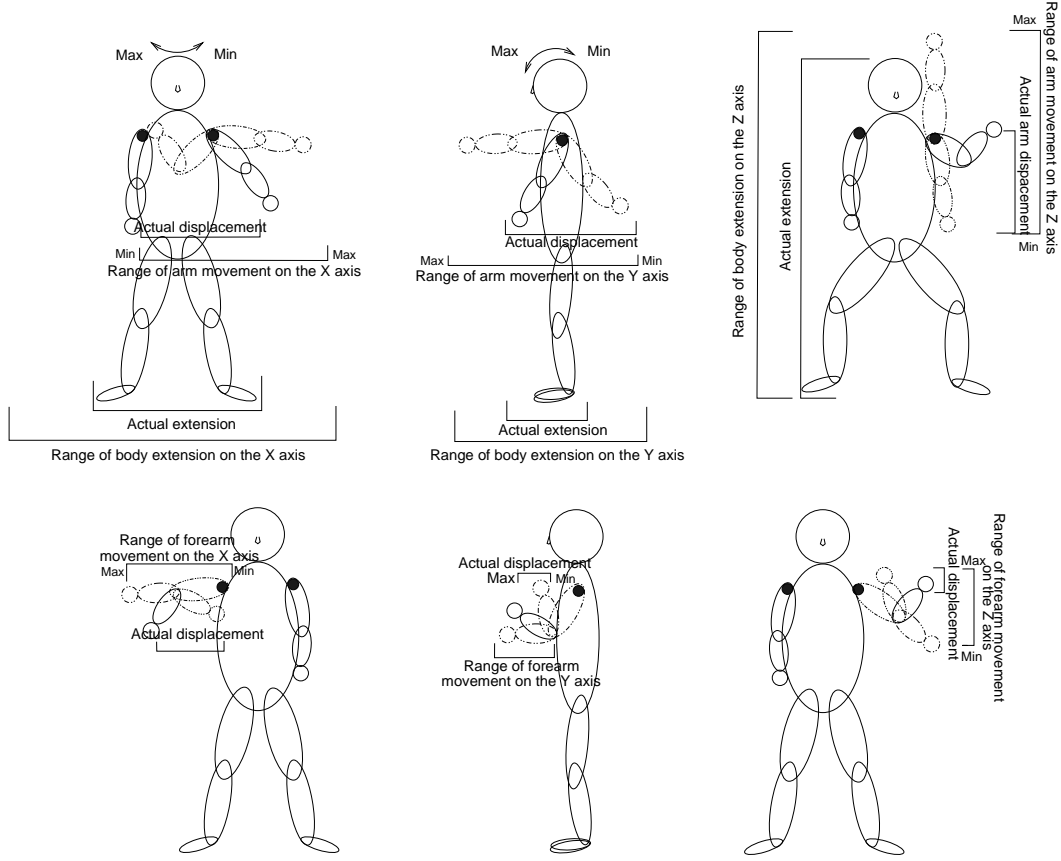


Fig. 3. The figure depicts examples of the ranges of normalization for the posture distance features. The upper part shows the rotation of the head, and the relation between the range of motion of the arm and its actual displacement along the three axes. The lower part shows the relation between the range of motion of the forearm and its actual displacement along the three axes.

postures, and the American model correctly classifies (according to the American observers) 78% of the postures.

As the models perform quite well, we can use them to ground the differences between cultures on the features. Hence, we analyzed the equations of the discriminant functions within each of the three MDA models, and extracted the set of features that were identified by MDA as most relevant for the discrimination process. We observed that the three models share only three features, while the overall set of important features is different for each model. Table 8 shows the feature sets for each culture. From this we can see that the Sri Lankans have only one feature unique to their culture while the Japanese and Americans have several. This discovery may indicate that the Japanese and the Americans are at opposite ends of the spectrum in their way to perceive emotion from posture, with the Sri Lankans residing somewhere in-between.

In evaluating the feature sets for each culture, one interesting finding was that

Table 7

The table lists the set of posture features used. The “Code” column indicates the feature codes used in the paper. The following short-cuts are used: L: Left, R: Right, B: Back, F: Front.

Code	posture features	Code	posture features
V4	$Orientation_{XY}$: B.Head - F.Head axis	V5	$Orientation_{YZ}$: B.Head - F.Head axis
V6	$Distance_z$: R.Hand - R.Shoulder	V7	$Distance_z$: L.Hand - L.Shoulder
V8	$Distance_y$: R.Hand - R.Shoulder	V9	$Distance_y$: L.Hand - L.Shoulder
V10	$Distance_x$: R.Hand - L.Shoulder	V11	$Distance_x$: L.Hand - R.Shoulder
V12	$Distance_x$: R.Hand - R.Elbow	V13	$Distance_x$: L.Hand - L.Elbow
V14	$Distance_x$: R.Elbow - L.Shoulder	V15	$Distance_x$: L.Elbow - R.Shoulder
V16	$Distance_z$: R.Hand - R.Elbow	V17	$Distance_z$: L.Hand - L.Elbow
V18	$Distance_y$: R.Hand - R.Elbow	V19	$Distance_y$: L.Hand - L.Elbow
V20	$Distance_y$: R.Elbow - R.Shoulder	V21	$Distance_y$: L.Elbow - L.Shoulder
V22	$Distance_z$: R.Elbow - R.Shoulder	V23	$Distance_z$: L.Elbow - L.Shoulder
V24	$Orientation_{XY}$: Shoulders axis	V25	$Orientation_{XZ}$: Shoulders axis
V26	$Orientation_{XY}$: Heels axis	V27	$3D - Distance$: R.Heel - L.Heel

Table 8

The table shows the features selected by the three MDA models. The first and the third columns indicate the set of features, the second and the fourth columns the culture-models using them

Selected features	Cultures
V5 - $Orientation_{YZ}$: B.Head - F.Head	JA, SL, US
V11 - $Distance_x$: L.Hand - R.Shoulder	JA, SL, US
V15 - $Distance_x$: L.Elbow - R.Shoulder	JA, SL, US
V13 - $Distance_x$: L.Hand - L.Elbow	JA, SL
V16 - $Distance_z$: R.Hand - R.Elbow	JA, SL
V27 - $3D - Distance$: R.Heel - L.Heel	JA, SL
V7 - $Distance_z$: L.Hand - L.Shoulder	SL, US
V18 - $Distance_y$: R.Hand - R.Elbow	SL, US
V25 - $Orientation_{XZ}$: Shoulders axis	SL, US
V6 - $Distance_z$: R.Hand - R.Shoulder	JA
V8 - $Distance_y$: R.Hand - R.Shoulder	JA
V9 - $Distance_y$: L.Hand - L.Shoulder	JA
V17 - $Distance_z$: L.Hand - L.Elbow	JA
V22 - $Distance_z$: R.Elbow - R.Shoulder	SL
V4 - $Orientation_{XY}$: B.Head - F.Head	US
V20 - $Distance_y$: R.Elbow - R.Shoulder	US
V26 - $Orientation_{XY}$: Heels axis	US

the arm stretched along the body (V6) and the head bent (V5) are necessary features for the Japanese in recognizing sadness in the avatar. Indeed, (V6) is considered to be important only to the Japanese, and seems to reflect a typical posture they frequently use to express sadness or remorse. Refer to

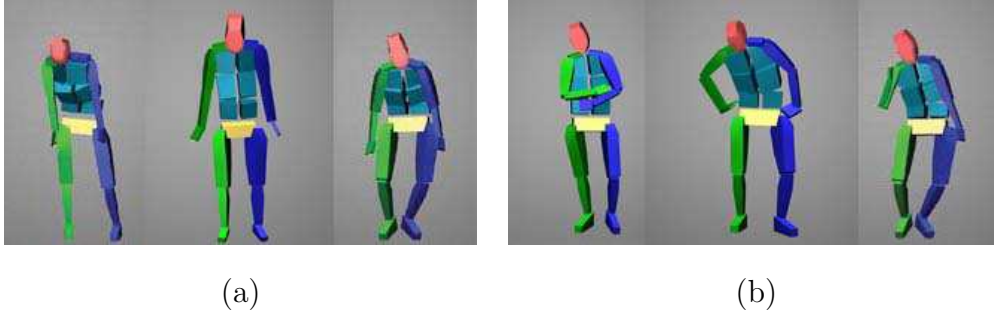


Fig. 4. Examples of the 3D affectively expressive avatars in which (a) all three cultures said sad and (b) only the Sri Lankans and the Americans said sad

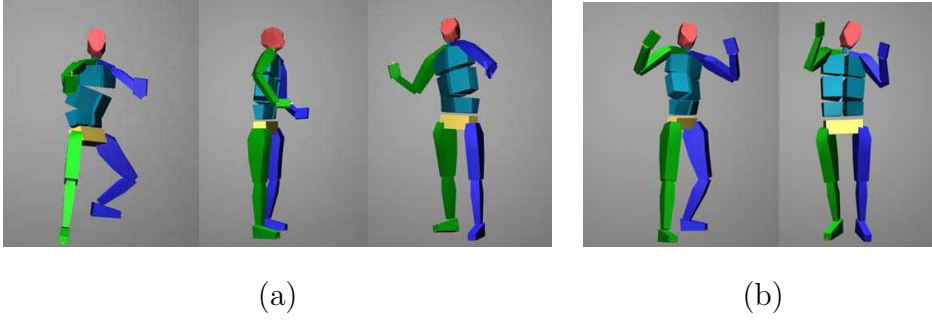


Fig. 5. Examples of the 3D affectively expressive avatars in which the Japanese said (a) fear and (b) happy while the Sri Lankans said angry

Figure 4(a) to see examples. However, while the other two cultures generally associated this type of posture to sadness, many avatars with the face raised a little and the hands close to the chest (a combination of (V18) and (V22)) were also considered as sad (Figure 4(b)). We can see that this feature is shared between the Sri Lankans and the Americans, while it is not considered important by the Japanese. Another relevant finding is that the Japanese seem to attribute happy to postures in which the arms remain close to the body's side (again (V6)), and both fear and happy (Figure 5(a) and (b), respectively) when the arms are raised to mid-level, whereas the Sri Lankans more often appear to consider these postures as angry. In general, fear and angry for all three cultures consist of a wide variety of postures (as referenced by Figure 1(a)(b)). Specifically, in the case of fear for the Sri Lankans, the elbow is always below the shoulder, and very low in the case of sad. This feature, (V22), is the single unique feature for the Sri Lankans. Furthermore, we notice that when the Japanese associated a posture with fear, the Sri Lankans did not agree. In fact, we see that angry was typically selected as the most frequent label by the Sri Lankans for these postures. This result seems to correlate to the distance between the heels (refer to Figure 1(b)) (V27), which is a feature shared by the Japanese and the Sri Lankans, however, it appears to be used differently within those cultures. We can see this difference in the Sri Lankan selected angry postures, which have a wider stance and a tilted or turned head.

5 Conclusions

The results of our analyses are useful for guiding the design and implementation of embodied agents, and indicate a need for considering culture as one specific design factor. First, we found that the three cultures obtained a moderate level of agreement when judging posture according to a set of predefined emotion categories. Yet, while some similarities do exist amongst cultures in how they convey, recognize, and attribute emotional meaning to posture, there are also some differences. Our results provide evidence to indicate that emotions are both universal and culturally specific. In evaluating cultural dimensions, some cultures are more similar than others. According to the individualism dimension, the Sri Lankans and the Japanese have quite similar ratings (37 and 43, respectively), while the Americans rank significantly higher (91) (Hofstede (2003)). This may help to prove differences between these cultures in recognizing emotion. According to the uncertainty dimension, it is the Sri Lankans and the Americans who are very similar (42 and 43, respectively), while the Japanese are drastically different (92). This may prove our results that the Sri Lankans appear to be in the middle of the other two cultures if you consider an “emotion-perception” spectrum.

Second, in further analysis, significant cultural differences were observed when considering intensity ratings, specifically, that the Japanese easily assigned a stronger intensity to the animated body postures than did the other cultures. Two conclusions may be drawn from this. One, that the Japanese are typically less animated in bodily *expressions* of emotion, supporting Scherer et al. (1988) findings discussed in the Introduction. Two, that in *perceiving* bodily expressions of emotion, the Japanese may believe that the feeling experienced by the poser is more intense than what s/he is actually displaying, thus, attributing a higher intensity rating. These were the findings of Matsumoto et al. (2002) for low-intensity facial expressions (i.e., less animated).

Third, in evaluating how the cultures use the emotion words, we found that the Japanese tended to use the joy and depressed labels more often than the Sri Lankans, initially indicating that they may assign very different meanings to each label within the word pairs. However, we found that the semantic patterns for both cultures were quite similar. Based on this, we may be able to conclude that the Japanese and the Sri Lankans have similar lexicons for these emotional categories. This could be a result of these cultures being considered significantly collectivistic. In a linguistic study by Semin et al. (2002), results showed that the use of emotion terms may be a result of social differences, individual or group, across cultures.

Fourth, we grounded the observed perceptual differences into a set of low-level posture features from which a separate model was created for each culture.

We found that classification rates were quite significant when testing these models on our set of affective avatars.

From this study we have identified several issues to be addressed in future research. One, the collection of the set of postures should include relatively equal numbers of actors from each of the cultures in order to study the existence of dialects within cultures. For this study it was difficult to find enough Sri Lankan and American actors as the study was conducted in Japan. Another issue in the collection of postures was caused by a problem of occlusion of some of the markers, most typically the markers placed on the chest. A larger variety of postures could be collected using a different motion capture system. Two, it would be interesting to explore more than the basic set of universal emotions since these non-basic emotions may occur more in human-computer interactions. In fact, it could be expected that emotion categories that are considered to be more ambiguous (greater variability) may exhibit a much stronger cultural bias.

6 Acknowledgment

This study was supported by a Grants-in-Aid for Scientific Research from the Japanese Society for the Promotion of Science granted to N. Bianchi-Berthouze.

References

- Altman, D. G., 1991. *Practical Statistics for Medical Research*. Chapman and Hall, England.
- Ballin, D., Gillies, M., Crabtree, I., 2004. A framework for interpersonal attitude and non-verbal communication in improvisational visual media production. In: *First European Conference on Visual Media Production*, IEEE. London.
- Becker, C., Prendinger, H., Ishizuka, M., I., W., 2005. Evaluating affective feedback of the 3d agent max in a competitive card game. In: Tao, J., Tan, T., Picard, R. (Eds.), *First International Conference on Affective Computing and Intelligent Interaction*. Springer-Verlag, pp. 466–473.
- Bianchi-Berthouze, N., Kleinsmith, A., 2003. A categorical approach to affective gesture recognition. *Connection Science special issue on Epigenetic Robotics - Modeling Cognitive Development in Robotic Systems* 15 (4), 259–269.
- Brandt, M., Boucher, J., 1986. Concepts of depression in emotion lexicons of eight cultures. *International Journal of Intercultural Relations* 10, 321–346.

- Chi, D., Costa, M., Zhao, L., Badler, N., 2000. The emote model for effort and shape. In: *Proceedings of SIGGRAPH 2000, ACM Computer Graphics Annual Conference*. New Orleans, pp. 173–182.
- Cohen, J., 1960. A coefficient of agreement for nominal scale. *Educational Psychol Measure* 20, 37–46.
- de Gelder, B., Snyder, J., Greve, D., Gerard, G., Hadjikhani, N., 2003. Fear fosters flight: A mechanism for fear contagion when perceiving emotion expressed by a whole body. *Proc. of the National Academy of Science* 101 (47), 16701–16706.
- Dunn, P., Marinetti, A., 2002. Cultural adaptation: Necessity for eLearning. <http://www.linezine.com/7.2/articles/pdamca.htm>.
- Ekman, P., 1994. Strong evidence for universals in facial expressions: A reply to russell’s mistaken critique. *Psychological Bulletin* 115, 268–287.
- Ekman, P., Friesen, W., 1975. *Unmasking the Face: A Guide to Recognizing Emotions from Facial Expressions*. Prentice Hall.
- Ekman, P., Friesen, W., Ellsworth, P., 1972. *Emotion in the human face: Guidelines for research and an integration of findings*. Pergamon Press, New York.
- Elfenbein, H., Ambady, N., 2002. On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin* 128, 205–235.
- Elfenbein, H., Mandal, M., Ambady, N., Haizuka, S., 2002. Cross-cultural patterns in emotion recognition: Highlighting design and analytical techniques. *Emotion* 2 (1), 75–84.
- Field, A., 2000. *Discovering statistics: Using SPSS for WIndows*. Sage, London.
- Friesen, W., 1972. *Cultural Differences in Facial Expressions in a Social Situation: An Experimental Test of the Concept of Display Rules*. Doctoral dissertation, University of California, San Francisco.
- Hastie, T., Tibshirabi, R., 1996. Discriminant analysis by gaussian mixture. *Journal of the Royal Statistical Society B* (58), 155–176.
- Hofstede, G., 2003. *Geert Hofstede Cultural Dimensions*. <http://www.geert-hofstede.com/>.
- Keltner, D., Ekman, P., 2003. Expression of emotion. In: Davidson, R., Scherer, K., Goldsmith, H. (Eds.), *Handbook of Affective Sciences*. Oxford University Press, pp. 411–414.
- Keltner, D., Ekman, P., Gonzaga, G. C., Beer, J., 2003. Facial expression of emotion. In: Davidson, R., Scherer, K., Goldsmith, H. (Eds.), *Handbook of Affective Sciences*. Oxford University Press, New York, pp. 415–431.
- Kleinsmith, A., Fushimi, T., Bianchi-Berthouze, N., 2005. An incremental and interactive affective posture recognition system. In: *User Modeling 2005 Workshop: Adapting the Interaction Style to Affective Factors*. <http://www.di.uniba.it/intint/UM05/list-ws-um05.html>.
- Kopp, S., Gesellensetter, L., Kramer, N., Wachsmuth, I., 2005. A conversational agent as museum guide – design and evaluation of a real-world application. In: Tao, J., Tan, T., Picard, R. (Eds.), *Intelligent Virtual Agents*. Springer-Verlag, pp. 329–343.

- Krenn, B., Neumayr, B., Gstrein, E., Grice, M., 2004. Lifelike agents for the internet: A cross-cultural case study. In: Payr, S., Trappl, R. (Eds.), *Agent Culture: Human-Agent Interaction in a Multicultural World*. Lawrence Erlbaum Associates, NJ, pp. 197–229.
- Kudoh, T., Matsumoto, D., 1985. Cross-cultural examination of the semantic dimensions of body postures. *Journ. of Personality and Social Psychology* 48 (6), 1440–1446.
- Lamolle, M., Mancini, M., Pelachaud, C., Abrilian, S., Martin, J.-C., Devillers, L., 2005. Contextual factors and adaptative multimodal human-computer interaction: Multi-level specification of emotion and expressivity in embodied conversational agents. In: *Modeling and Using Context: 5th International and Interdisciplinary Conference*. Springer-Verlag, pp. 225–239.
- Lewis, R. D., 1999. *When cultures collide: Managing successfully across cultures*. Nicholas Brealey, London.
- Lim, M., Aylett, R., Jones, C., 2005. Affective guide with attitude. In: Tao, J., Tan, T., Picard, R. (Eds.), *First International Conference on Affective Computing and Intelligent Interaction*. Springer-Verlag, pp. 72–79.
- Matsumoto, D., 1989. Cultural differences on the perception of emotion. *Journal of Cross-Cultural Psychology* 20, 92–105.
- Matsumoto, D., 1992. American-japanese cultural differences in the recognition of universal facial expressions. *Journal of Cross-Cultural Psychology* 23, 72–84.
- Matsumoto, D., 2005. Culture and nonverbal behavior. In: Manusov, V., Patterson, M. (Eds.), *Handbook of Nonverbal Communication* (in press). Sage, Thousand Oaks, CA.
- Matsumoto, D., Consolacion, T., Yamada, H., Suzuki, R., Franklin, B., Paul, S., Ray, R., Uchida, H., 2002. American-japanese cultural differences in judgments of emotional expressions of different intensities. *Cognition and Emotion* 16, 721–747.
- Maya, V., Lamolle, M., Pelachaud, C., 2004. Influences and embodied conversational agents. In: *Third International Conference on Autonomous Agents and Multi-Agent Systems*. ACM, pp. 1306–1307.
- Meeren, H., van Heijnsbergen, C., de Gelder, B., 2005. Rapid perceptual integration of facial expression and emotional body language. *PNAS* 102 (45), 16518–16523.
- Mehrabian, A., Friar, J., 1969. Encoding of attitude by a seated communicator via posture and position cues. *Journal of Consulting and Clinical Psychology* 33, 330–336.
- Mesquita, B., 2003. Emotions as dynamic cultural phenomena. In: Davidson, R., Scherer, K., Goldsmith, H. (Eds.), *Handbook of Affective Sciences*. Oxford University Press, New York, pp. 871–890.
- Neill, S., Caswell, C., 1993. *Body language for competent teachers*. Routledge, London.
- Picard, R., 1998. *Toward Agents that Recognize Emotion*. Actes Proc. IMAGINA, Monaco.

- Prendinger, H., Dohi, H., Ishizuka, M., 2004. Eempathic embodied interfaces: Addressing users' affective state. In: Tao, J., Tan, T., Picard, R. (Eds.), Tutorial and Research Workshop on Affective Dialogue Systems. Springer-Verlag, pp. 53–64.
- Russell, J. A., 1994. Is there universal recognition of emotion from facial expressions? a review of the cross-cultural studies. *Psychological Bulletin* 115, 102–141.
- Scherer, K. R., Johnstone, T., Klasmeyer, G., 2003. Vocal expression of emotion. In: Davidson, R., Scherer, K., Goldsmith, H. (Eds.), *Handbook of Affective Sciences*. Oxford University Press, New York, pp. 433–456.
- Scherer, K. R., Wallbott, H. G., Matsumoto, D., Kudoh, T., 1988. Emotional experience in cultural context: A comparison between europe, japan, and the united states. In: Scherer, K. R. (Ed.), *Faces of Emotions*. Erlbaum, Hillsdale, New Jersey.
- Schimmack, U., 1996. Cultural influences on the recognition of emotion by facial expressions. *Journal of Cross-Cultural Psychology* 27, 37–50.
- Semin, G., Gorts, C., Nandram, S., Semin-Goossens, A., 2002. Cultural perspectives on the linguistic representation of emotion and emotion events. *Cognition and Emotion* 16, 11–28.
- von Laban, R., 1988. *The mastery of movement*. Princeton.
- Wierzbicka, A., 1999. *Emotions across Languages and Cultures: Diversity and Universals*. Cambridge University Press, London.