

Recognizing Affective Dimensions from Body Posture

Andrea Kleinsmith and Nadia Bianchi-Berthouze

UCL Interaction Centre, University College London, London, WC1E 7DP, UK
a.kleinsmith@cs.ucl.ac.uk, n.berthouze@ucl.ac.uk

Abstract. The recognition of affective human communication may be used to provide developers with a rich source of information for creating systems that are capable of interacting well with humans. Posture has been acknowledged as an important modality of affective communication in many fields. Behavioral studies have shown that posture can communicate discrete emotion categories as well as affective dimensions. In the affective computing field, while models for the automatic recognition of discrete emotion categories from posture have been proposed, to our knowledge, there are no models for the automatic recognition of affective dimensions from static posture. As a continuation of our previous study, the two main goals of this study are: i) to build automatic recognition models to discriminate between levels of affective dimensions based on low-level postural features; and ii) to investigate both the discriminative power and the limitations of the postural features proposed. The models were built on the basis of human observers' ratings of posture according to affective dimensions directly (instead of emotion category) in conjunction with our posture features.

1 Introduction

The role of computers and other technologies in many facets of today's society explains the importance of creating systems to be capable of interacting well with humans. Affective communication may be used to provide developers with a rich source of information for achieving this goal. Posture in particular has been recognized as an important modality for affective communication [20][1][8][5]. In fact, cognitive neuroscience studies have shown the importance of body posture over facial expressions in cases of incongruent affective displays [7].

Behavioral studies have shown that posture can communicate both affective dimensions and discrete emotion categories [17][5]. In the affective computing field, models for the automatic recognition of discrete emotion categories from static body postures have been proposed [2][5][15]. Other studies have shown that body motions extracted from dance sequences [3][24][14] can convey emotion in terms of discrete categories effectively. To our knowledge, what are missing still are models for the automatic recognition of affective dimensions from posture as there are for motion [4]. This is important as a single label may not reflect the complexity of the affective state conveyed by the posture. Indeed, Ekman

and Friesen [10] initially posit that while they consider the face to be the foremost modality for expressing discrete emotion categories, the body is better at communicating dimensions of affect.

While affective posture recognition systems continue to focus on discriminating between emotion categories, our previous work [18] has attempted to understand if we can go beyond these categories and recognize levels of affective dimensions. Using a multidimensional scaling technique, affective dimensions were extracted that support the categorical evaluation of affective posture. The extracted dimensions are arousal, valence, and action tendency. The goals of the new study are: i) to build automatic recognition models to discriminate between levels of affective dimensions based on low-level postural features; and ii) to investigate both the discriminative power and the limitations of the postural features proposed.

The remainder of the paper is organized as follows: Section 2 describes the affective dimension recognition survey and the automatic recognition models built upon the survey data and our low-level postural features. The models and the statistically determined most relevant features of each are evaluated in Section 3. Section 4 provides a discussion on some further testing that was carried out and future directions to be explored.

2 Affective Dimension Recognition

2.1 Human Recognition of Affective Dimensions

As our goal is to build posture recognition models based on affective dimensions, an online survey was conducted to collect data from human observers. To carry out the survey, the same set of affective posture images that were used in the previous study were used in the current study. The reader is directed to [17] for a detailed explanation of the posture collection and stimulus preparation processes.



Fig. 1. Examples of the affectively expressive avatars reconstructed from motion capture data.

The survey was conducted online as a series of webpages. The stimuli comprised 111 affective posture images, and were presented separately (one posture

per page), in a randomized order that differed for each participant (called observers hereafter). Five anonymous observers, two males and three females participated in the study. In contrast to the previous study, in the current survey, human observers judged postures according to affective dimensions directly instead of judging postures according to discrete emotion categories only. Specifically, observers were asked to rate each posture according to a seven-point Likert scale for each of four affective dimensions: *valence* (pleasure), *arousal* (alertness), *potency* (control), and *avoidance* (avoid/attend to). Valence, arousal, and potency (or dominance) were chosen based on psychological research throughout the last century which asserts that these three dimensions cover the majority of affect variability [25][23][6] [22]. Furthermore, valence and arousal were two of the dimensions identified in our initial study, and we consider potency a form of action tendency, which was the third dimension we identified previously.

Next, we analyzed the observers' data to determine the level of agreement between observers in rating the affective dimensions. Using the normalized observers' rating for each posture for each dimension, the estimated marginal means for each observer were calculated for each dimension. Results showing differences between observers across the four affective dimensions are presented in Figure 2. The *x-axis* denotes the four affective dimensions and the *y-axis* denotes the estimated marginal means for each observer. As we can see, the greatest variability occurs with observer two across the arousal, potency, and avoidance dimensions. According to the means, across all observers, the arousal dimension contains the least amount of variability while the greatest amount of variability occurs for the *avoidance* dimension. Cronbach's α was used to test the reliability of the observers' agreements across the four dimensions. The results reflect the above findings with arousal showing the highest reliability ($\alpha = 0.85$, $r = 0.55$), and avoidance showing the lowest reliability ($\alpha = 0.35$, $r = 0.11$).

Next, we calculated the average percentage of error across all observers for each dimension, considering each observer's rating for each posture separately. The results obtained show an average error of 19% for valence, 15% for arousal, 19% for potency, and 25% for avoidance. We consider these differences to reflect the variability that may typically occur in human-human affective communication.

2.2 Automatic Recognition of Affective Dimensions

To build automatic affective dimension recognition models, each static posture is associated with a vector of 24 postural configuration features (explained below), and four values indicating the observers' normalized average ratings for each affective dimension: *valence*, *arousal*, *potency*, and *avoidance*. The same features were used to build emotion category recognition models in previous studies [2][16][9].

The postures are described according to the 24 features introduced in our previous research [2][8], which are listed in Table 1, and shown in Figure 3. These features, calculated using the numerical data obtained from the motion capture,

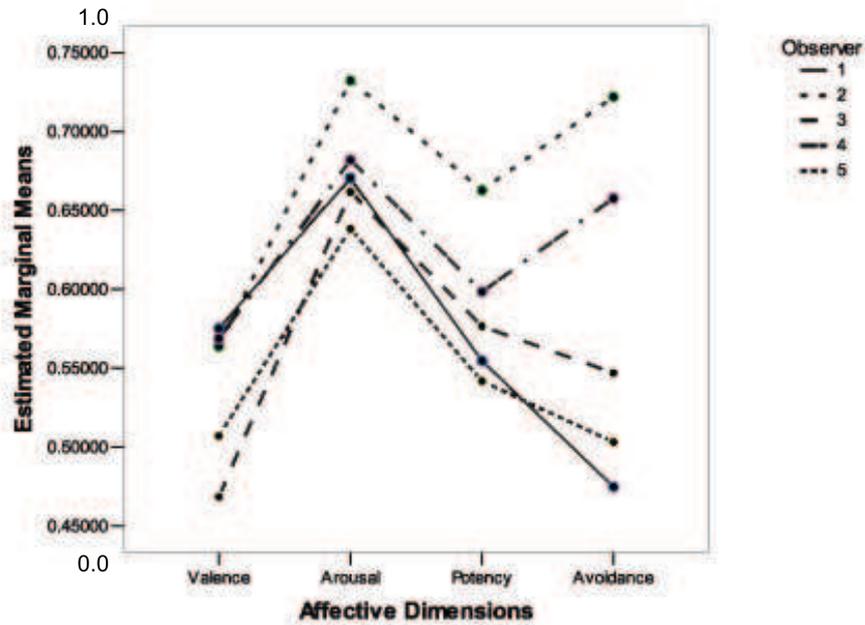


Fig. 2. Results of observer ratings for each affective dimension. The affective dimensions comprise the x-axis, and the estimated marginal means for each observer comprise the y-axis.

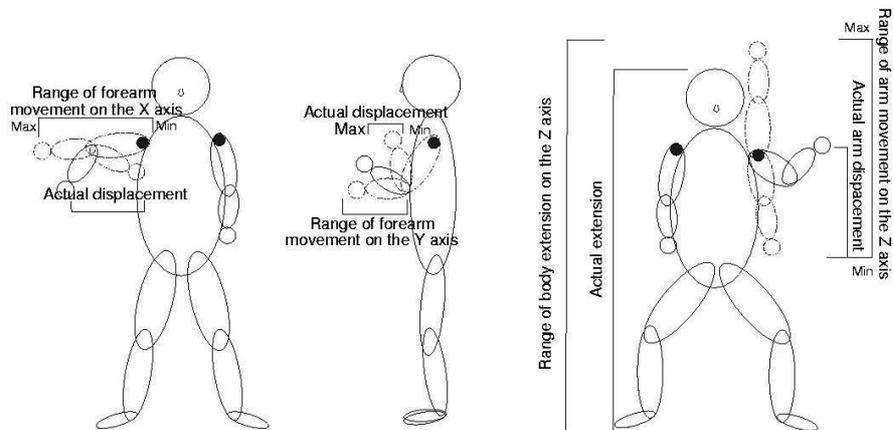


Fig. 3. Visual representation of some of the postural configuration features. These features were computed in the frontal view by projecting 3D motion captured data on the 3 orthogonal planes to measure direction and volume of the body according to the lateral, frontal, and vertical extensions of the body, and body orientation.

were chosen because they are low-level and context-independent. Precisely, direction and volume of the body were described by projecting each marker on the three orthogonal planes and measuring the lateral, frontal, and vertical extension of the body, body torsion, and the inclination of the head and shoulders.

Table 1. The table lists the set of posture features proposed. The Code column indicates the feature codes used in the paper. The following short-cuts are used: L: Left, R: Right, B: Back, F: Front.

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

A backpropagation algorithm was used to build a separate model for each affective dimension. One reason for choosing backpropagation is its effective handling of data comprising continuous values. Although other algorithms may perform better, finding one is not the goal. Instead, the goal is to examine our set of features, and to determine whether or not they are effective for discriminating between levels of affective dimensions from posture. Furthermore, backpropagation is a well accepted algorithm used in recognition tasks.

Table 2. Performance levels of backpropagation for the 111 postures for each affective dimension model using the complete set of 24 posture features.

Affective Dim. Models	Error %: 24 Features
Valence	21%
Arousal	21%
Potency	19%
Avoidance	19%

The topology of the backpropagation network was determined empirically. It consists of one input layer with 24 nodes (corresponding to the 24 posture features), one hidden layer with 12 nodes (the number of input nodes/2), and

an output layer with one node (corresponding to the observers’ ratings for each dimension, separately). Both the learning rate and the momentum were set to 0.01. Each network was trained with 10,000 epochs, and tested using five fold cross-validation. The performance error obtained for each of the backpropagation models, shown in Table 2, is 21% for both the *valence* and *arousal* dimensions, and 19% for both the *potency* and *avoidance* dimensions. Comparable to the results of the human observers discussed in Section 2.1, these results using backpropagation are quite promising.

3 Grounding Affective Dimensions into Posture Features

In the first part of this paper, we proposed a set of low-level postural features and tested the possibility of building models for recognizing levels of affective dimensions. As mentioned above, these results are quite promising as they are similar to the level of agreement between our observers. This second part of the paper tackles two issues. One, we look at how humans may use postural features to discriminate between levels of affective dimensions. Two, we assess the limitations of our features and hence, which features may be necessary to add in future studies. To answer the first question, we extract the most relevant features for each affective dimension by applying non-linear mixture discriminant analysis (MDA) [19][12] to the postures for which agreement between the observers was high. To answer the second question, we examine the postures for which agreement between the observers was very low and discuss them according to models presented in other studies.

MDA was applied to each of the four affective dimensions separately. MDA is a statistical technique used to discriminate between two or more categories or groups. Mixtures of Gaussian distributions are used to model the groups. Moreover, the separation between groups is maximized while the variance within the groups is minimized. Postures that received an average observer rating of < 3.8 were labeled *low*. Postures that received an average rating between $3.8 - 4.2$ were labeled *neutral*. Finally, postures that received an average rating of > 4.2 were labeled *high*. Given that the majority of the postures were labeled as either *low* or *high*, the number of subclasses used by the MDA algorithm for classification was two for low-rated postures, one for neutral-rated postures, and two for high-rated postures. Furthermore, in order to discern the most discriminating feature sets, only the postures which obtained high agreement between observers (> 0.80) were used. Thus, after discarding the low agreement postures from the original 111, 91 postures remained for the valence dimension, 101 postures remained for the arousal dimension, 93 postures remained for the potency dimension, and 68 postures remained for the avoidance dimension. The MDA algorithm allows us to ascertain the most relevant features because it uses an iterative process to create the models based on linear combinations of the most discriminating features.

Separate models were built to discriminate between pairs of affective dimension levels (e.g., low vs. high, low vs. neutral, etc.). The models obtained are

Table 3. Important features for discriminating between pairs the affective dimension levels are presented. The last column contains the percentage of error for each model. The following short-cuts are used: Dim: Dimension, H: High, L: Low, N: Neutral.

Affective Dim. Model	Head		Vertical					Frontal					Lateral					Heels		MDA Error
	V5	V6	V7	V16	V22	V23	V8	V9	V19	V20	V21	V24	V10	V11	V13	V14	V15	V26	V27	
Valence HL	3	1																2		16%
Arousal HL	1	2				5							3			4				5%
Potency HL	3		2	1				4						5	6					3%
Avoidance HL	4		1			2								3						7%
Valence HN	4					1					2							3		27%
Arousal HN			1		4		3							2						9%
Potency HN	4		3	1			5				2			8			6	7		19%
Avoidance HN	2				5	1							4	3						14%
Valence LN			3					1	2											28%
Arousal LN		3			1					4	5					2				12%
Potency LN	3	4		1											2					20%
Avoidance LN		1											2							18%

shown in Table 3. The last column of the Table reports the percentage of classification error for the models. The remaining columns report the features used by the discrimination functions of each model. The number in each entry of the Table represents the rank of importance of the features. Lower numbers correspond to higher discriminative power. For conciseness, we report in this Table only the features selected by at least one of the four models.

An examination of the Table shows that the head ($V5$) is important for distinguishing between low and high for all dimensions and high and neutral for all dimensions except arousal. For arousal, to distinguish between high and neutral, and low and neutral levels, openness of the body seems to be important as these models rely on at least one vertical, frontal, and lateral feature. In the case of potency, the most discriminating feature for all three models is $V16$, the vertical extension of the arm. Frontal features do not play a role in any of the avoidance models. Similarly, lateral features are not represented in any of the valence models.

A visual examination of the the labels assigned to the subclasses for low- and high-rated avatars for all four dimensions indicates that the relevant features identified by MDA do aid in discriminating between postures. Refer to Figures 4, 5, 6, and 7 for examples. Low subclasses of valence (Figure 4(a-d)) appear to be separated by the vertical extension of the arm ($V6$) and the 3D distance between the heels ($V27$). In one subclass (Figure 4(a,b)), the arms are stretched down alongside the body and the heels remain close together (the body is quite closed), while in the other subclass (Figure 4(c,d)), the arms are raised to face level and the heels are further apart. Whether the arms are raised well above the head (Figure 4(e,f)) or fall within the torso area ($V6$) (Figure 4(g,h)) seems to separate postures within the two high subclasses of valence.

Lateral features ($V10$ and $V14$) are used for distinguishing between high levels of arousal. As we can see from Figure 5, the body in one subclass (Figure

5(c,d)) is significantly more open laterally than in the other high subclass (Figure 5(a,b)).

In the case of potency, the vertical ($V7$ and $V16$) and lateral ($V11$ and $V13$) openness of the body seem to classify postures within the two low-rated subclasses (Figure 6(a-d)). In the high subclasses of potency, the vertical ($V7$ and $V16$) and frontal ($V9$) extension of the arms are the most distinguishing features. In one high subclass (Figure 6(e,f)), hands rest on the hips, whereas in the other high subclass (Figure 6(g,h)), the arms are more raised and extended out in front of the body.

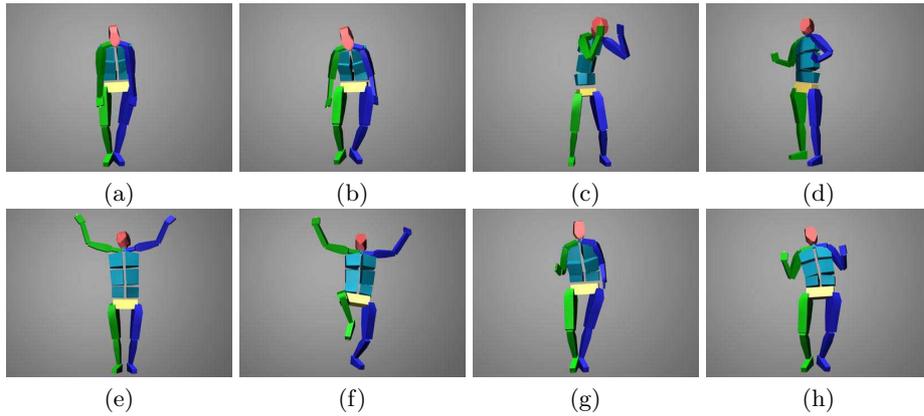


Fig. 4. Examples of the 3D affectively expressive avatars for low and high **valence** subclasses. (a)(b) Low subclass 1, (c)(d) Low subclass 2, (e)(f) High subclass 1 (g)(h) High subclass 2

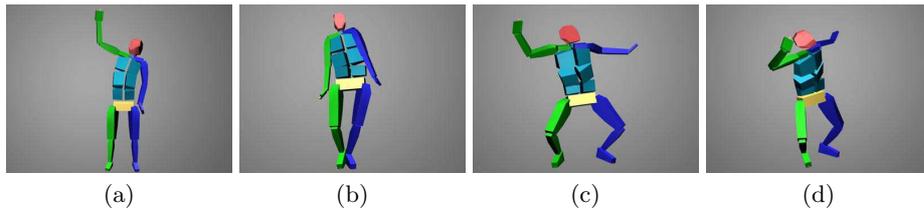


Fig. 5. Examples of the 3D affectively expressive avatars for high **arousal** subclasses. (a)(b) High subclass 1 (c)(d) High subclass 2

Vertical ($V7$ and $V23$) and lateral ($V10$ and $V11$) features are important for distinguishing between two levels of low avoidance, as well as two levels of high

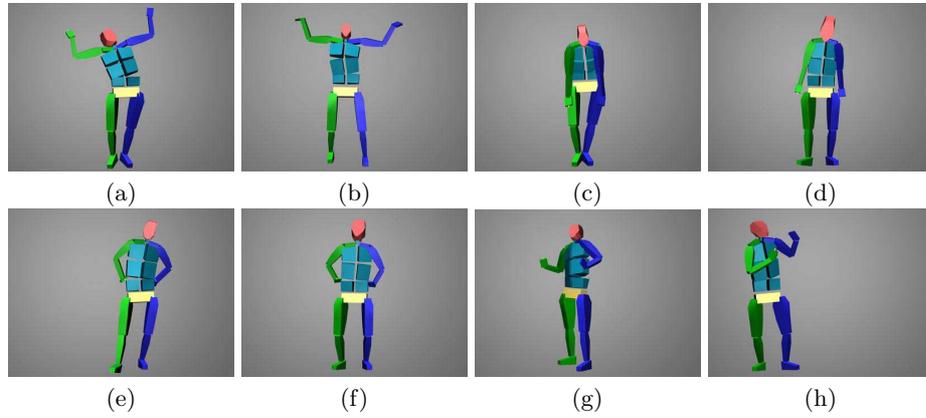


Fig. 6. Examples of the 3D affectively expressive avatars for low and high **potency** subclasses. (a)(b) Low subclass 1, (c)(d) Low subclass 2, (e)(f) High subclass 1 (g)(h) High subclass 2

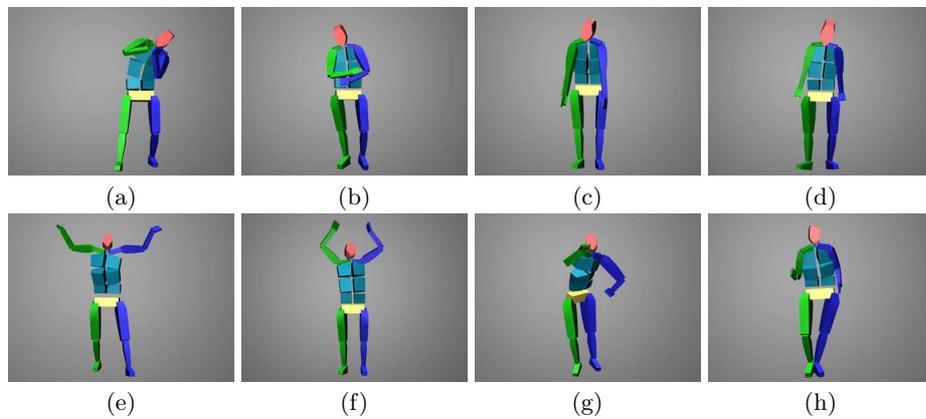


Fig. 7. Examples of the 3D affectively expressive avatars for low and high **avoidance** subclasses. (a)(b) Low subclass 1, (c)(d) Low subclass 2, (e)(f) High subclass 1 (g)(h) High subclass 2

avoidance. In particular, arms are folded across the body or near the face for one low subclass (Figure 7(a,b)), while in the other low subclass (Figure 6(c,d)), the arms are stretched down along the side of the body. In the case of high subclasses, one subclass (Figure 6(e,f)) contains postures in which the arms are raised and open laterally, while postures in which the arms rest at torso level both vertically and laterally are represented in the second high subclass (Figure 6(g,h)).

Interestingly, an examination of the misclassified postures reveals a possible inconsistency in perception by the observers, and not in classification by the automatic recognition models. In the majority of the cases, the postures actually seem to be classified correctly by MDA when considering postural features.

4 Discussion

The results presented in this paper are quite interesting and promising. Our set of low-level posture configuration features seem quite suitable for building automatic recognition models for affective dimensions. In fact, backpropagation was used for further evaluation using reduced feature sets to build a new set of affective dimension models. These models were built with the set of features determined by MDA to be the most discriminating features for each separate affective dimension, and tested on the entire set of 111 postures. The reduced number of features required a reduction in the number of nodes for both the input layer and the hidden layer of the architecture. Specifically, there were five nodes in the input layer and three nodes in the hidden layer for the valence, potency, and avoidance models; and four nodes in the input layer and two nodes in the hidden layer for the arousal model. The results obtained from testing are positive and lend further support to the effectiveness of our features, as a marked decrease can be seen in the error percentages across all four dimensions. There was a 12% error percentage for valence, 10% for both arousal and potency, and 11% in the case of avoidance.

While the recognition models presented here were tested with standing postures, the low-level, context-independent nature of our features has also been assessed with a combination of standing and seated postures. In that situation, our affective posture recognition system [16] was used to classify posture according to nine affective categories (*angry, confused, fear, happy, interest, relaxed, sad, startled, and surprised*) chosen to represent different types of emotion situations. On average, 70% of the postures were correctly classified.

However, there are more situations and conditions to examine. As mentioned in Section 3, other types of features, (e.g., amplitude, speed of movement, direction, etc.) may be necessary for achieving better recognition of some affective states such as fear [5][16]. Indeed, an evaluation of the postures that were discarded due to low agreement on affective dimension ratings between observers reveals that fear and surprise were the most frequently chosen emotion categories.

Moreover, currently unrepresented static features may add to the performance of the models. In fact, a 1932 behavioral study by James [13] found that Approach-Withdrawal (leaning direction) and Expansion-Contraction of the body are important for attributing attitude to static posture. This importance of leaning direction also has been evidenced by more recent studies [11][21].

References

1. Argyle, M., *Bodily Communication*, Methuen & Co. Ltd, London, (1988)
2. Bianchi-Berthouze, N. and Kleinsmith, A., A categorical approach to affective gesture recognition, *Connection Science*, **15**, (2003), 259-269
3. Camurri, A., Hashimoto, S., Suzuki, K., Trocca, R., Kansei analysis of dance performance, In: *IEEE Int'l Conf. on Systems, Man and Cybernetics*, (1999), 327-332
4. Camurri, A. and Volpe, G. and De Poli, G. and Leman, M., Communicating Expressiveness and Affect in Multimodal Interactive Systems, *IEEE Multimedia*, **12**, (2005), 43-53
5. Coulson, M., Attributing emotion to static body postures: recognition accuracy, confusions, and viewpoint dependence, *Jour. of Nonv. Behav.*, **28**, (2004), 117-139
6. Davitz, J., Auditory correlates of vocal expression of emotional feeling, In Davitz, J., ed.: *In The Communication of emotional Meaning*, New York, McGraw-Hill, (1964), 101-112
7. de Gelder, B., Snyder, J., Greve, D., Gerard, G., Hadjikhani, N., Fear fosters flight: A mechanism for fear contagion when perceiving emotion expressed by a whole body, *Proc. of the National Academy of Science*, **101**(47), (2003), 16701-16706
8. de Silva, P.R. and Bianchi-Berthouze, N., Modeling human affective postures: An information theoretic characterization of posture features, *Journal of Computer Animation and Virtual Worlds*, **15**, (2004), 269-276
9. de Silva, P., Kleinsmith, A., Bianchi-Berthouze, N., Towards unsupervised detection of affective body posture nuances, In: *Proceedings of the First International Conference on Affective Computing and Intelligent Interaction*, Springer-Verlag, (2005), 32-39
10. Ekman, P., Friesen, W., Head and body cues in the judgment of emotion: A reformulation, *Perceptual and Motor Skills*, **24**, (1967), 711-724
11. Harrigan, J., Rosenthal, R.: Physicians head and body positions as determinants of perceived rapport, *Journal of Applied Social Psychology*, **13**(6), (1983), 496-509
12. Hastie, T. and Tibshirabi, R., Discriminant analysis by Gaussian mixture, *Journal of the Royal Statistical Society*, **B:58**, (1996), 155-176
13. James, W.T., A study of the expression of bodily posture, *Journal of General Psychology*, **7** (1932), 405-437
14. Kamisato, S., Odo, S., Ishikawa, Y., Hoshino, K., Extraction of motion characteristics corresponding to sensitivity information using dance movement, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, **8**(2), (2004), 167-178
15. Kapoor, A. and R. Picard, and Y. Ivanov., Probabilistic combination of multiple modalities to detect interest, *Proc. of the 17th International Conference on Pattern Recognition*, **3**, (2004), 969-972
16. Kleinsmith, A., Fushimi, T., and Bianchi-Berthouze, N., An incremental and interactive affective posture recognition system, *Proc. Workshop on Adapting the Interaction Style to Affective Factors*, held in conjunction with UM'05, Edinburgh, July, (2005).

17. Kleinsmith, A., de Silva, P., Bianchi-Berthouze, N., Cross-cultural differences in recognizing affect from body posture, *Interacting with Computers* **18** (2006), 1371–1389
18. Kleinsmith, A., de Silva, P., Bianchi-Berthouze, N., Grounding affective dimensions into posture features, In: *Proceedings of the First International Conference on Affective Computing and Intelligent Interaction*, Springer-Verlag, (2005), 263–270
19. Lachenbruch, P.A., *Discriminant Analysis*, NY, Hafner, (1975).
20. Mehrabian, A., Friar, J., Encoding of attitude by a seated communicator via posture and position cues, *Journal of Consulting and Clinical Psychology*, **33**, (1969), 330–336
21. Mehrabian, A., Inference of attitude from the posture, orientation, and distance of a communicator. *Journal of Consulting and Clinical Psychology*, **32** (1968), 296–308
22. Mehrabian, A., Russell, J., *An Approach to Environmental Psychology*, MIT Press, Cambridge, (1974)
23. Osgood, C., Suci, G., Tannenbaum, P., *The measurement of meaning*, University of Illinois Press, Chicago, (1957)
24. Woo, W., Park, J. and Iwadate, Y. Emotion analysis from dance performance using time-delay neural networks, *Proc. of the JCIS-CVPRIP*, **2**, (2000), 374–377
25. Wundt, W., *Outlines of psychology*, Wilhelm Englemann, Leipzig, (1907)