

# Affective Body Expression Perception and Recognition: A Survey

Andrea Kleinsmith and Nadia Bianchi-Berthouze

**Abstract**— Thanks to the decreasing cost of whole-body sensing technology and its increasing reliability, there is an increasing interest in, and understanding of, the role played by body expressions as a powerful affective communication channel. The aim of this survey is to review the literature on affective body expression perception and recognition. One issue is whether there are universal aspects to affect expression perception and recognition models or if they are affected by human factors such as culture. Next we discuss the difference between form and movement information as studies have shown that they are governed by separate pathways in the brain. We also review psychological studies that have investigated bodily configurations to evaluate if specific features can be identified that contribute to the recognition of specific affective states. The survey then turns to automatic affect recognition systems using body expressions as at least one input modality. The survey ends by raising open questions on data collecting, labeling, modeling, and setting benchmarks for comparing automatic recognition systems.

**Index Terms**— Affective body posture, affective body movement, affective recognition systems, cross-cultural differences, spatio-temporal affective body features

## 1 INTRODUCTION

In recent years, there has been a growing interest in the development of technology that has the crucial ability to recognize people's affective states [1] as the role played by affect in human development and everyday functioning is now well recognized [2]. Increasing attention is being paid to the possibility of using body expressions to build affectively aware technologies. Three possible reasons for this attention are scientific, technological and social. First, as will be discussed in Section 2, differently from what was previously thought [6], [7], more and more studies from various disciplines have shown that body expressions are as powerful as facial expressions in conveying emotions [7]-[12]. Second, as technologies encountered by the average person on a day-to-day basis become more and more ubiquitous [1], they afford a multimodal interaction in which body expressions are assuming an important role that goes beyond that of gesture. A typical example is offered by whole-body computer games (e.g., Nintendo Wii and Microsoft Kinect) where body movement is not only a means to control the interaction between us and the games, but also a way to capture and affect our own emotional and cognitive performances [16]-[19]. As such, there is an increasing need to better understand and fully exploit this channel in human-computer interaction [13]-[15].

Third, the relevance of body expressions and the benefits of developing applications into which affect perception can be integrated is evident in many areas of society, such as security, law enforcement, games and entertainment, education and health care. For example, teachers are taught how to read affective aspects of students' body language and how to react appropriately through their body language and actions [20] in an effort to help students maintain motivation. Stu-

dents lose motivation when high levels of affective states such as frustration, anxiety, fear of failure, etc., are experienced [21], [22]. In chronic pain rehabilitation, [23], [24], specific movements and postural patterns (called "guarding behaviour") inform about the emotional conflict experienced by the patients and their level of ability to relax [25], [26]. Clinical practitioners make use of such information to tailor their support to patients during therapy.

While there is a clear need to create technologies that exploit the body as an affective communication modality, there is a less clear understanding on how these systems should be built, validated and compared [27], [28]. This paper aims at reviewing the literature on affective body expression perception and recognition models and raises some open questions on data collecting, labeling, modeling, and setting benchmarks for comparing automatic recognition models. The focus is on whole body static postures and whole body movements rather than gestures as there is already an extensive literature focusing on this aspect of body expressions [29], [30], [31], [32]. However, some studies on gestures will be reviewed as they bring up more general issues important to the design of automatic affective body expression recognition systems.

The paper is organized as follows: Section 2 discusses the motivation for investigating affective body expression and recognition. Section 3 describes the universality argument with the focus on bodily affect recognition and conveyance. Section 4 briefly explains the roles of both form and motion information and surveys research aimed at mapping specific cues of bodily expressions to specific affective states and dimensions. Section 5 reports on the state-of-the-art of automatic affective bodily expression recognition systems. Section 6 provides a discussion on some of the issues identified from the literature that exist for creating such automatic systems. Finally, Section 7 provides a summary of the issues discussed throughout the paper and areas within affective

- A. Kleinsmith is with the Department of Computing, Goldsmiths, University of London, UK (email: a.kleinsmith@gold.ac.uk)
- N. Bianchi-Berthouze is with the UCL interaction Centre, UCL, UK (email: n.berthouze@ucl.ac.uk)

computing that still lack sufficient research.

## 2 THE IMPORTANCE OF THE BODY IN EMOTION CONVEYANCE AND PERCEPTION

*“Considering the emotional value of bodily expressions, it is somewhat surprising that the study of perception of whole-body expressions lags so far behind that of facial expressions.” [12].*

Affect expression occurs through combinations of verbal and nonverbal communication channels such as eye gaze, facial expressions, and bodily expressions [33]. Despite this wide range of modalities, the majority of research on nonverbal affect recognition has concentrated on facial expressions in particular [34], [35], [36]. Thus, a fair amount is known and accepted about affective facial expressions, such as some ways in which they are conveyed and recognized, their neurobiological bases [37] and an understanding about how to code them [3]. There is a well established coding system for facial expressions, FACS, developed by Ekman and Friesen [3] over the course of a decade [38]. The examination of facial expression perception has been the basis for learning how humans process affect neurologically [39].

The same cannot be said for affective bodily expressions. Only recently has affective computing research and related disciplines focused on body movement and posture. Indeed, in a 2009 article, de Gelder [34] states that 95% of the studies on emotion in humans have been conducted using facial expression stimuli while research using information from voice, music and environmental sounds make up the majority of the remaining 5%, with research on whole-body expressions comprising the smallest number of studies. Hence the question is: What role does bodily information play in affect recognition?

Bodily expressions have been recognized as more important for nonverbal communication than was previously thought [6], [7]. According to Mehrabian and Friar [6] and Wallbott [40], changes in a person’s affective state are also reflected by changes in body posture. Mehrabian and Friar found that bodily configuration and orientation are significantly affected by a communicator’s attitude toward her/his interaction partner. Ekman and Friesen [41], [42] conjecture that postural changes due to affective state aid a person’s ability to cope with the experienced affective state.

### 2.1 Body expressions vs. facial expressions

A number of studies have been carried out to understand the importance of body expressions with respect to facial expressions. Indeed, some affective expressions may be better communicated by the body than the face [7], [8], [37]. De Gelder [37] postulates that for fear specifically, by evaluating body posture, it is possible to discern not only the cause of a threat but also the action to be carried out (i.e., the action tendency). The face communicates only that there is a threat.

Darwin [43] surmised that people are able to control bodily movements during felt emotions. Ekman and Friesen [44] refer to this as the “*face>body leakage hypothesis*”. However,

they conclude that there is a lack of evidence to support Darwin’s claim by stating that “*most people do not bother to censor their body movements*” and instead people make a conscious attempt to control their facial expressions [45]. On the other side of the argument, Hocking and Leathers [46] believe that facial expressions are more difficult to control due to the expresser’s inability to view them, while body expressions can be visually monitored. Furthermore, they argue that due to a stereotyped expectation that deceivers in particular will display more body movements, there is a greater attempt to control them. Indeed, a number of studies have shown that fewer finger, hand and lower limb movements are used during deceptive situations [174], [175], [176].

The above research indicates that one modality is not necessarily more important than another modality in detecting deception, but that both the face and the body play important roles. Our purpose in discussing this issue is to highlight the importance of considering the body in the area of deception detection, specifically. Ultimately a multimodal approach may be the most effective as there are many factors to consider, e.g., low- vs. high-stakes deception, studies with student participants vs. criminal populations, etc.

Studies have also examined the role played by the body in communicating emotions when observers are presented with affective displays containing a combination of facial expressions and posture or movement. According to studies by de Gelder and colleagues, body expressions may provide more information than the face when discriminating between fear and anger [11] or fear and happiness [12]. In [11], de Gelder et al. examined incongruent displays of posture and facial expressions. The stimuli were created using validated databases [51], [52], [53]. The findings indicate that when the affective information displayed by the two channels is incongruent, body posture is the influencing factor over the recognized emotion. There was a significant decrease in facial expression recognition when face and body information were incongruent than when they were congruent. The results were replicated in a more recent study [12] aimed at extending the set of emotions by investigating fear and happy congruent and incongruent face-body images.

Preliminary studies by Pollick and colleagues [54], [55] examined high, low and neutral saliency facial expressions combined with motion captured arm movements representing knocking motions for angry, happy, sad and neutral. The results showed that when the modalities were viewed separately, movement information for angry was more heavily weighted than facial information [54]. Furthermore, angry knocking motions were perceived as more intense and with higher recognition rates than low saliency angry facial expressions [55].

A study by Tuminello and Davidson [56] reported higher recognition rates by children for afraid and angry expressions when body posture information was added to facial expression photographs. More complex patterns of interferences between incongruent face and body expressions were identified in Willis, et al [57]. The authors suggest that the valence of the two expressions appears to have a major role

Ying Gao's can't dress requires people to control their body movement, this is an interactive attempt to change the affective state of the audience engaging in the

The research gap in 2009

in prioritizing one over the other in order to address social and threat issues.

The studies presented throughout this section show that the body does indeed play an important role in the expression and perception of affect. In fact, for some affective states in particular, more attention is paid to body expressions than facial expressions.

## 2.2 Body Expressions and affective dimensions

While many studies presented throughout the survey describe body expressions in terms of *discrete* emotions, fewer studies have attempted to classify them in terms of affective dimensions. Ekman and Friesen [41] indicated that the body may be better for communicating broader dimensions of affect than discrete categories. Paterson and colleagues [58] aimed to map head and arm movements to an affective space. They examined not only how well affect may be recognized but also the structure of the representation of affect. Observers viewed acted affective knocking motions and judged the emotion displayed. A 2D affective space was obtained by applying a statistical technique to the observer judgments. The mapping was shown to reflect a circumplex model of affect with levels of arousal depicted on the first dimension and levels of valence depicted on the second dimension. These results show that similar to research on the structural representation of *experienced* affect, valence and arousal dimensions are also used by human observers when *describing* affective body expressions. A significantly higher percentage of variance was covered by the arousal dimension which may indicate that arousal is better conveyed by the body than valence for the knocking action considered.

Similar results were obtained in studies by Kleinsmith et al [59], [60] and Karg et al [61]. Kleinsmith et al examined affective dimensions of whole body posture with acted postures first [59] and recently followed up with non-acted postures [60]. First, acted static postures were mapped to an affective space which identified arousal first, valence second and action tendency third as the main discriminative dimensions. Their subsequent study [60] examined non-acted postures in a video game situation and also found higher agreement between observers for arousal over valence. Karg et al [61] examined acted whole body gait patterns according to three levels of arousal, valence, and dominance. Again, observer agreement was highest for arousal. These results may indicate that arousal is more easily identified from bodily expressions than valence. Indeed, findings by Clavel et al [178] appear to validate that assumption. In their study, face only and posture only levels of arousal and valence of an affective virtual agent were judged by observers. The results showed that arousal was better perceived than valence.

Identifying bodily expressions as combinations of discrete labels and levels of affective dimensions may provide a more complete description of the affective state exhibited; a single label may not always be enough to reflect the complexity of the affective state conveyed. Indeed, in the realm of affective computing, research is now focusing on an integration of the discrete emotions and affective dimensions approaches [177], [178].

The studies presented throughout Section 2 show that the

body is an important nonverbal communication channel. The body has also been shown to be more communicative than other modalities for some emotions and contexts. However, as Picard [33] points out, the manner in which humans convey emotion or affective messages in general is affected by many factors, such as age, gender, culture, and context. One factor that is being given attention by the affective computing community is culture.

## 3 THE UNIVERSALITY OF AFFECTIVE BODY EXPRESSIONS

*“Perhaps no issue has loomed larger or permeated the study of bodily communication than the extent to which such expressions are universal, which implies that they have a common genetic or neurological basis that reflects an evolutionary heritage shared by all humans, or relative, which implies that their form, usage, and interpretation are tied to individual cultures and contexts”* [62].

Matsumoto defines culture as *“a shared system of socially transmitted behaviour that describes, defines, and guides people’s ways of life”* [63]. The need for understanding how different people and cultures recognize and express affective body language has become more and more important in a number of real-life affective computing situations. For example, embodied museum agents are gaining much attention [64], [65]. Due to the diversity of people visiting, museums are a particularly appropriate arena in which to have an agent capable of recognizing differences due to personality, culture, etc. E-Learning systems may also benefit by taking into account various human factors. For instance, one study found high dropout rates due to ‘culturally insensitive content’ [66]. As systems replace humans, it is important that how they express and perceive non-verbal behaviors in a multi-cultural community is as natural as possible so that the user is not made uncomfortable.

There is evidence that the way in which affective states are expressed and controlled [6], as well as the interpretation of affect [67] is shaped by culture. Many researchers have used cross-cultural emotion recognition studies to validate evidence in favor of emotion universality as stated by Elfenbein et al [68]. The majority of the research on emotion universality has concentrated on the recognition of facial expressions using still photographs [69], [70], [71]. For some emotions, cross-cultural research has suggested universality of many modes of non-verbal behavior, including face, voice and body expressions, as well as changes in a person’s physiology.

Elfenbein and Ambady [73] have proposed the concept of emotional ‘dialects’. They consider the idea that emotional expression is a universal language, and that different dialects of that universal language exist across cultures. It is also hypothesized that accuracy rates in emotion identification will be higher with in-group (i.e., same culture) members because there is more familiarity with them [72], [74] as well as more motivation to understand the expressions of the members with whom they spend more time [75]. Therefore, the more time spent with members of a different culture may also lead to a better understanding of how to de-



code their affective expressions, i.e., out-group effects [56]. In-group effects in African American and European American children in recognizing emotion from facial and body expressions were investigated by Tuminello and Davidson [56]. In-group effects were found for some emotions for the European Americans whereas higher out-group effects were found for the African American children who spent more time with European American children.

Matsumoto and Kudoh carried out two studies designed to examine cross-cultural differences between Japanese and Americans in judging body posture according to a set of semantic dimensions [76], [77]. Based on other research [78], [79] (as cited in [77]), Kudoh and Matsumoto assert that differences reported between Japanese and Americans are almost always due to status being a more important aspect of the Japanese culture than the American culture. They further argue that postures can be a main dimension through which the semantic dimensions are interpreted [77]. Matsumoto and Kudoh's first study [76] investigated judgments of a corpus of verbal posture expressions from Japanese participants. To create the corpus, Japanese students provided written descriptions of postures from situations encountered in everyday life. Using the same methodology, the second study [77] investigated judgments from American participants. The researchers found that the same factors were extracted from the two sets of participants, but that the factor order was different between the two cultures. While cultural differences were found as expected, the authors questioned whether cultural differences would be found with posture *images* instead of verbal descriptions of postures.

To address this issue, Kleinsmith et al [80] examined differences between Japanese, Sri Lankan and American observers in perceiving emotion from whole body postures of a 3D faceless, cultureless, genderless 'humanoid' avatar. Both similarities and differences were found in how the cultures conveyed, recognized and attributed emotional meanings to the postures. For all three cultures, the sad/depressed category showed the highest agreement between actor and observer labels. This was expected according to a study which showed that the cultures share similar lexicons for depression-type words [81]. However, differences were found in how the cultures assigned intensity ratings for the emotions. In particular, the Japanese consistently assigned higher intensity ratings to more animated postures than did the Sri Lankans or the Americans. The authors asserted that, similar to the findings of Matsumoto et al. [82] for facial expressions, the Japanese may believe that the emotion being expressed is more intense than what is actually portrayed.

Although not conclusive, the results discussed may indicate a need for taking culture into account in various aspects of affect recognition research, such as labeling, how affect is both expressed and perceived by members of different cultures, and computational models for affect recognition.

## 4 MAPPING BODY EXPRESSIONS INTO AFFECT

This section provides a survey of research mapping body posture and movement into affect. The first issue discusses what bodily information is necessary for recognizing the affective state displayed. Sections 4.2 and 4.3 focus on psy-

chological studies aimed at examining bodily expressions to evaluate if specific features of the body can be identified that contribute to the recognition of specific affective states and dimensions. These studies have sought to understand these features according to aspects of body expressions, i.e., form and movement, and two main levels of bodily detail, i.e. high- and low-level descriptions. The remainder of this section is structured according to these three different aspects of body expressions and an overview of each study is presented.

The studies listed in Tables 1 and 2 will be used to support the discussion throughout the section. Although we do not present an exhaustive list of studies, our aim is to provide a general overview of the field. Table 1 lists the details of how the studies were carried out and the features examined. Table 2 lists the studies' findings, highlighting the set of features that characterize the affective states and dimensions investigated. A final discussion on the overall lessons learned from the studies is provided after Section 4.3.

TABLE 1

DESIGN DETAILS OF THE RESEARCH STUDIES AIMED AT MAPPING BODILY FEATURES INTO AFFECT (OBS. = OBSERVERS; + = ATTITUDE, NOT AFFECTIVE STATES; CG = COMPUTER GENERATED)

TABLE 2

DESCRIPTIONS OF SPECIFIC BODY FEATURES AND HOW THEY MAY BE ATTRIBUTED TO THE RECOGNITION OF SPECIFIC AFFECTIVE STATES ACCORDING TO INDIVIDUAL STUDIES OUTLINED IN TABLE 1

### 4.1 Body form vs movement in affect perception

According to neuroscience studies by Giese and Poggio [83] and Vaina et al [84], there are two separate pathways in the brain for recognizing biological information, one for form information (i.e., the description of the configuration of a stance) and one for motion information. The study of Lange and Lappe [85] makes even stronger claims by stating that "*...a model that analyses global form information and then integrates the form information temporally*" can better explain results from psychophysical experiments of biological motion perception. They argue that information about the temporal development of the movement is only used if necessary to resolve inconsistencies and if it is essential to the type of task. This argument is supported by previous research findings indicating that form information can be instrumental in the recognition of biological motion [86], [87], [88]. Hirai and Hiraki [86] showed that spatial scrambling of point light configuration stimuli had a stronger effect in the brain area involved in biological motion perception than temporal scrambling of the information. A neuropsychological study by McLeod et al [88] found that a brain-damaged patient who had a specific deficit in detecting moving stimuli, referred to as 'motion blind', was still able to detect a wide range of human actions (e.g., walking, cycling, etc.) from point-light displays by extracting form from motion information.

A recent study by Atkinson and colleagues [89] (refer to row 1 of Table 1) determined that both form and motion signals are assessed for affect perception from the body. Specifically, the authors concluded that motion signals can

be sufficient for recognizing basic emotions, but that recognition accuracy is significantly impaired when the form information is disrupted by inverting and reversing the clip. Through a systematic approach, the work by Omlor and Giese [90] and its more comprehensive follow up study [91] also suggest the existence of emotion specific spatio-temporal motor primitives that characterize human gait. Details of the study are discussed in Section 4.3.

These studies indicate that both form and motion information is useful and important for perceiving affect from body expressions. While movement can add information in some cases, it may be partially redundant to form information. Analyzing posture cues aids in discriminating between emotions that are linked with similar dynamic cues or movement activation [91]. As shown in Tables 1 and 2, both types of features have been explored in more detail with respect to different affective states, supporting their respective relevance to the recognition of affect in body expressions.

## 4.2 High-level description

One approach used in modelling affective body expressions is to investigate the relationship between affective states and a high-level description of either movement or form. Using acted ballet movements and postures, Aronoff and colleagues [92] (row 2 of Table 1) concluded that angular and diagonal configurations can be adopted to signify threatening behavior, while rounded postures demonstrate warmth. Other studies have acknowledged the important role that leaning direction plays in affect perception [93], [94], [95]. In an early study, James [95] (row 3 of Table 1) discovered the importance of more specific whole body features of posture, such as leaning direction, openness of the body and head position, (e.g., up, down, and tilted) for discriminating between affective states.

Dahl and Friberg [96] (row 4 of Table 1) explored to what extent the emotional intentions of a musician could be recognized from their body movements. The results showed that happiness, sadness and anger were well communicated, while fear was not. In the same study, movement cues were also examined and obtained similar results. The movement ratings indicated that observers used well defined cues to distinguish between intentions. For instance, anger is indicated by large, fairly fast and jerky movements while sadness is exhibited by fluid, slow movements. Their results also showed that the expression of the same emotion may vary strongly according to the instrument played (refer to happiness and fear in Table 2). We can also see from the Table that some of the emotions in their study (e.g., anger and happiness) share general patterns but differ in the qualifiers used (e.g., very slow vs. slow). Castellano et al [97] (Table 1, row 5) examined the quality of motion of the upper body and the velocity of head movements of a pianist across performances played with a specific emotional intention. Differences were found mainly between sad and serene, especially in the velocity of head movements, similar to Dahl and Friberg's results. Furthermore, they identified a relationship between the temporal aspects of a gesture and the emotional expression it conveys. They concluded by highlighting the need for more analysis of such features.

One goal of the study by Gross et al [27] was to establish a qualitative description of the movement qualities associated with specific emotions for a single movement task (knocking). Details of the study are presented in row 6 of Table 1. A qualitative analysis of the movement showed that motion perception was predicted most strongly for the high activation emotions (pride, angry and joyful). The analysis of the ratings showed interesting patterns; however, these need to be cautiously treated given that only 15 expressions were analyzed. Another aim of this study was to quantitatively assess the value of different emotions on different body expressions. The results were positive, meaning that a quantitative comparison of the expressions was possible. As one example, they found that the arm was raised at least 17 degrees higher for angry movements than for other emotions. This shows that it may be necessary to quantify features as body expressions may differ according to the presence or absence of a particular feature as well as the quantitative value of each feature.

Glowinski et al [98] (row 7 of Table 1) hypothesized that the use of a reduced set of features, upper body only, would be sufficient for classifying a large amount of affective behavior. Acted upper body emotional expressions from the GEMEP corpus [99] were statistically clustered according to the four quadrants of the valence-arousal plane. The authors concluded that 'meaningful groups of emotions' could be clustered in each quadrant and that the results are similar to existing nonverbal behavior research [40], [101], [131].

Another important point that can be observed from Table 2 is that there are similarities between core elements of body expressions related to specific emotion families. For instance, Gross et al [27] found that expanded limbs and torso signified both content and joy. Moreover, some affective states that are similar along a particular affective dimension appear to share some of the same characteristics. For example, from Table 2 we can see that sadness and shame are both characterized by slow, low energy movements [27], [96], [97], [101]. However, they tend to present other differences e.g., there appears to be a 'stepping back movement' in shame that is not present in sadness.

## 4.3 Low-level description

More recently and thanks to the possibility to automatize the analysis of body expressions, researchers have tried to ground affective body expressions into low-level descriptions of configurations to examine which postural cues afford humans the ability to distinguish between specific emotions. In Wallbott's study [40] (row 8 of Table 1), a category system was constructed which consisted of body movements, postures and movement quality. Differences were found to exist in how people evaluate posture in order to distinguish between emotions. Specific features were found to be relevant in discriminating between emotions. However, Wallbott himself stated that this was an initial study and asserted that additional studies needed to be carried out. In particular, he stressed the need for studies examining non-

<sup>1</sup> Cluster 1) high arousal-positive valence: elation, amusement, pride; cluster 2) high arousal-negative valence: hot anger, fear, despair; cluster 3) low arousal-positive valence: pleasure, relief, interest; 4) low arousal-negative valence: cold anger, anxiety, sadness.

acted expressions and cross-cultural issues.

In another study, Coulson [100] (listed in row 9 of Table 1) attempted to ground basic emotions into low-level static features that describe the configuration of posture. Computer generated avatars expressing Ekman's [42] basic emotions were used. His proposed body description comprises six joint rotations. Judgment survey results showed that observers reached high agreement in associating angry, happy, and sad labels to some postures. Coulson statistically evaluated the role played by each joint rotation in determining which emotion label was associated to each posture. All of the postures were kinematically plausible, however according to Coulson himself, *"the complexity of the stimuli meant that some postures looked rather unusual"* [100]. The posture features associated with each of the six emotions examined are described in Table 2.

DeMeijer [101] carried out a study to examine if specific body movements were indicative of specific emotions and which movement features accounted for these attributions. To this aim, seven movement dimensions listed in row 10 of Table 1 were utilized. Dancers were videotaped while performing specific characteristics of movements instead of explicitly enacting emotions. A separate group of observers rated each movement according to its compatibility with each emotion. The results showed that specific movements were attributed to each emotion category except for disgust and that specific features could be attributed to specific movements as listed in Table 2. Trunk movement (ranging from stretching: performed with straight trunk and legs; to bowing: performed with the trunk and head bowed and the knees bent slightly) was the most predictive for all emotions except anger and was found to distinguish between positive and negative emotions. For instance, surprise is characterized by a straight trunk and legs, backward stepping, and fast velocity movements; whereas fear is characterized by a bowed trunk and head, slightly bent knees, downward, backward, fast body movement and tensed muscles.

Using an information-based approach, De Silva and Berthouze [102] (row 11 of Table 1) investigated the relevance of body posture features in conveying and discriminating between four basic emotions. 24 features were used to describe upper-body joint positions and the orientation of the shoulders, head and feet to analyze affective postures from the UCLIC affective database [80]. A statistical analysis showed that few dimensions were necessary to explain the variability in form configuration. Similar results were obtained by clustering the postures according to the average observer labels. The vertical features were the most informative for separating happy from sad. Specifically, the hands were raised for happy and remained low along the body for sad. The features indicating the lateral opening of the body were the second most informative with the hands significantly more extended for happy and fear over sad. Using the same feature set, Kleinsmith and Berthouze [103] (third to last row of Table 1) extended this analysis by investigating how the features contributed to the discrimination between different levels of four affective dimensions.

Roether et al. [91] (second to last row of Table 1) carried out a three-step process to extract and validate the minimum set of spatio-temporal motor primitives that drive the per-

ception of particular emotions in gait. Through validation by creating walking patterns that reflect these primitives, they showed that perception of emotions is based on specific changes of joint angle amplitudes with respect to the pattern of neutral walking. In a third step, they investigated whether adaptation to natural affective gait patterns biased observer judgments of subsequent artificial patterns towards affectively ambiguous patterns. This is known as an 'after-effect' and is a tool commonly used in face perception (e.g., Leopold et al [179]). The purpose of employing this technique in the Roether et al study was to determine whether the extracted feature set sufficiently captured the important information for the perception of emotion. The results showed that there were after-effects in the perception of sad and happy movements, indicating that the feature sets are indeed complete.

In a recent study, Kleinsmith et al [60] (last row of Table 1) have taken steps towards addressing the issue of obtaining non-acted affective postures. They collected motion capture data from people playing sports games with the Nintendo Wii (part of the UCLIC affective database [80]) and used static postures from the data after a point in the game was won or lost. The average percentage of agreement between observers across multiple trials was set as the benchmark. Next, each posture was associated with a vector containing a low-level description of the posture. A statistical analysis of the features showed that the most important features were mainly the arms and upper body. While there was significant discrimination between the four separate emotions, greater discrimination was obtained between the more 'active' affective states (frustrated and triumphant) and the less 'active' states (concentrating and defeated). For instance, the shoulders were slumped forward with the arms were extended down and diagonally across the body for concentrating and defeated. Frustrated and triumphant postures were indicated with shoulders straight up or back and the arms raised and laterally extended.

In general, the analysis carried out in Section 4 indicates that the body may allow for discrimination between levels of affective dimensions as well as discrete emotion categories. However, this is far from meaning that there is a unique relationship between a discrete emotion and a body expression. A further review of Table 2 shows that most of the emotion studies also appear to have quite discriminative patterns when considering combinations of features over individual features. For example, Wallbott shows that the arms are crossed in front for both disgust and pride, but the discriminating feature appears to be head position (bent forward for disgust and bent backward for pride). Typically, the core elements of body expressions for sadness are the arms straight down, close to the side of the body [100], [102], [80], [91].

While some emotion categories (or emotion families) do share a core set of body expression characteristics, they also exhibit a number of variations for other parts of the body. For example, happiness and elated joy share the characteristic of the head bent back across several studies [100], [80], [40]. However, while the arms are raised in several cases [100], [102], [80], they remain straight down Roether et al's study [91] which may be due to the contextual factor of gait.

Throughout Sections 4.2 and 4.3, we have also seen various examples where more than one body expression pattern may be associated with the same emotion category. In some cases, the patterns shares some core features (as shown above for happiness). In other cases, the patterns within the same emotion class appear to be very different from each other (e.g., the variation in expressions of happiness and fear according to the musical instrument played [96]). This lends further support to the idea that there are contextual factors that may affect the way an emotional state is expressed.

All of this seems to be in line with Russell [180], who argues that prototypical expressions are actually quite rare. From this perspective, Table 2 may only highlight a limited amount of the variability that may be present in real situations. In fact, as shown in Table 1, most of these studies are based on acted expressions. According to Russell, the components that make up an emotion expression are not fixed, and each emotion reaction is not unique. By increasing the number of non-acted studies, less distinct, yet still discriminative patterns may emerge when context is not considered as shown by Kleinsmith et al [60].

Another important issue that becomes immediately apparent from an examination of Table 1 is the lack of a common vocabulary used by researchers for describing features (last 2 columns). The feature descriptions often appear to be based on subjective, qualitative evaluations, and hence difficult to compare across studies. Moreover, the high-level features are very context-dependent and difficult to compare without decomposing and interpreting the terms. Overall, for both high- and low-level descriptions, a systematic use of common, and possibly numerical descriptors is needed in order to more objectively compare body expressions as shown by Gross et al [27].

## 5 AUTOMATIC AFFECT RECOGNITION

There has been an increased pursuit of affective computing within the last few years in particular, as evidenced by recently published surveys in the field [104], [28], [35], [105]. The majority of the automatic affect recognition systems have focused mainly on using facial expressions [106], [107], [108] and voice [109], [110], [111], [112] as the input modality. Only recently have systems been built that center on the automatic recognition of bodily expressions monomodally [113], [114], [115], [116], [117], [103], [60], [118], [119], [120] and multi-modally [121], [22], [122], [123], [124]. Table 3 lists study details of systems that use affective bodily expressions as at least one input modality. Similar to the behavioral studies discussed in Section 4, most automatic recognition systems, independent of modality, rely on corpora that have been acted. Furthermore, many of these systems rely on the actors' labels to provide the ground truth. More recent studies are now addressing the problem of modeling non-acted and more subtle body expressions. Regardless of the modality examined, a lot of the studies validate the automatic affect recognition results by comparing them with a baseline computed on observers.

Some of the studies aimed at creating affective bodily expression recognition systems are reviewed in the remainder of this section. For completeness, multi-modal recognition

systems are also reviewed but only to highlight the contribution made by the body. A full discussion of multi-modal recognition remains outside of the scope of this paper.

TABLE 3

AUTOMATIC AFFECTIVE RECOGNITION SYSTEMS FOR BODY AND MULTIMODAL EXPRESSIONS. GR. TRUTH = GROUND TRUTH; BASIC = ANGER, DISGUST, FEAR, HAPPINESS, SADNESS, SURPRISE; SVM = SUPPORT VECTOR MACHINE; CALM = CATEGORIZING AND LEARNING MODULE; K-NN = K NEAREST NEIGHBOUR; MLP = MULTILAYER PERCEPTRON; DIFF = DIFFERENT; GP = GAUSSIAN PROCESS; \* = RECOGNITION RATE FOR POSTURE MODALITY ALONE; F = FRAME-LEVEL LABELLING; S = SEQUENCE-LEVEL LABELLING; B = BIASED; U = UNBIASED; II = INTERINDIVIDUAL; PD = PERSON-DEPENDENT; V = VALENCE; A = AROUSAL; D = DOMINANCE; # = RECOGNITION OF SMALL GROUP BEHAVIOURS TRIGGERED BY EMOTION, NOT EMOTION RECOGNITION DIRECTLY; CPR = CORRELATION PROBABILITY OF RECURRENCE

### 5.1 Affective Body Expressions

The majority of today's affective recognition systems of body posture and movement (top part of Table 3) have focused on extracting emotion information from dance sequences [125], [126], [127], [113]. Camurri and colleagues [113], [128] examined cues and features involved in emotion expression in dance for four affective states. After removing facial information, a set of motion cues was extracted and used to build automatic recognition models. The recognition of fear was the worst, achieving below chance level classification rates. Fear was most often misclassified as anger. This is an intriguing result because body movement was used as opposed to static postures, and as postulated by Coulson [100], dynamic information may help to increase recognition rates of fear in particular. Other automatic misclassifications occurred between joy and anger, and joy and grief. The misclassification of grief as joy is also interesting given the authors' examination of the quality of motion feature, which showed joy movements to be very fluid and grief movements to be quite the opposite. Kapur et al [114] used acted dance movements from professional and non-professional dancers. Observers correctly classified the majority of the movements and automatic recognition models achieved comparable recognition rates.

The use of dance movements for building affect recognition systems is interesting; however, these movements are exaggerated and purposely geared toward conveying affect. Body movements and postures that occur during day-to-day human interactions and activities are typically more subtle and not overtly emotionally expressive.

Turning to non-dance-based automatic affective body expression recognition, Pollick and colleagues [115] carried out a study in which they compared automatic affect recognition model performance with human recognition performance in recognizing different movement styles in terms of affectively performed knocking, lifting and waving actions. In particular, are human observers able to make use of the available movement information? The results indicated that the system was able to discriminate between affective states more consistently than the human observers.

Karg et al [61] examined automatic affect recognition for discrete levels of valence, arousal and dominance in affect-

tive gait patterns. Recognition rates were best for arousal and dominance, and worst for valence. The results were significantly higher than observer agreement on the same corpus of affective gait patterns reported in Section 2.2. Sanghvi et al [132] also used the recognition rates of observers as a baseline for system evaluation. They extracted posture and movement features from body only videos of children playing chess with an iCat robot in an attempt to recognize levels of engagement. A user study indicated that both posture configuration features and spatio-temporal features may be important for detecting engagement. The best automatic models achieved recognition rates that were significantly higher than the average human baseline.

Kleinsmith and Bianchi-Berthouze have examined automatic recognition of affect from whole body postures using a low-level posture description in an acted situation first [117], [103], progressing to a non-acted situation most recently [60]. In their earliest work [117] on acted postures, they built an automatic recognition model for three discrete categories, and achieved a very high average classification rate. As a second step, automatic models were built for recognizing levels of four affective dimensions [103]. While these models also achieved high classification levels, they were somewhat lower than the models for the discrete categories. In their most recent work [60] using non-acted postures and subtle affective states in video games, their models achieved recognition rates lower than their acted studies, but similar to the target rate set by computing the level of agreement between *sets* of observers (described in detail in Section 6.2).

These studies have shown that by using either static or dynamic features, the systems achieve results that are similar to the target set by either a self-reported ground truth or the level of agreement between observers. Most recently, Kleinsmith et al [60] raise the question of how to set the target rate for the evaluation process. They argue that this target should not be based on the level of agreement between the observers' judgments used to build the system, but instead it should be based on an unseen set of observers. Learning and testing systems on the same set of observers may produce results that do not take into account the high variability that may exist between observers (especially if they are not experts). Hence, their approach is to test recognition systems for their ability to generalize not only to new *postures* but also to new *observers*.

Another important issue related to both building and evaluating recognition systems was raised in Bernhardt and Robinson's work [116]: the existence of individual differences between the expressers. They made the point that not only is affect readily seen in body movement, but individual idiosyncrasies are also noticeable, which can make classification more difficult. Bernhardt and Robinson, and more recently Gong et al [119] tested the differences between models with personal biases removed and with personal biases remaining. Both studies used Pollick et al's motion capture database [130]. The automatic recognition rates achieved in both studies were considerably higher with personal biases removed over the rates for the biased motions. Their results were compared with the observers' agreement from Pollick et al's study [131] to obtain a baseline on which

to validate their models. The results indicated that the automatic models [116], [119] and the observers' rates [131] were comparable. From this Bernhardt and Robinson concluded that "*even humans are far from perfect at classifying affect from non-stylised body motions*", suggesting that creating a 100% accurate affect recognition system is unlikely given that humans are not 100% accurate. In a more recent study, Bernhardt and Robinson [120] extended their system using motion capture of additional actions from the same database to create a system to detect emotion from connected action sequences. In this case, the average recognition was similar to the previous system with personal biases removed.

Using affective whole body gait patterns, Karg et al [61] built automatic recognition models to examine the differences between inter-individual and person-dependent recognition accuracies for emotion categories. Similar to the results of Gong et al [119] and Bernhardt and Robinson [116], the inter-individual recognition accuracies were much lower than the person-dependent recognition accuracies. However, automatic recognition accuracies were higher than the observer agreement rate which was used as a baseline.

Savva et al. [118] investigated these issues in a non-acted situation. They proposed a system based on dynamic features to recognize emotional states of people playing Wii tennis. Individual idiosyncrasies were removed by normalizing each expression according to the minimum and maximum values of the features for that participant. The best results were obtained using angular velocity, angular frequency and amount of movement. Overall, the system was able to correctly classify a high percentage of both the high and low intensity negative emotion expressions and the happiness expressions, but considerably fewer of the concentration expressions. The results for inter-individual and person-dependent models were very similar and just slightly below the observer agreement. Two main reasons were hypothesized for this difference. First, differently from other studies [61], [119], [116], the level of agreement set as the target in Savva and Berthouze was based on a new set of observers using a simplified version of the method proposed by Kleinsmith et al [60] to take into account the high variability between observers. Second, a non-acted dataset was used in this study. An analysis of the results highlighted the high variability between expressions belonging to the same category which could justify the lower performances with respect to the acted situations discussed previously. The high variability was due to the diversity of the players' playing styles. Some participants played the game using only their hand/wrist in comparison with other participants who used their arm and shoulder as well. In the former case, high negative emotions may be expressed by jerky and fast movements of the wrists while the rest of the body is very controlled. In the latter case, affective states may be expressed by jerky movements of a larger part of the body.

The existence of strategy differences in playing full-body games is consistent with the results of other studies [181], [182]. Interviews and quantitative analysis of body movements in these latter studies showed that game strategy differences were due not only to differences in game skills and experience levels, but also to players' game playing motiva-



tions (e.g., winning vs. role-play experience). Again, this highlights how critical it is to study non-acted situations in which various factors contribute to high variability both in the way people express emotions (e.g., body strength of the player) and also in the way people perform an action. In order to increase the performance, we need to consider models that take into account these factors and also optimize the choice of features on the basis of individual differences.

This section shows that automatic recognition of affect using acted and non-acted expressions achieve results well above chance level and comparable to, or above observers' agreement. Both postural and configurational features appear to contribute to the positive results; however, a systematic comparison of the contribution made by these two types of features has not been carried out. The section also highlights the importance of taking into account individual differences in the expressions when building and evaluating the recognition systems. Also, the use of non-experts in labeling data and an increase in the number of applications using non-acted data may require evaluation methods that take into account variability between observers.

## 5.2 Multimodal Expressions

The bottom part of Table 3 lists multimodal automatic affect recognition systems which include body posture or movement information as one of the modalities examined. Two of these systems have been designed by Picard's group at MIT [133], [121], [22]. Focused on non-acted affect, their system models a description of the body and attempts to recognize discrete levels of a child's interest [121] and self-reported frustration [22] from postures detected through the implementation of a chair embedded with pressure sensors, facial expressions, and task performance. Their postures were defined by a set of eight coarse-grained posture features (e.g., leaning forward, sitting on the edge, etc). Of the three types of input examined, the highest recognition accuracy was obtained for posture activity over game status and individual Facial Action Units [121]. Accuracy rates for posture alone as an input modality for recognizing frustrated were not reported in [22]. A potential issue with using a chair to sense body expressions is that the recognition situations are limited to specifically seated contexts. Technologies today are ubiquitous; not limited to only seated situations. Furthermore, as the posture description is dependent on seated postures, important information from the body may be missing. For instance, at the time of their research in 2004, they did not have features to describe the position of the head, hands or feet. More recently however, in 2007, while still employing a posture sensing chair, head position (shown by [40], [80], [134] to be an important feature for discriminating between affective states) and velocity were added to the list of features for the system built to recognize learner frustration [22].

A system by Varni et al [135] focused on the analysis of real-time multimodal affective nonverbal social interaction. The inputs to the system are the same posture and movement features from Camurri et al [113], [128] as well as physiological signals. As opposed to other systems for which direct emotion recognition is the aim, Varni et al's system aims to detect the synchronization of affective behavior and leader-

ship as triggered by emotions. As a first test of the system, violin duos enacted four emotions and neutral in music performances. The percentage of synchronization was highest for pleasure and lowest for anger. This work is interesting as it is one of the first steps towards group emotions (rather than an individual's emotions) and related social phenomena. This is important as technology is used more and more to mediate group collaboration in various contexts.

The automatic recognition system of Gunes and Piccardi [122] is bi-modal, recognizing video sequences of facial expressions and upper-body expressions. They examined the automatic recognition performance of each modality separately before fusing information from the two modalities into a single system. The automatic recognition performance was highest for the upper body sequences, compared to the facial expression sequences. The authors attributed this outcome to the fact that facial movements are much smaller in comparison to the upper body movements, and that even though high resolution video was used, it may not be sufficient enough for perfect recognition. In a more recent implementation of the system using the same database, Gunes and Piccardi [123] exploited temporal dynamics between facial expressions and upper-body gestures to improve the reliability of emotion recognition. Both the temporal phases of an expression and its emotion labels were used to code each modality independently. The apex phases of the emotion for each modality were used to perform a low-level fusion of the features. Interestingly, the best bi-modal classification performances were comparable to the body only classification performances and the bi-modal system outperformed the unimodal system based on facial expressions. The same database was also used by Shan et al [124]. They tested the recognition of individual modalities. In this case, facial expression recognition was slightly better than body expression recognition. The issue with these systems presented is that the expressions were scripted, and therefore the high automatic recognition rates are not surprising. A question that needs to be addressed now is: what happens when spontaneous, unscripted expressions are used?

As evidenced by the results presented throughout Section 5 and listed in Table 3, there are significant variations between the studies such as whether the expressions were acted or spontaneous (non-acted), labeled according to the expresser's intention or observers' agreement, the corpus used, the features computed, context, target affective states and dimensions, testing method, and finally, automatic modeling technique. So many differences make it difficult to compare the system performances properly as recognized by Gunes and Pantic [28]. The following section addresses some of the issues that affect the creation and evaluation of automatic affect recognition systems.

## 6 DISCUSSION

This section discusses some of the issues that emerged from the studies investigated and issues that require consideration for creating affective body expression recognition models. While some issues are general to all affective channels, given the complexity of body expressions, these issues are particularly important to its modeling. First, we discuss

the modeling process itself with a specific focus on the contribution played by form, dynamic and temporal information and propose moving towards systems that are action-independent. Second, how should we establish the ground truth of affective body expressions? We focus specifically on situations in which observers are used to label the expressions; reasoning that self-report is not reliable [22] in most cases and hence the objective ground truth does not exist. This automatically raises a third question: how should such systems be evaluated, i.e., what benchmark should be used when a real ground truth is not available? We conclude with a summary of existing affective body expression corpora and a discussion of how they should be obtained and used.

### 6.1 What should be modelled?

The studies surveyed have shown that the body is an important modality for recognizing affect, independently of the body actions performed (e.g., walking, playing tennis, studying, dancing, standing, gesturing, etc.). Furthermore, these studies show that body expressions associated with the same emotion category do generally share a core set of features independent of the action performed [91]. Hence, given the large variability of possible actions in which the body may be involved, it becomes important to investigate the possibility of an action independent model of how affect is expressed rather than building a recognition system for each type of action. One question that needs to be asked in relation to this challenge is the contribution played by static and dynamic body features; i.e., what type of information should be modelled and how should the features be integrated?

Although many studies (see Section 4.1) have shown the importance of form information with dynamic features used to solve uncertainties, there is also evidence that dynamic features on their own may have a strong discriminative power. In fact, even when form information was disrupted, the recognition of affect from dynamic only features remained above chance level [89]. This means that dynamic information may be not only complementary to form but also partially redundant to it. These observations may indicate that more effort should be dedicated to developing feature extraction algorithms and fusion models that take into account the role that each feature (or combinations of them) plays in the classification process; this being a discriminative, a reinforcing or an inconsistency-resolving role.

To add to this, it is also possible that the role of form and dynamic features may depend not only on the emotions expressed but also on the type of action performed. This raises another issue, i.e., the separation of the temporal relationship between the movement phases characterizing a body action (either cycled or not) and the temporal characteristics of its expressive content. Various studies have in fact shown that kinematic and form-from-motion features are more relevant to discriminate non-instrumental actions (e.g., locomotory actions) rather than instrumental actions (i.e., goal directed) or social actions (e.g., emotional expressions) [89], [136], [137]. Furthermore, Atkinson et al.'s study [137] on autism spectrum disorders shows that emotion recognition seems to depend more on global motion and global form features, whereas non-instrumental, and instrumental actions depend on relatively local motion and form cues. This suggests that

affect recognition systems may benefit from the investigation of feature representation spaces that allow for the separation of affect recognition tasks from idiosyncrasy tasks as well as (non-) instrumental action tasks. In fact, perceptual and neuroscience studies provide evidence for the existence of separate neural structures for the processing of these three tasks [138], [139], [140].

This separation may facilitate the optimization and generalization of the former tasks. Finally, to fully address these challenges, it is important that more systematic studies are carried out using datasets of natural body expressions that go beyond gestures and gait. This will help to understand the role and importance of these features and how they should be modelled to build action-independent affect recognition systems that can easily generalize to different situations.

### 6.2 Establishing the ground truth

When establishing the ground truth of an emotion expression we often refer to observer coders for various reasons. First, discerning what a person really feels when her expression is recorded is not always feasible or reliable [22], [141]. Second, the need to label each modality separately to create more accurately labeled training and testing sets may require post-processing the data without contextual information. Finally, in certain situations, the aim of the recognition system is to model the observer rather than the expresser.

A problem with using observers to build the ground truth is that a high level of disagreement may arise in the decoding of each expression [60], [121]. One way to overcome this is to use expert coders but often this is not feasible or desirable. However, even when high variability is present, a typical approach is to use the 'most frequent label'. Unfortunately, it is generally difficult to obtain a large number of evaluations across a large number of observers [60]; hence there is no statistical rationale to consider the most frequent values as the most probable. Hence, new methods are necessary to measure either the validity of the most frequent label selected or to model the recognition system so that it takes into account the variability between samples.

To address the former issues, Kleinsmith et al. [60] have proposed to create a more complex estimate of how well the observers agree. In their method, the group of observers is split into three groups and the first two groups are used to estimate the level of agreement. The third group is then used to estimate the labels to be used to build the system. The system is then tested against the first group of observers, i.e., the labels set according to that group. The complete process is repeated a number of times to simulate a cross-validation approach using random repeated sub sampling with replacement. They argue that this approach better approximates the variability between humans in recognizing expressed affect as the number of observers recruited in these studies is generally small.

Whereas this approach tries to overcome the issue of smaller sets of observers, there is still the need to overcome the limitation of forcing the attribution of one label to a body expression. Instead, multi-labeling techniques could be adopted. One method being employed in the artificial intelligence [142], [143] and machine learning [144] fields is preference learning. In the field of automatic affect recogni-

tion it is used to construct computational models of affect based on users' preferences. To this aim, observers or expressers are asked to view two stimuli (e.g., two postures) and indicate which stimulus better represents a certain affective state (e.g., happy). This process is repeated for each pair of stimuli. The approach attempts to model the order of preferences instead of an absolute match. It can reduce the noise caused by a forced choice approach in which the observers or expressers are obliged to provide an absolute judgment. Multi-labeling techniques raise the need of applying evaluation metrics that take into account the intrinsic variability internal to each group, such as the frequency of use of each label and the ranking between applied labels [145]. According to the type of application, interesting biases could also be added to these kinds of approaches by combining them with an observer profile based approach in which weights are attached to the labels according to the observer's level of empathy. This is considered important for the recognition of another person's emotional state [146].

To increase the reliability of multi-labeling approaches, crowdsourcing could be seen as a promising and low-cost method. This approach is largely exploited by the information retrieval community to improve the labeling of datasets, and is particularly useful when only a subjective ground truth exists [147]. The idea is that noise could be partially cancelled out over a large number of observers and that probabilistic modeling approaches could be used to weight observers' labeling skills [148]. Although this approach comes with problems [149], [150], it would be an interesting source of information not only for improving the labeling process but also for investigating various contextual factors that affect the perception of affective body expressions.

Contextual factors are indeed very critical to the perception of emotional expressions. For example, Gendron et al [183] have provided evidence of the fact that language shapes the way we interpret emotional expressions. This is supported by cross-cultural studies showing that certain expressions are recognized differently by people of different cultures [80]. Research on embodied cognition (e.g., [185]) has also challenged previous views on conceptual knowledge. According to this view, an emotion category (e.g., happiness) is represented in the sensorimotor cortex. Hence, the perception of an affective expression requires a partial re-enactment of the sensorimotor events associated with that affective state [185]. Following this view, Lindquist et al. [184] argue that the fact that over-exposing observers to a particular emotion-word reduces their ability to recognize prototypical expressions of that emotion (i.e., an after-effect) may be due to the inhibition of the motor system necessary to enact that emotion. It follows that the emotional state of the observers may also bias the perception of another person's expression as it may inhibit or facilitate access to the sensorimotor information necessary to re-enact that expression [184]. Evidence from other studies also shows that such biases may be triggered by the valence associated with the postural stance of the observer (e.g., [186], [187], [14]). Given this evidence, it is critical that factors that may affect the labeling process, such as observer profiling (e.g., empathic skills) and observer contextual factors (e.g., mood, posture) are taken into account when establishing the ground

truth and its validity.

### 6.3 Affective expression corpora

Finally, the question to ask is what type of corpora should be used. An issue surrounding affective expression corpora is whether to use acted or non-acted corpora. Acted affective corpora are signified as actions that have been deliberately and knowingly expressed, whereas non-acted or naturalistic affective corpora are expressions that have been expressed naturally, without intention for the experimental procedure. The longstanding argument about acted vs. non-acted affective corpora concerns the reliability of using acted stimuli for studying emotion/affect perception [71], [40]. The early affective databases were acted or posed and focused on face and voice [151], [152], [153], [154]. The difficulty and necessity of obtaining naturalistic, non-acted stimuli has been discussed for more than two decades, being described as "*one of the perennial problems in the scientific study of emotion*" [155].

Using material obtained from actors who are explicitly instructed to express specific affective states is considered to be unnatural and contrived/artificial [156]. However, Banzinger and Scherer [157] argue that the use of acted affective expressions that have been well-designed can be very useful given the practical and ethical problems of inducing genuine and intense emotions in a lab setting. Castellano et al [158] explain that while an advantage of acted expressions is that they can be clearly defined and enable multiple emotions to be recorded from a single individual; they are not genuine emotions and are typically devoid of context. Furthermore, whereas they may allow for the identification of general features that characterize affective body expressions, they may produce a very narrow understanding of them and of the temporal dynamics between action and emotion content. Studies presented in Section 4 have in fact shown that the same emotion category can be associated with quite different types of expressions (e.g., [96]). Russell [180] considers the prototypical expressions of emotions just a subset of the way we express emotions. Since the expression of an emotional state depends on various components, a larger variety of expressions may be provided in response to a particular stimulus. This implies that studying acted emotions can be useful for the reasons discussed above, but it leads to the creation of datasets and computational models that are very limited in their usefulness in real life situations. The research trend is now on naturally occurring affective expressions [159], [160] (yet, they are still focused mainly on facial expressions). The combination of the two types of data collection should facilitate a more comprehensive and systematic study of affective body expressions.

As described in Section 5, until recently, much of the research on body expressions has focused on dance, often using video recordings of ballets and other dance performances for analysis of affective behavior. This means that research groups aiming to examine more natural, day-to-day affective bodily behaviors are required to create their own corpora. An issue here is that unless the affective corpora are made available for research, the use of different datasets makes it difficult to compare and evaluate systems and methods properly [28]. Moreover, datasets should be described

according to continuous, numerical descriptors as much as possible as this should make the analysis of affective body expressions less prone to subjective interpretation.

While there are several databases of affective facial expression corpora available (e.g., [70], [141], [167], [168], [169], [170], [171], [172], [173]), there are fewer databases that include affective body expressions [80], [130], [161], [100], [162], [163], [99]. Based on the variety of recent research presented throughout this paper, it is apparent that providing databases of affective, whole body postures and movements, acted and non-acted, could reduce (if not eliminate) the time-consuming task of developing a new corpus for each research endeavour. This would allow researchers to focus on the main goal of understanding and automating affect recognition from body expressions.

Furthermore, there is no doubt that the analysis and modeling of affective expressions would strongly benefit from multi-modal data collection. However, a modality that is rarely used but particularly important in body expression analysis is muscle activation. Electromyograms (EMG) have been used for facial expression analysis but rarely for the study of body expressions even though a few medical studies have found evidence of a relationship between patterns of activation in body muscles and emotional states [164], [165], [166]. For example, fear of movement in people with back pain may cause them to freeze their muscles and produce guarded movements. Muscle tension has also been examined by DeMeijer [101] and Gross et al [27]; however, these ratings were based on visual subjective perception from videos. Whereas muscle activation affects the way a movement is performed, unfortunately these effects may not always be easily detected through motion capture systems and/or video cameras. Hence, even if EMG data provides various challenges from a modeling process perspective, they could be valuable pieces of information that may help solve misclassifications between various affective states.

## 7 CONCLUSIONS

The purpose of this paper has been to examine the use of body expressions as an important modality for affective communication and for creating affectively aware technology. To this end, we have reviewed studies from different fields in which affective body expressions have been investigated. Behavioral science research has shown that body expressions are more important for nonverbal communication than was previously thought. Moreover, several perception studies have shown that there are spatio-temporal body features that are responsible for conveying affect and that these features appear to be the same across different types of tasks.

Unfortunately, most of these studies have relied on a limited set of acted body expressions (e.g., dance, gesture, posture, and gait). There is a need to go beyond this focus and investigate the possibility of creating systems that are able to recognize emotions independently of the action the person is doing. This is very important given the high degrees of freedom of the body and given that these systems can be ubiquitously deployed (e.g., a social robot). This requires a more systematic investigation of the types of fea-

tures (e.g., local vs. global) that change the affective message carried out by the body independently from the semantic meaning (action) that it conveys. Furthermore, as the data labeling process is time consuming and highly subjective, it is necessary to move towards crowdsourcing types of labeling processes. This may allow for the generation of labeled datasets for which the reliability and generality can be more accurately computed.

This paper has focused mainly on the analysis and importance of the body alone in conveying affect and has only briefly touched upon the issues of modeling multi-modal expressions. This raises other very important and interesting issues that would require significantly more time and space to be addressed and therefore, have been left for a later publication.

## ACKNOWLEDGMENT

This study has been supported by the Marie Curie International Reintegration Grant "AffectME" (MIRG-CT-2006-046343).

## REFERENCES

- [1] N. Fragopanagos and J.G. Taylor, "Emotion recognition in human-computer interaction," *Neural Net.*, vol. 18, no. 4, pp. 389-405, 2005.
- [2] C.E. Izard, P.B. Ackerman, K.M. Schoff, and S.E. Fine, "Self-organization of discrete emotions, emotion patterns, and emotion-cognition relations," *Emotion, Development, and Self-Organization: Dynamic Systems Approaches to Emotional Development*, D.L. Marc and I. Granic (eds), pp. 15-36, Cambridge Univ. Press, 2002.
- [3] P. Ekman and W. Friesen. *Manual for the facial action coding system*. Consulting Psychology Press, 1978.
- [4] R. von Laban. *The mastery of movement*. MacDonald & Evans Ltd, 1971.
- [5] R.H. Rozensky and L. Feldman-Honor, "Notation systems for coding nonverbal behavior: A review," *J. Behav. Assess.*, vol. 4, no. 2, pp. 119-132, 1982.
- [6] A. Mehrabian and J. Friar, "Encoding of attitude by a seated communicator via posture and position cues," *Journal of Consulting and Clinical Psychology*, vol. 33, pp. 330-336, 1969.
- [7] M. Argyle. *Bodily Communication*. Methuen & Co. Ltd, 1988.
- [8] P.E. Bull. *Posture and Gesture*. Pergamon, 1987.
- [9] P. Ekman and W. Friesen, "Detecting deception from the body or face," *J. Pers. and Soc. Psych.*, vol. 29, no. 3, pp. 288-298, 1974.
- [10] L. McClenney and R. Neiss, "Post-hypnotic suggestion: A method for the study of nonverbal communication," *J. Nonv. Behav.*, vol. 13, pp. 37-45, 1989.
- [11] H. Meeren, C. van Heijnsbergen, and B. de Gelder, "Rapid perceptual integration of facial expression and emotional body language," *Proc. Nat. Acad. Sci. of the USA*, vol. 102, no. 45, pp. 16518-16523, 2005.
- [12] J. Van den Stock, R. Righart, and B. de Gelder, "Body expressions influence recognition of emotions in the face and voice," *Emotion*, vol. 7, no. 3, pp. 487-494, 2007.
- [13] A.N. Antle, G. Corness, M. Droumeva, "What the body knows: Exploring the benefits of embodied metaphors in hybrid physical digital environments," *Interacting with Computers*, vol. 21, pp. 66-75, 2009.
- [14] N. Bianchi-Berthouze, "Understanding the role of body movement in player engagement," *Human-Computer Interaction*, In press. <http://web4.cs.ucl.ac.uk/uclhc/people/n.berthouze/BerthouzeHCI12.pdf>
- [15] F. Muller, N. Bianchi-Berthouze, "Evaluating Exertion Games," *HCI Series, Part 4*, pp. 187-207, 2010.



- [16] P.M. Niedenthal, L.W. Barsalou, P. Winkielman, S. Krauth-Gruber, and F. Ric, "Embodiment in attitudes, social perception, and emotion," *Pers. and Social Psych. Review*, vol. 9, pp. 184–211, 2005.
- [17] L. Barsalou, "Grounded cognition," *Ann. Rev. Psych.*, vol. 59, pp. 617–645, 2008.
- [18] S. Goldin-Meadow, H. Nusbaum, S.D. Kelly, and S. Wagner, "Explaining math: Gesturing lightens the load," *Psych.Sci.* vol. 12, no. 6, pp. 516–522, 2001.
- [19] J. Chandler and N. Schwarz, "How extending your middle finger affects your perception of others: Learned movements influence concept accessibility," *J. Exp. Soc. Psych.*, vol. 45, no. 1, pp. 123–128, 2009.
- [20] S. Neill and C. Caswell. *Body language for competent teachers*. Routledge, 1993.
- [21] D.H. Jonassen and B.L. Grabowski. *Handbook of individual differences, learning, and instruction*. Erlbaum, 1993.
- [22] A. Kapoor, W. Bursleson, and R.W. Picard, "Automatic prediction of frustration," *Int. J. HC Studies*, vol. 65, no. 8, pp. 724–736, 2007.
- [23] A. Kvåle, A.E. Ljunggren, and T.B. Johnsen, "Examination of movement in patients with long-lasting musculoskeletal pain: reliability and validity," *Physiotherapy Res Int* vol. 8, pp. 36–52, 2003.
- [24] G.K. Haugstad, T.S. Haugstad, U.M. Kirste, S. Leganger, S. Wojniusz, I. Klemmetsen, U.F. Malt, "Posture, movement patterns, and body awareness in women with chronic pelvic pain," *J. Psychosomatic Research*, vol. 61, no. 5, pp. 637–644, 2006.
- [25] B. Bunkan, A.E. Ljunggren, S. Opjordsmoen, O. Moen and S. Friis, "What are the dimensions of movement?" *Nord J Psychiatry*, vol. 55, pp. 33–40, 2001.
- [26] J.W.S. Vlaeyen, S.J. Linton, "Fear-Avoidance and its Consequences in Musculoskeletal Pain: A State of the Art," *Pain*, vol. 85, no. 3, pp. 317–332, 2000.
- [27] M.M. Gross, E.A. Crane, B.L. Fredrickson, "Methodology for Assessing Bodily Expression of Emotion," *J Nonv Behav*, vol. 34, pp. 223–248, 2010.
- [28] H. Gunes, M. Pantic, "Automatic, Dimensional and Continuous Emotion Recognition", *Int'l J Synth Emotion*, vol.1, no.1, pp. 68–99, 2010.
- [29] D. McNeill, *Hand and Mind—What Gestures Reveal about Thought*. The University of Chicago Press, 1992.
- [30] B. DeCarolis, C. Pelachaud, I. Poggi, M. Steedman, "APML, a markup language for believable behavior generation," *Life-like Characters: Tools, Affective Functions and Applications*. H. Prendinger, M. Ishizuka (eds), Springer, pp.65–85, 2004.
- [31] J. Cassell, "Body language: lessons from the near-human," *Genesis Redux: Essays on the history and philosophy of artificial life*. J. Riskin, (ed.), pp. 346–374, University of Chicago Press, 2007.
- [32] C. Pelachaud, "Studies on gesture expressivity for a virtual agent," *Speech Commun*, vol. 51, no. 7, pp. 630–639, 2009.
- [33] R. Picard, "Toward agents that recognize emotion," *Proc. of IMAGINA*, pp. 153–165, Springer-Verlag, 1998.
- [34] B. de Gelder, "Why bodies? Twelve reasons for including bodily expressions in affective neuroscience," *Philosophical Transactions of the Royal Society*, vol. 364, no. 3, pp. 3475–3484, 2009.
- [35] W. Zhao, R. Chellappa, A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, vol. 35 pp. 399–458, 2003.
- [36] M. Pantic and L.J.M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424–1445, 2000.
- [37] B. de Gelder, "Towards the neurobiology of emotional body language," *Nature Reviews Neurosci.* vol. 7, no. 3, pp. 242–249, 2006.
- [38] P. Ekman and W. Friesen, "Felt, false and miserable smiles," *Journal of Nonverbal Behavior*, vol. 6, no. 4, pp. 238–252, 1982.
- [39] R. Adolphs, "Neural systems for recognizing emotion," *Current Opinion in Neurobiology*, vol. 12, no. 2, pp. 169–177, 2002.
- [40] H.G. Wallbott, "Bodily expression of emotion," *European Journal of Social Psychology*, vol. 28, pp. 879–896, 1998.
- [41] P. Ekman and W. Friesen, "Head and body cues in the judgment of emotion: A reformulation," *Perc and Motor Skills*, vol. 24, pp. 711–724, 1967.
- [42] P. Ekman and W. Friesen, "The repertoire of non-verbal behavioral categories: Origins, usage and coding," *Semiotica*, vol. 1, pp. 49–98, 1969.
- [43] C. Darwin. *The expression of the emotions in man and animals*. Murray, 1872.
- [44] P. Ekman and W. Friesen, "Nonverbal leakage and clues to deception," *Psychiatry*, vol. 32, pp. 88–105, 1969.
- [45] P. Ekman, "Mistakes When Deceiving," *Annals of the New York Academy of Sciences*, vol. 364, pp. 269–278, 1981.
- [46] J. E. Hocking and D.G. Leathers, "Nonverbal indicators of deception: A new theoretical perspective," *Communication Monographs*, vol. 4, no. 2, pp. 119–131, 1980.
- [47] R. Kraut, "Verbal and Nonverbal Cues in the Perception of Lying," *Journal of Personality and Social Psychology*, vol. 36, no. 4, pp. 380–391, 1978.
- [48] Hocking, J. E. *Detecting deceptive communication from verbal, visual, and paralinguistic cues: An exploratory experiment*. Unpublished doctoral dissertation, Michigan State University, 1976.
- [49] G.E. Littlepage and M.A. Pineault, "Detection of deceptive factual statements from the body and the face," *Personality and Social Psychology Bulletin*, vol. 5, pp. 325–328, 1979.
- [50] R.E. Riggio and H.S. Friedman, "Individual Differences and Cues to Deception," *Journal of Personality and Social Psychology*, vol. 45, no. 4, pp. 899–915, 1983.
- [51] P. Ekman and W. Friesen. *Pictures of facial affect*. Consulting Psychologists Press, 1976.
- [52] N. Hadjikhani and B. de Gelder, "Seeing fearful body expressions activates the fusiform cortex and amygdala" *Curr. Biol.*, vol. 13, pp. 2201–2205, 2003.
- [53] B. de Gelder, J. Snyder, D. Greve, G. Gerard, and N. Hadjikhani, "Fear fosters flight: A mechanism for fear contagion when perceiving emotion expressed by a whole body," *Proc Natl Acad Sci U S A*, vol. 101, no. 47, pp.16701–6, 2004.
- [54] H.M. Paterson, F.E. Pollick, and E. Jackson, "Movement and faces in the perception of emotion from motion," *Perception, ECVF Glasgow Supplemental*, vol. 31, no. 118, pp. 232–232, 2002.
- [55] F.E. Pollick, H. Paterson, and P. Mamassian, "Combining faces and movements to recognize affect," *J Vis*, vol. 4, no. 8, pp. 232–232, 2004.
- [56] R. Tuminello & D. Davidson, "What the face and body reveal: In-group emotion effects and stereotyping of emotion in African-American and European-American children," *Journal of Experimental Child Psychology: Special Issue on Emotion Assessment in Children and Adolescents*, vol. 110, pp. 258–274, 2011
- [57] M.L. Willis, R. Palermo, D. Burke, "Judging Approachability on the Face of It: The Influence of Face and Body Expressions on the Perception of Approachability," *Emotion*, vol. 11, no. 3, pp. 514–523, 2011.
- [58] H.M. Paterson, F.E. Pollick, and A.J. Sanford, "The role of velocity in affect discrimination," *Proc. 23rd Annual Conference of the Cognitive Science Society*, pp. 756–761, Lawrence Erlbaum Associates, 2001.
- [59] A. Kleinsmith and N. Bianchi-Berthouze, "Grounding affective dimensions into posture features," *LNCS: Proc. 1st Int. Conference on Affective Computing and Intelligent Interaction*, pp. 263–270, 2005.
- [60] A. Kleinsmith, N. Bianchi-Berthouze, and A. Steed, "Automatic Recognition of Non-Acted Affective Postures," *IEEE Trans. on Systems, Man, and Cybernetics Part B*, vol. 41, no. 4, pp. 1027–1038, 2011.
- [61] M. Karg, K. Kuhlennz, and M. Buss, "Recognition of affect based on gait patterns," *IEEE Trans on Systems, Man, and Cybernetics, Part B*, vol. 40, no. 4, pp. 1050–1061, 2010.
- [62] J.K. Burgoon, M.L. Jensen, T.O. Meservy, J. Kruse, and J.F. Nuna-maker, "Augmenting human identification of emotional states in video," *Proc. Int. Conference on Intelligent Data Analysis*, 2005.
- [63] D. Matsumoto, "Culture and nonverbal behavior," *Handbook of Non-verbal Comm.*, V. Manusov and M. Patterson (eds), Sage, 2005.
- [64] S. Kopp, L. Gesellensetter, N. Kramer, and I. Wachsmuth, "A conversational agent as museum guide - design and evaluation of a real-world application," *Intelligent Virtual Agents*, J. Tao, T. Tan, and R. Picard (eds), pp. 329–343, Springer-Verlag, 2005.

- [65] W. Swartout, D. Traum, R. Artstein, D. Noren, P. Debevec, K. Bronnenkant, J. Williams, A. Leuski, S. Narayanan, D. Piepol, C. Lane, J. Morie, P. Aggarwal, M. Liewer, J.-Y. Chiang, J. Gerten, S. Chu, K. White, "Ada and Grace: Toward Realistic and Engaging Virtual Museum Guides," *LNAI: Intelligent Virtual Agents*, pp. 286–300, 2010.
- [66] P. Dunn and A. Marinetti, *Cultural adaptation: Necessity for eLearning*. <http://www.linezine.com/7.2/articles/pdamca.htm>, Retrieved January 2006.
- [67] D. Keltner, P. Ekman, G. C. Gonzaga, and J. Beer, "Facial expression of emotion. *Handbook of Affective Sciences*, R. Davidson, K. Scherer, and H. Goldsmith (eds), Oxford University Press, 2003.
- [68] H. A. Elfenbein, M. K. Mandal, N. Ambady, S. Harizuka, and S. Kumar, "Cross-cultural patterns in emotion recognition: Highlighting design and analytical techniques," *Emo*, vol. 2, no. 1, pp. 75–84, 2002.
- [69] P. Ekman, "Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique," *Psych. Bull.*, vol. 115, pp. 268–287, 1994.
- [70] L. Yin, X. Wei, Y. Sun, J. Wang, and M.J. Rosato, *The Binghamton University 3D Facial Expression Database*. <http://www.cs.binghamton.edu/>
- [71] J.A. Russell, "Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies," *Psychological Bulletin*, vol. 115, pp. 102–141, 1994.
- [72] B. Mesquita, "Emotions as dynamic cultural phenomena," *Handbook of Affective Sciences*, R. Davidson, K. Scherer, and H. Goldsmith (eds.), Oxford University Press, 2003.
- [73] H. Elfenbein, N. Ambady, "On the universality and cultural specificity of emotion recognition: A metaanalysis," *Psychological Bulletin*, vol. 128, pp. 205–235, 2002.
- [74] L. Ducci, L. Arcuri, T. Georgis, and T. Sineshaw, "Emotion recognition in Ethiopia: The effects of familiarity with Western culture on accuracy of recognition," *J Cross-Cultural Psych*, vol. 13, pp. 340–351, 1982.
- [75] U. Hess, S. Senecal, and G. Kirouac, "Recognizing emotional facial expressions: Does perceived sociolinguistic group make a difference?" *Int'l Journal of Psychology*, vol. 31, pp. 93, 1996.
- [76] T. Kudoh and D. Matsumoto, "Cross-cultural examination of the semantic dimensions of body postures," *J. Pers. Social Psych.*, vol. 48, no. 6, pp. 1440–1446, 1985.
- [77] D. Matsumoto and T. Kudoh, "Cultural similarities and differences in the semantic dimensions of body postures," *Journal of Nonverbal Behavior*, vol. 11, no. 3, pp. 166–179, 1987.
- [78] M.H. Bond, H. Nakazato, and D. Shiraishi, "Universality and distinctiveness in dimensions of Japanese person perception," *Journal of Cross-Cultural Psychology*, vol. 6, pp. 346–357, 1975.
- [79] D. Matsumoto and H. Kishimoto, "Developmental characteristics in judgments of emotion from nonverbal vocal cues," *International Journal of Intercultural Relations*, vol. 7, pp. 415–424, 1983.
- [80] A. Kleinsmith, P.R. de Silva, and N. Bianchi-Berthouze, "Cross-cultural differences in recognizing affect from body posture," *Interacting with Computers*, vol. 18, pp. 1371–1389, 2006.
- [81] M. Brandt, J. Boucher, "Concepts of depression in emotion lexicons of eight cultures," *Int. J. Intercult. Relat.*, vol. 10, pp. 321–346, 1986.
- [82] D. Matsumoto, T. Consolacion, H. Yamada, R. Suzuki, B. Franklin, S. Paul, R. Ray, H. Uchida, "American-japanese cultural differences in judgments of emotional expressions of different intensities," *Cognition and Emotion* vol. 16, pp. 721–747, 2002.
- [83] M.A. Giese and T. Poggio, "Neural mechanisms for the recognition of biological movements," *Neuroscience*, vol. 4, pp. 179–191, 2003.
- [84] L.M. Vania, M. Lemay, D.C. Bienfang, A.Y. Choi, and K. Nakayama, "Intact biological motion and structure from motion perception in a patient with impaired motion mechanisms: A case study," *Visual Neuroscience*, vol. 5, pp. 353–369, 1990.
- [85] J. Lange and M. Lappe, "The role of spatial and temporal information in biological motion perception," *Advances in Cognitive Psychology*, vol. 3, no. 4, pp. 419–428, 2007.
- [86] M. Hirai and K. Hiraki, "The relative importance of spatial versus temporal structure in the perception of biological motion: An event-related potential study," *Cognition*, vol. 99, pp. B15–B29, 2006.
- [87] M.V. Peelen, A.J. Wiggett, and P.E. Downing, "Patterns of fMRI activity dissociate overlapping functional brain areas that respond to biological motion," *Neuron*, vol. 49, pp. 815–822, 2006.
- [88] P. McLeod, W. Dittrich, J. Driver, D. Perret, and J. Zihl, "Preserved and impaired detection of structure from motion by a "motion-blind" patient," *Visual Cognition*, vol. 3, pp. 363–391, 1996.
- [89] A.P. Atkinson, W.H. Dittrich, A.J. Gemmell, and A.W. Young, "Evidence for distinct contributions of form and motion information to the recognition of emotions from body gestures," *Cognition*, vol. 104, pp. 59–72, 2007.
- [90] L. Omlor and M.A. Giese, "Extraction of spatio-temporal primitives of emotional body expressions," *Neurocomputing*, vol. 70, no. 10–12, pp. 1938–1942, 2007.
- [91] C. Roether, L. Omlor, A. Christensen, M.A. Giese, "Critical features for the perception of emotion from gait," *Journal of Vision*, vol. 8, no. 6:15, pp. 1–32, 2009.
- [92] J. Aronoff, B.A. Woike, and L.M. Hyman, "Which are the stimuli in facial displays of anger and happiness? Configurational bases of emotion recognition," *J Pers and Soc Psych*, vol. 62, pp.1050–1066, 1992.
- [93] J.A. Harrigan and R. Rosenthal, "Physicians head and body positions as determinants of perceived rapport," *Journal of Applied Social Psychology*, vol. 13, no. 6, pp. 496–509, 1983.
- [94] A. Mehrabian, "Inference of attitude from the posture, orientation, and distance of a communicator," *Journal of Consulting and Clinical Psychology*, vol. 32, pp. 296–308, 1968.
- [95] W.T. James, "A study of the expression of bodily posture," *Journal of General Psychology*, vol. 7, pp. 405–437, 1932.
- [96] S. Dahl and A. Friberg, "Visual perception of expressiveness in musicians' body movements," *Music Perception*, vol. 24, no. 5, pp. 433–454, 2007.
- [97] G. Castellano, M. Mortillaro, A. Camurri, G. Volpe, and K. Scherer, "Automated analysis of body movement in emotionally expressive piano performances," *Music Perc*, vol. 26, no. 2, pp. 103–120, 2008.
- [98] D. Glowinski, N. Dael, A. Camurri, G. Volpe, M. Mortillaro, K. Scherer, "Towards a Minimal Representation of Affective Gestures," *IEEE Trans. on Affective Computing*, vol.2, no. 2, pp. 106–118, 2011.
- [99] T. Banziger and K.R. Scherer, "Chapter Blueprint for Affective Computing: A Sourcebook," *Introducing the Geneva Multimodal Emotion Portrayal Corpus*, pp. 271–294, Oxford Univ. Press, 2010.
- [100] M. Coulson, "Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence," *Journal of Nonverbal Behavior*, vol. 28, pp. 117–139, 2004.
- [101] M. de Meijer, "The contribution of general features of body movement to the attribution of emotions," *Journal of Nonverbal Behavior*, vol. 13, pp. 247–268, 1989.
- [102] R. De Silva and N. Bianchi-Berthouze, "Modeling human affective postures: An information theoretic characterization of posture features," *J. Comp Anim and Virtual Worlds*, vol. 15, no.3–4, pp. 269–276, 2004.
- [103] A. Kleinsmith, N. Bianchi-Berthouze, "Recognizing Affective Dimensions from Body Posture," *Proc Int. Conf. of Affective Computing and Intelligent Interaction*, LNCS 4738, pp. 48–58, 2007.
- [104] R.A. Calvo and S. D'Mello, "Affect Detection: An Interdisciplinary Review of Models, Methods, and their Applications", *IEEE Trans. on Affective Computing*, vol. 1, no. 1, pp. 18–37, 2010.
- [105] H. Gunes, M. Piccardi, M. Pantic, "From the Lab to the Real World: Affect Recognition using Multiple Cues and Modalities", *Affective Computing: Focus on Emotion Expression, Synthesis, and Recognition*, J. Or (ed.), pp. 185–218, 2008.
- [106] M. Pantic and L.J.M. Rothkrantz, "Automatic analysis of facial expressions: The state of the art," *IEEE Trans on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1424–1445, 2000.
- [107] M.S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Recognizing facial expression: Machine learning and application to spontaneous behavior," *Proc. Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 568–573, 2005.

- [108] M. Rosenblum, Y. Yacoob, and L.S. Davis, "Human emotion recognition from motion using a radial basis function network architecture," *Proc. Workshop on Motion of Non-rigid & Articulated Objects*, 1994.
- [109] P.Y. Oudeyer, "The production and recognition of emotions in speech: Features and algorithms," *International Journal of Human-Computer Studies*, vol. 59, pp. 157-183, 2003.
- [110] V. Kostov and S. Fukuda, "Emotion in user interface, voice interaction system," *IEEE Int'l Conf on Systems, Man, and Cybernetics*, pp. 798-803, 2007.
- [111] C. Lee, S. Narayanan, and R. Pieraccini. Classifying emotions in human-machine spoken dialogs. In *Proceedings of the International Conference on Multimedia and Expo*, 2002.
- [112] V. Petrushin. Emotion in speech: Recognition and application to call centers. *Proc. Artif. Neural Networks in Engineering*, pp. 7-10, 1999.
- [113] A. Camurri, B. Mazarino, M. Ricchetti, R. Timmers, and G. Volpe, "Multimodal analysis of expressive gesture in music and dance performances," *Gesture-based Communication in HCI*, pp. 20-39, 2004.
- [114] A. Kapur, A. Kapur, N. Virji-Babul, G. Tzanetakis, and P.F. Driessen, "Gesture-based affective computing on motion capture data," In *Proc First Int Conf on Aff Comp and Intelligent Interaction*, pp. 1-7, 2005.
- [115] F.E. Pollick, V. Lestou, J. Ryu, and S-B. Cho, "Estimating the efficiency of recognizing gender and affect from biological motion," *Vision Research*, vol. 42, pp. 2345-2355, 2002.
- [116] D. Bernhardt and P. Robinson, "Detecting affect from non-stylised body motions," *LNCS: Procs 2nd Int Conf on Affective Computing and Intelligent Interaction*, pp. 59-70, 2007.
- [117] N. Bianchi-Berthouze and A. Kleinsmith, "A categorical approach to affective gesture recognition," *Conn Sci.* vol. 15, pp. 259-269, 2003.
- [118] N. Savva, N. Bianchi-Berthouze, "Automatic recognition of affective body movement in a video game scenario," *International Conference on Intelligent Technologies for interactive entertainment*, 2011
- [119] L. Gong, T. Wang, C. Wang, F. Liu, F. Zhang, and X. Yu, "Recognizing affect from non-stylized body motion using shape of Gaussian descriptors," *Proc. of ACM Symposium on Applied Computing*, 2010.
- [120] D. Bernhardt, P. Robinson, "Detecting Emotions from Connected Action Sequences, Visual Informatics: Bridging Research and Practice," *LNCS: 5857*, pp. 1-11, 2009.
- [121] A. Kapoor, R.W. Picard, and Y. Ivanov, "Probabilistic combination of multiple modalities to detect interest," *Proc 17th Int Conf on Pattern Recognition*, vol. 3, pp. 969-972, 2004.
- [122] H. Gunes and M. Piccardi. Bi-modal emotion recognition from expressive face and body gestures," *Journal of Network and Computer Applications*, vol. 30, pp. 1334-1345, 2007.
- [123] H. Gunes, M. Piccardi, "Automatic Temporal Segment Detection and Affect Recognition From Face and Body Display," *IEEE Transon Systems, Man, and Cybernetics, Part B*, vol. 39, no.1, pp. 64-84, 2009.
- [124] CF. Shan, S.G. Gong, P.W. McOwan, "Beyond Facial Expressions: Learning Human Emotion from Body Gestures," *Proc. of the British Machine Vision Conference*, 2007.
- [125] H. Park, J. Park, U. Kim, and W. Woo, "Emotion recognition from dance image sequences using contour approximation," *LNCS: Proc. Int'l Wkshp on Structural, Syntactic, and Stat Patt Rec*, pp. 547-555, Springer-Verlag, 2004.
- [126] A. Camurri, R. Trocca, and G. Volpe, "Interactive systems design: A KANSEI-based approach," *Proc Conf on New Interfaces for Musical Expression*, pp. 1-8, 2002.
- [127] S. Kamisato, S. Odo, Y. Ishikawa, and K. Hoshino, "Extraction of motion characteristics corresponding to sensitivity information using dance movement," *Journal of Advanced Computational Intelligence and Intelligent Informatics*, vol. 8, no. 2, pp. 167-178, 2004.
- [128] A. Camurri, I. Lagerlof, and G. Volpe, "Recognizing emotion from dance movement: Comparison of spectator recognition and automated techniques," *Int. J. of Human-Computer Studies*, vol. 59, no. 1-2, pp. 213-225, 2003.
- [129] R. von Laban, *Modern educational dance*. MacDonald & Evans, Ltd., 1963.
- [130] Ma, H.M. Paterson, and F.E. Pollick, "A motion capture library for the study of identity, gender, and emotion perception from biological motion," *Behavior Res Methods*, vol. 38, no. 1, pp. 134-141, 2006.
- [131] F.E. Pollick, H.M. Paterson, A. Bruderlin, and A.J. Sanford, "Perceiving affect from arm movement," *Cognition*, vol. 82, pp.51-61, 2001.
- [132] J. Sanghvi, G. Castellano, I. Leite, A. Pereira, P.W. McOwan, A. Paiva, "Automatic analysis of affective postures and body motion to detect engagement with a game companion," *Proc. of Int. Conf. on Human-Robot Interaction*, 2011.
- [133] A. Kapoor, S. Mota, and R.W. Picard, "Towards a learning companion that recognizes affect," *Tech Report 543*, MIT Media Lab, 2001.
- [134] R. el Kaliouby and P. Robinson, "Real-time inference of complex mental states from facial expressions and head gestures," *Proc. IEEE Int. Workshop on Real Time Computer Vision for Human Computer Interaction at CVPR*, 2004.
- [135] G. Varni, G. Volpe, and A. Camurri, "A system for real-time multimodal analysis of nonverbal affective social interaction in user-centric media," *IEEE Trans on Multimedia*, vol. 12, no. 6, pp. 576-590, 2010.
- [136] W.H. Dittrich, "Action categories and the perception of biological motion," *Perception*, vol. 22, no. 1, pp. 15-22, 1993.
- [137] A.P. Atkinson, "Impaired recognition of emotions from body movements is associated with elevated motion coherence thresholds in autism spectrum disorders," *Neuropsychologia*, vol. 47, no. 13, pp. 3023-3029, 2009
- [138] H.L. Gallagher and C.D. Frith, "Dissociable neural pathways for the perception and recognition of expressive and instrumental gestures," *Neuropsychologia*, vol. 42, no. 13, pp. 1725-1736, 2004.
- [139] A.J. Calder, A.M. Burton, P. Miller, A.W. Young, S. Akamatsu, "A principal component analysis of facial expressions," *Vision Research*, vol. 41, no. 9, pp. 1179-1208, 2001.
- [140] U. Martens, H. Leuthold and S.R. Schweinberger, "On the temporal organization of facial identity and expression analysis: Inferences from event-related brain potentials," *Cognitive, Affective, & Behavioral Neuroscience*, vol. 10, no. 4, 505-522, 2010.
- [141] P. Ekman and W.V. Friesen, *Pictures of Facial Affect*. <http://www.paulekman.com/researchproducts.phpt>, Retrieved October 2008.
- [142] J. Fürnkranz and E. Hüllermeier, "Preference learning," *Kunstliche Intelligenz*, vol. 19, no. 1, pp. 60-61, 2005.
- [143] J. Doyle, "Prospects for preferences," *Computational Intelligence*, vol. 20, no. 2, pp. 111-136, 2004.
- [144] G.N. Yannakakis, "Preference learning for affective modelling," *Proc 3rd Int'l Conf on Aff Comp and Intell Interaction*, pp. 126-131, 2009.
- [145] H. Meng, A. Kleinsmith, N. Bianchi-Berthouze, "Multi-score Learning for Affect Recognition: the Case of Body Postures," *Int'l Conf on Aff Comp and Intel. Interaction*, In press.
- [146] A. Mehrabian and N. Epstein, "A measure of emotional empathy," *Journal of Personality*, vol. 40, pp. 525-543, 1972.
- [147] V. S. Sheng, F. Provost, and P. G. Ipeirotis, "Get another label? Improving data quality and data mining using multiple, noisy labellers," *Proc of the 14th ACM SIGKDD Int'l Conf on Knowledge Disc and Data Mining*, pp. 614-622, 2008.
- [148] V. Raykar, S. Yu, L. Zhao, G. Valadez, C. Florin, L. Bogoni, and L. Moy, "Learning from crowds," *Journal of Machine Learning Research*, vol. 11, no. 7, pp. 1297-1322, 2010.
- [149] J. Howe, *Crowd sourcing: Why the Power of the Crowd Is Driving the Future of Business*. 2008.
- [150] A. Tarasov, C. Cullen, S.J. Delany, "Using crowdsourcing for labelling emotional speech assets," *W3C Wkshp Emo Markup Lang*, 2010.
- [151] P. Juslin and P. Laukka, "Communication of emotions in vocal expression and music performance," *Psych. Bulletin*, vol. 129, no. 5, pp. 770-814, 2002.

- [152] M. Kienast and W.F. Sendlmeier, "Acoustical analysis of spectral and temporal changes in emotional speech," *Speech and emotion: Proc ISCA workshop*, pp. 92-97, 2000.
- [153] L. Leinonen and T. Hiltunen, "Expression of emotional-motivational connotations with a one-word utterance," *Journal of the Acoustical Society of America*, vol. 102, no. 3, pp. 1853-1863, 1997.
- [154] S. Yacoub, S. Simske, X. Lin, and J. Burns, "Recognition of emotions in interactive voice response systems," *Proc. Eurospeech*, 2003.
- [155] H.G. Wallbott and K.R. Scherer, "Cues and channels in emotion recognition," *J. Pers Social Psych.*, vol. 51, no. 4, pp. 690-699, 1986.
- [156] N. Sebe, Y. Sun, E. Bakker, M.S. Lew, I. Cohen, and T.S. Huang, "Towards authentic emotion recognition," *IEEE Int'l Conf on Systems, Man and Cybernetics*, pp. 623-628, 2004.
- [157] T. Bänziger, K. R. Scherer, "Using Actor Portrayals to Systematically Study Multimodal Emotion Expression: The GEMEP Corpus," *Proc. Int'l Conf Aff Comp Intell Interaction*, pp. 476-487, 2007.
- [158] G. Castellano, I. Leite, A. Pereira, C. Martinho, A. Paiva, and P. McOwan, "Affect recognition for interactive companions: Challenges and design in real-world scenarios," *Journal on Multimodal User Interfaces*, vol. 3, no. 1-2, pp. 89-98, 2010.
- [159] A. Ashraf, S. Lucey, J. Cohn, T. Chen, K. Prkachin, and P. Solomon, "The painful face: Pain expression recognition using active appearance models," *Image and Vision Comp*, vol. 27, pp. 1788-1796, 2009.
- [160] S. Afzal and P. Robinson, "Natural affect data - collection and annotation in a learning context," *Proc 3rd Int Conf on Affective Computing and Intelligent Interaction*, pp. 22-28, 2009.
- [161] N. Dael, M. Mortillaro, and K.R. Scherer, "Introducing the Body Action and Posture coding system (BAP): Development and reliability," *Manuscript submitted for publication*.
- [162] H. Gunes, M. Piccardi, "A Bimodal Face and Body Gesture Database for Automatic Analysis of Human Nonverbal Affective Behavior," *ICPR*, pp. 1148-1153, 2006.
- [163] M. Karg, R. Jenke, W. Seiberl, K. Kuhnlenz, A. Schwirtz, and M. Buss, "A comparison of PCA, KPCA, and LDA for feature extraction to recognize affect in gait patterns," *Proc In'l Conf Aff Comp Intell Interaction*, pp. 195-200, 2009.
- [164] P.J. Watson, C.K. Booker, C.J. Main, A.C. Chen, "Surface electromyography in the identification of chronic low back pain patients: The development of the flexion relaxation ratio," *Clinical Biomechanics*, vol. 12, no. 3, pp. 165-71, 1997.
- [165] M. Pluess, A. Conrad, F.H. Wilhelm, "Muscle tension in generalized anxiety disorder: A critical review of the literature," *Journal of Anxiety Disorders*, vol. 23, no. 1, pp. 1-11, 2009.
- [166] M.E. Geisser, A.J. Haig, A.S. Wallbom, E.A. Wiggert, "Pain-related fear, lumbar flexion, and dynamic EMG among persons with chronic musculoskeletal low back pain," *Clin J Pain*, vol. 20, no. 2, pp. 61-9, 2004.
- [167] M. Kamachi, M. Lyons, and J. Gyoba, *The Japanese Female Facial Expression Database*. <http://www.kasrl.org/jaffe.html>, Retrieved October 2008.
- [168] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," *IEEE Int'l Conf Multimedia and Expo*, 2005.
- [169] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, *The MMI Face Database*. <http://www.mmifacedb.com/>, Retrieved October 2008.
- [170] S. Baron-Cohen, O. Golan, S. Wheelwright, and J. Hill, *The Mindreading DVD*. <http://www.jkp.com/mindreading/>, Retrieved October 2008.
- [171] *The Psychological Image Collection at Stirling*. <http://pics.psych.stir.ac.uk/>, Retrieved October 2008.
- [172] Center for the Study of Emotion and Attention. *The International Affective Picture System*. <http://csea.phhp.u.edu/media/iapsmessage.html>, Retrieved Oct. 2008.
- [173] T. Kanade, J.F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," *Proc IEEE Int'l Conf Automatic Face and Gesture Recognition*, pp. 46-53, 2000.
- [174] J.K. Burgoon, M.L. Jensen, T.O. Meservy, J. Kruse, J.F. Nunamaker Jr., "Augmenting Human Identification of Emotional States in Video," *Proc. Int'l Conf on Intelligent Data Analysis*, 2005.
- [175] P. Ekman, "Lying and deception," In Stein, N.L., Ornstein, P.A., Tversky, B., & Brainerd, C. *Memory for Everyday and Emotional Events*. Lawrence Erlbaum Associates, 1997.
- [176] B.M. DePaulo, J.J. Lindsay, B.E. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper, "Cues to deception," *Psychological Bulletin*, vol. 129, pp.74-112, 2003.
- [177] H. Gunes and M. Piccardi, "Observer annotation of affective display and evaluation of expressivity: Face vs. face-and-body," *Proc HCSNet Workshop on Use of Vision in HCI*, pp. 35-42, 2006.
- [178] C. Clavel, J. Plessier, J-C. Martin, L. Ach, and B. Morel, "Combining facial and postural expressions of emotions in a virtual character," *LNCS: Proc 9th Int'l Conf on Intell Virtual Agents*, pp. 387-300, 2009.
- [179] D.A. Leopold, A.J. O'Toole, T. Vetter, and V. Blanz, "Prototype-referenced shape encoding revealed by high-level aftereffects," *Nature Neuroscience*, vol. 4, pp. 89-94, 2001.
- [180] J.A. Russell, "Core affect and the psychological construction of emotion," *Psychological Review*, vol. 110, pp. 145-172, 2003.
- [181] M. Pasch, N. Bianchi-Berthouze, B. van Dijk, B., and A. Nijholt, "Movement-based Sports Video Games: Investigating Motivation and Gaming Experience," *Ent Comp*, vol. 9, no. 2, pp. 169-180, 2009.
- [182] J. Nijhar, N. Bianchi-Berthouze, G. Boguslawski, "Does Movement Recognition Precision affect the Player Experience in Exertion Games?" *Int'l Conf on Intelligent Technologies for Interactive Entertainment*, In press.
- [183] M. Gendron, K.A. Lindquist, L. Barsalou, and L.F. Barrett, "Emotion words shape emotion percepts," *Emotion*, In press.
- [184] K.A. Lindquist, L.F. Barrett, E. Bliss-Moreau, and J.A. Russell, "Language and the perception of emotion," *Emo*, vol. 6, pp. 125-138, 2006.
- [185] L.W. Barsalou, "Perceptual symbol systems," *Behavioral & Brain Sciences*, vol. 22, pp. 577-660, 1999.
- [186] P.M. Niedenthal, L.W. Barsalou, P. Winkelman, S. Krauth-Gruber, and F. Ric, "Embodiment in Attitudes, Social Perception, and Emotion," *Pers and Social Psych Rev*, vol. 9, no. 3, pp. 184-211, 2005.
- [187] J. Chandler and N. Schwarz, "How extending your middle finger affects your perception of others: Learned movements influence concept accessibility," *Journal of Experimental Social Psychology*, vol. 45, no. 1, pp. 123-128, 2009.

**Andrea Kleinsmith** received a B.A. degree in psychology from the University of Oregon, USA in 1995, a MSc. Degree in computer science and engineering from the University of Aizu, Japan in 2004 and a PhD degree in computer science in the area of affective computing from UCL, UK in 2010. She is currently a post-doctoral researcher in the Department of Computing at Goldsmiths, University of London, UK investigating performance based expressive virtual characters. Dr. Kleinsmith's main research interests are in affective human computer interaction and automatic affect recognition systems with a focus on body posture and movement.

**Nadia Bianchi-Berthouze** received the Laurea with Honours in computer science in 1991 and her PhD degree in science of biomedical images in 1996 from the University of Milano, Italy. From 1996-2000, she was a post-doctoral fellow at the Electrotechnical Laboratory of Tsukuba (Japan), when she then became Lecturer at the University of Aizu (Japan). Since 2006, she has been at UCL. She is currently a Senior Lecturer at the UCL Interaction Centre, University College London, UK. Her current research focuses on studying body movement as a medium to induce, recognize, and measure the quality of experience of humans interacting and engaging with/through whole-body technology. In 2006, she was awarded an EU FP6 International Marie Curie Reintegration Grant to investigate the above issues in the clinical and entertainment contexts. Dr Bianchi-Berthouze is member of the editorial board of International Journal of Creative Interfaces and Computer Graphics Journal on Multimodal User Interfaces.



TABLE 1

| Refs                  | Affective states or dimensions each study examined   | Posture/ Movement | Acted/ Non-acted | Percep Study obs. | Stimuli             | No of samples | Ground truth | Features  | Feature details   |
|-----------------------|--|-------------------|------------------|-------------------|---------------------|---------------|--------------|---|---|
| Atkinson et al [89]   | (5) Anger, disgust, fear, happiness, sadness   | Movement          | Acted            | 32                | Patch- & full-light | 60&60         | Actor        | Biological movement   | Upright, upside-down, forward-moving, reversed  |
| Aronoff et al [92]    | (2) Warm, threatening  | Both              | Acted            | 6                 | Video               | Not reported  | Actor        | Diagonal poses:<br>Arabesque:<br>Arms:<br>Movement:         | -<br>-<br>% round, % straight, %angular<br>% round, % straight, %angular  |
| James [95]            | Not defined  | Posture           | Acted            | 3                 | Photos              | 347           | Observer     | Head, trunk, feet, knees, arms                              |   |
| Dahl & Friberg [96]   | (4) Angry, fear, happy, sad  | Movement          | Acted            | 20                | Video               | 32            | Actor        | Amount:<br>Speed:<br>Fluency:<br>Regularity:                | None – large<br>Slow – fast<br>Jerky – smooth<br>Irregular – regular  |
| Castellano et al [97] | (5) Emotion intentions: personal, sad, allegro, serene, overexpressive   | Movement          | Acted            | --                | Video               | 75            | Actor        | Quality of motion of upper body, velocity of head movements |   |
| Gross et al [27]      | (7) Angry, anxious, content, joyful, proud, sadness, neutral   | Movement          | Acted            | 35                | Video               | 42            | Actor        | Torso:<br>Limb:<br>Space:<br>Energy:<br>Time:<br>Flow:      | Contracted, bowed, shrinking - Expanded, stretched, growing<br>Moves close to body, contracted - Moves close to body, expanded<br>Indirect, wandering, diffuse - Direct, focused, channelled<br>Light, delicate, buoyant - Strong, forceful, powerful<br>Sustained, leisurely, slow - Sudden, hurried, fast<br>Free, relaxed, uncontrolled - Bound, tense, controlled   |
| Glowinski et al [98]  | (12) Amusement, anxiety, cold/hot anger, despair, elation, interest, panic, pleasure, pride, relief, sad                           | Movement          | Acted            |                   | Video               | 120           |              | Head & hands:   | Velocity and acceleration, energy, spatial extension, smoothness/jerkiness, Symmetry, forward/backward leaning (head)   |
| Wallbott [40]         | (14) Elated joy, happiness, sadness, despair, fear, terror, cold anger, hot anger, disgust, contempt, shame, guilt, pride, boredom | Posture           | Acted            | 14                | AV                  | 224           | Actor        | Upper body:<br>Shoulders:<br>Head:<br>Arms:<br><br>Hands:   | Away, collapsed<br>Up, backward, forward<br>Downward, backward, turned sideways, bent sideways<br>Lateralized hand/arm movements, stretched out frontal, stretched out sideways, crossed in front of chest, crossed in front of belly, before belly, stemmed to hips<br>Fists, opening/closing, back of hands sideways, emblem, self-manipulator, illustrator, pointing |

|                              |   |          |           |     |            |     |          |  |  |
|------------------------------|---|----------|-----------|-----|------------|-----|----------|--|--|
| Coulson [100]                | (6) Anger, disgust, fear, happiness, sadness  | Posture  | CG        | 61  | Images     | 528 | Observer | Abdomen twist, chest bend, head bend, shoulder swing, shoulder adduct/abduct, elbow bend, weight transfer            | A low-level approach using joint angles. Two different degrees for each feature except weight transfer. Degrees used are dependent on emotion label.                         |
| De Meijer [101]              | (9 & 3+) Joy, grief, anger, fear, surprise, disgust, interest, shame, contempt, sympathy, antipathy, admiration | Movement | Acted     | 85  | Video      | 96  | Observer | Trunk movement:<br>Arm movement:<br>Vertical direction:<br>Sagittal direction:<br>Force:<br>Velocity:<br>Directness: | Stretching - bowing<br>Opening - closing<br>Upward - downward<br>Forward - backward<br>Strong (muscles tensed) – light (muscles relaxed)<br>Fast - slow<br>Direct - indirect |
| De Silva & Berthouze [102]   | (4) Angry, fear, happy, sad   | Posture  | Acted     | 109 | Images     | 109 | Actor    | Head, shoulders, elbows, hands, heels  | The lateral, frontal and vertical extension of the upper body, body torsion, and the inclination of the head and shoulders   |
| Kleinsmith & Berthouze [103] | (4) Arousal, valence, potency, avoidance  | Posture  | Acted     | 5   | Images     | 111 | Observer | Head, shoulders, elbows, hands, heels  | The lateral, frontal and vertical extension of the upper body, body torsion, and the inclination of the head and shoulders   |
| Roether et al [91]           | (4) Angry, fear, happy, sad   | Both     | Acted     | 21  | Animations | 388 | Actor    | Head, neck, spine, and right and left clavicle, shoulder, elbow, wrist, hip, knee and ankle                          | Joint angles (Euler) around the flexion, abduction, rotation axes  |
| Kleinsmith et al [60]        | (4) Concentrating, defeated, frustrated, triumphant   | Posture  | Non-acted | 8   | Images     | 103 | Observer | Head, neck, collar, shoulders, elbows, wrists, torso, hips and knees   | Joint angles (Euler) around the flexion, abduction, rotation axes  |

TABLE 2

| Aff states/dims      | Study                   | Discriminating features  |
|----------------------|-------------------------|--|
| <b>Anger</b>         | Coulson [100]           | Head bent back, no backward chest bend, no abdominal twist, arms raised forward & upward   |
|                      | De Meijer [101]         | Bowed trunk & head, knees slightly bent, slow velocity, strong force, downward body movement, stepping backward, arms open frontally |
|                      | Dahl & Friberg [96]     | Both instruments: Large, very jerky, somewhat fast movements   |
|                      | Gross et al [27]        | High energy, expanded limbs, tense and controlled flow   |
|                      | Kleinsmith et al [80]   | Head bent forward, elbows bent and laterally extended  |
|                      | Roether et al [91]      | Head bent forward, elbows bent   |
| <b>Cold anger</b>    | Wallbott [40]           | Lateralized hand/arm movements, arms stretched out frontal   |
| <b>Hot anger</b>     | Wallbott [40]           | Shoulders lifted, lateralized hand/arm movements, arms stretched out frontal   |
| <b>Anxious</b>       | Gross et al [27]        | Low energy, slow movement, somewhat expanded limbs and torso   |
| <b>Arousal</b>       | Kleinsmith et al [103]  | Low-arousal: head bent forward, hands close to the body. High arousal: head bent backward, hands vertically extended                 |
| <b>Avoidance</b>     | Kleinsmith et al. [103] | Vertical extension and lateral opening of the body for high avoidance  |
| <b>Boredom</b>       | Wallbott [40]           | Collapsed upper body, head bent backwards  |
| <b>Contempt</b>      | De Meijer [101]         | Bowed trunk & head, knees slightly bent, stepping backward   |
| <b>Concentrating</b> | Kleinsmith et al [60]   | Shoulders slumped forward, the arms extended down and diagonally across the body   |
| <b>Defeated</b>      | Kleinsmith et al [60]   | Shoulders slumped forward, the arms extended down and diagonally across the body   |
| <b>Despair</b>       | Wallbott [40]           | Shoulders forward  |
| <b>Disgust</b>       | De Meijer [101]         | Bowed trunk & head, knees slightly bent  |
|                      | Wallbott [40]           | Shoulders forward, head bent forward, arms crossed in front  |
| <b>Fear</b>          | Coulson [100]           | Backward head bend, no abdominal twist, forearms raised, weight shift backward   |
|                      | De Meijer [101]         | Bowed trunk & head, knees slightly bent, downward, backward fast movement, muscles tensed  |
|                      | Dahl & Friberg [96]     | Saxophone: regular, smooth & slow movements. Bassoon: jerky & somewhat fast movements  |
|                      | Kleinsmith et al [100]  | Head straight up or bent back slightly, elbows bent, arms lateral  |
|                      | Roether et al [91]      | Head upright, elbows bent  |
|                      | Wallbott [40]           | Shoulders forward  |
| <b>Frustrated</b>    | Kleinsmith et al [60]   | Shoulders straight up or back, arms raised and extended laterally  |
| <b>Grief</b>         | De Meijer [101]         | Bowed trunk & head, knees slightly bent, slow velocity, downward body movement, arms folded across chest                             |
| <b>Content</b>       | Gross et al [27]        | Expanded limbs and torso, low energy   |
| <b>Happiness</b>     | Coulson [100]           | Head bent back, no forward chest movement, arms raised above shoulder, straight at elbow   |
|                      | Dahl & Friberg [96]     | Saxophone: Large, regular, fluid, somewhat slow movements. Bassoon: Large, fairly regular, jerky, fast movements                     |
|                      | De Silva et al. [102]   | Vertical and lateral extension of the arms, opening of the shoulders   |
|                      | Kleinsmith et al [80]   | Head bent back, elbows bent, arms raised   |
|                      | Roether et al [91]      | Head upright, straight spine, arms straight  |
| <b>Joy</b>           | De Meijer [101]         | Straight trunk & legs, upward, forward, fast body movement, muscles tensed, arms open frontally                                      |
|                      | Gross et al [27]        | Expanded limbs and torso   |
| <b>Elated joy</b>    | Wallbott [40]           | Shoulders lifted, head bent backwards, arms stretched frontal  |
| <b>Interest</b>      | De Meijer [101]         | Straight trunk & legs, stepping forward, muscles relaxed, slow velocity, arms open frontally   |
| <b>Potency</b>       | Kleinsmith et al. [103] | Low potency: hands along body. High potency: hands raised to shoulder & extended frontally   |
| <b>Pride</b>         | Wallbott [40]           | Head bent backwards, arms crossed in front   |
|                      | Gross et al [27]        | Expanded limbs and torso, high energy, hurried, tense and controlled flow  |
| <b>Sadness</b>       | Coulson [100]           | Forward head bend, forward chest bend, no abdominal twist, arms at side of the trunk   |
|                      | Dahl & Friberg [96]     | Small, very slow, very fluid, fairly regular movements; Bassoon: very little movement  |
|                      | Castellano et al [97]   | Low level of upper body movement, slow velocity of head movements  |
|                      | Gross et al [27]        | Low energy, slow, tense and controlled flow  |
|                      | De Silva et al. [102]   | Little to no vertical or lateral extension of the arms   |
|                      | Kleinsmith et al [80]   | Head bent forward, arms extended straight down alongside body  |
|                      | Roether et al [91]      | Head bent forward, arms straight   |
| <b>Serene</b>        | Wallbott [40]           | Collapsed upper body   |
|                      | Castellano et al [97]   | High velocity of head movements, high quality of motion  |

|                    |                         |  |
|--------------------|-------------------------|--|
| <b>Shame</b>       | De Meijer [101]         | Bowed trunk & head, knees slightly bent, downward body movement, light force (muscles relaxed), slow velocity, stepping backward |
|                    | Wallbott [40]           | Collapsed upper body   |
| <b>Surprise</b>    | Coulson [100]           | Backward head & chest bends, abdominal twisting, arms raised with forearms straight  |
|                    | De Meijer [101]         | Straight trunk and legs, backward stepping, fast velocity  |
| <b>Terror</b>      | Wallbott [40]           | Arms stretched sideways  |
| <b>Threatening</b> | Aronoff et al [92]      | Diagonality & angularity of both arms & movement, diagonal poses   |
| <b>Triumphant</b>  | Kleinsmith et al [60]   | Shoulders straight up or back, arms raised and extended laterally  |
| <b>Valence</b>     | Kleinsmith et al. [103] | High valence has vertical extension of the arms & greater 3D distance between heels  |
| <b>Warmth</b>      | Aronoff et al [92]      | Roundedness of both arms and body movement, more static and moving arabesques  |
| <b>Admiration</b>  | De Meijer [101]         | Straight trunk & legs, upward body movement, stepping forward, arms open frontally   |
| <b>Antipathy</b>   | De Meijer [101]         | Bowed trunk & head, knees slightly bent, stepping backward   |
| <b>Sympathy</b>    | De Meijer [101]         | Straight trunk and legs, stepping forward, arms open frontally, muscles relaxed, slow velocity                                   |

TABLE 3

| Modality    | Ref   | Affective states  | Acted/Non            | Stimuli                          | Gr. truth | Method                      | Accuracy                      |
|-------------|-------|---|----------------------|----------------------------------|-----------|-----------------------------|-------------------------------|
| Body        | [113] | (4) anger, fear, grief, joy   | 5 actors             | 20 videos                        | actor     | decision tree               | 36%                           |
|             | [115] | (2) angry, neutral  | 26 actors            | 1560 movements                   | actor     | MLP                         | 33% efficient                 |
|             | [114] | (4) anger, fear, joy, sad   | 30 actors            | 40 point-lights                  | actor     | 5 diff classifiers          | 62%-93%                       |
|             | [119] | (4) angry, happy, sad neutral   | 30 actors            | 1200 movements                   | actor     | SVM                         | 59% (B)<br>76% (U)            |
|             | [118] | (4) High and low intensity negative, happiness, concentration   | 9 non-actors         | 423 playing windows              | obs       | Recurrent NN                | 57%                           |
|             | [116] | (4) angry, happy, sad neutral   | 30 actors            | 1200 movements                   | n/a       | SVM                         | 50% (B) 81% (U)               |
|             | [120] | (4) angry, happy, sad, neutral  | 30 actors            | 1200 movements                   | n/a       | HMM                         | 81%                           |
|             | [61]  | (4) angry, happy, sad, neutral  | 13 actors            | 520 strides                      | actor     | Naïve Bayes, NN, SVM        | 69% (II)<br>95% (PD)          |
|             |       | (3) valence, arousal, dominance   |                      | 780 strides                      |           | NN                          | 88% (V)<br>97% (A)<br>96% (D) |
|             | [132] | (2) levels of engagement  | 5 non-actors         | 44 videos                        | obs       | ADTree, OneR                | 82%                           |
|             | [117] | (3) angry, happy, sad   | 13 actors            | 138 postures                     | actor     | CALM                        | 96%                           |
|             | [103] | levels of valence, arousal, potency, avoidance  | 13 actors            | 111 postures                     | actor     | BP                          | 79%-81%                       |
|             | [60]  | (4) concentrating, defeated, frustrated, triumphant<br>(2) valence and arousal                          | 11 non-actors        | 103 postures                     | obs       | MLP                         | 60%<br>84% (V)<br>87% (A)     |
| Multi-modal | [121] | (3) levels of interest  | 8 actors             | 262 multimodal                   | obs       | NN                          | 55%*                          |
|             | [22]  | (2) pre- or not pre-frustration   | 24 non-actors        | 24 multimodal                    | actor     | kNN, SVM, GP                | 79%                           |
|             | [122] | (6) 4 basic + anxiety, uncertainty  | 4 actors             | 27 face videos<br>27 body videos | actor     | BayesNet                    | 91% & 94%                     |
|             | [124] | (7) anger, anxiety, boredom, disgust, joy, puzzle, surprise   | 23 actors            | 262 videos                       | actor     | SVM, 1-NN                   | 82%-89%                       |
|             | [123] | (12) anger, anxiety, bored, disgust, fear, happy, pos & neg surprise, neutral, uncertainty, puzzle, sad | 23 actors            | 539 videos                       | obs       | Adaboost with Random Forest | 83%                           |
|             | [135] | (5) Anger, joy, sadness, pleasure, deadpan + # (synchronization, leadership)                            | 4 actors (musicians) | 54 videos                        | actor     | CPR                         | 45%                           |