Computational model and the human perception of emotional body language (EBL)

Marina Fridin[1], Avi Barliya[1], Edna Schechtman[3], Beatrice de Gelder[2], Tamar Flash[1]

Abstract. Emotions are undeniably a central component of human existence. In recent years, the importance of developing systems which incorporate emotions into human-computer interaction (HCI) has been widely acknowledged. However, research on emotion recognition has been dominated by studies of facial expression of emotion. In comparison, the study of EBL has received relatively little attention.

Here we study the phenomena of EBL, specifically of *static body postures expressing emotions,* from two different perspectives. First, we have built a computational model for the recognition of four basic emotions which achieves a relatively high recognition rate (70 %). Secondly, to study perception of EBL, we examined what body parts attract the observer's attention during the perception of EBL. This is done by tracking eye movements of human subjects during the observation of static postures expressing emotions. Although invaluable information can be inferred from motion, this study will show that information about static body posture is rich enough for both automatic recognition and human perception.

The present study contributes in an applicative way both to the development of automatic recognition systems of EBL and provides insight into the nature of human recognition of EBL.

1 INTRODUCTION

Research on EBL is rapidly emerging as a new field in Human-Computer Interaction (HCI) and affective computing.

Darwin in his groundbreaking work on the evolutionary significance of emotion stressed the importance of bodily movement [4]. Nonetheless, the dominant theory of Ekman [8] denied that specific gestures, body movements, or postures could indicate emotions. Contrary to the above claims, Wallbott a pioneering researcher in the field study of EBL [16] , has demonstrated that specific body movements, postures and gestures, generally allow to differentiate among different emotions.

Studies of EBL are still rare but it has been already repeatedly shown that distinct expressions of at least the basic emotions are readily recognized even in the absence of facial and vocal cues, when emotions are portrayed by static body postures and by whole body movements, as was described in [1], [7] and [16] . In recent years, Pollick and colleagues [14] have investigated the perception of EBL through movement and have shown that the speed of movement plays a major role in the perception of EBL. In the study by Cammuri [4] of emotional dance performance, hand configurations and locations with respect to the body were also indicated as important features for emotion discrimination. The neural mechanisms underlying perception of EBL were examined by de Gelder et al. [7] , where in a fMRI study they have shown that the amygdala is a key structure in the processing of static, fearful body expressions, along with the fusiform cortex.

In this study we propose a novel computational approach towards the understanding of EBL. The strength of the proposed model is that it is highly modular, that is, it is independent of the database being examined and the feature extraction processes.

We have also studied of the perception of EBL by investigating gaze patterns of human subjects when perceiving EBL of other individuals. In this study, the first to describe such gaze patterns, we examined whether fixation behavior is emotion-specific. Fixation behavior was described in terms of which body parts receive most visual attention. We also examined whether fixation behaviour remains stable under different task conditions, such as observation vs. recognition, for images which were correctly vs. incorrectly recognized, and as a function of the familiarity with the stimuli, which may result in perceptual learning

As it is impossible to cover a large spectrum of human emotions, in this study we have strived to characterize and model the four basic emotions: ***JOY***, ***SADNESS***, ***ANGER***, and ***FEAR***.

The paper is organized as follows. First, we describe the database of EBL that was used for both parts of the research (Section 2). Next, in Section 3 we discuss how EBL is perceived and recognized. In Section 4, we present the computational model of EBL. Section 5 presents our conclusions and contrasts this paper with related works. Section 6 provides directions for future work.

2 EBL DATABASE

The input material in both studies consisted of still images derived from video films in which 80 semi-professional actors and 47 ordinary subjects freely portrayed body postures expressing four basic emotions. The subjects were of different genders, cultures, ages and socioeconomic status. The images from both groups were randomly mixed together. Therefore the entire database consists of 508 images (127x4). In the eye tracking study the faces on the photographs were blurred.

[1] Dept. of Computer Science and Applied Math., Weizmann Institute of Science, Rehovot, Israel. Email: {marina.ousov-fridin, avi.barliy, tamar.flash}@weizmann.ac.il.

[2] Lab. of Cognitive and Affective Neuroscience, Tilburg Univ., the Netherlands. Email: {b.degelder}@uvt.nl.

[3] Dept. of Industrial Engineering and Management, Ben Gurion Univ. Beer Sheva, Israel. Email: {ednas}@ bgu.ac.il.

3 GAZE AND THE PERCEPTION OF EBL, AN EYE TRACKING STUDY

In this study we examined eye fixation patterns during the perception and recognition of EBL. Participants were asked to perform two tasks while wearing an eye-tracking system. In the first task (observation task), subjects were asked just to look at photographs and in the second (recognition task) to tell verbally which emotion they have recognized.

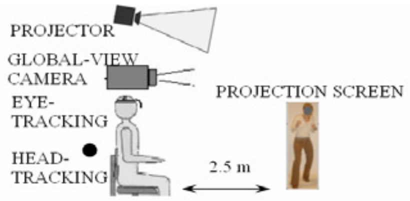

Figure 1. Schematic representation of the experimental set up.

An "ISCAN" eye-tracking apparatus was mounted on the participant's head (Figure 1) and was connected to the eye-tracking computer. To estimate global gaze direction, the eye-tracking apparatus was coupled with a Polhemus Liberty head-tracker. A separate computer projected the stimuli on the screen and captured the presentation data. A global view camera captured the scene observed by the subject and produced a video showing the scene and the estimated gaze directions

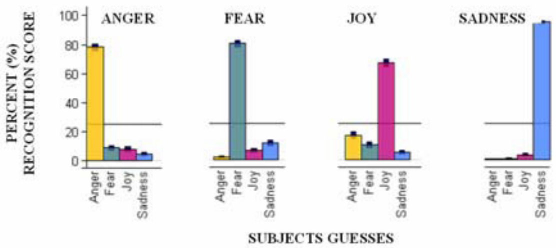

Figure 2. The recognition score (RS) for each emotion shown in static postures presented to the subjects. Subjects' errors in the recognition task are also shown. RS of all emotions were clearly well above the 25% chance level (black line).

We first evaluated subjects' recognition rate of EBL in the recognition task. Figure 2 shows the recognition accuracy for each emotion (when the emotion is actually shown) and the respective confusion rate. Choices were not made randomly: the emotions shown were very frequently recognized, with sadness being the most easily recognizable emotion, with a recognition rate of 94.83% $\pm$ 1.19%. The other emotions were also well recognized, when the most confused emotion being fear, which was confused with sadness (11.54% $\pm$ 1.77%), and joy, which was confused with anger (17.11% $\pm$ 2.13%).

Response times (RT) for emotions that were incorrectly recognized were significantly longer than the response times for correctly recognized emotions (means 38543 $\pm$ 29329 msec (n=134) and 27004 $\pm$ 22720 msec (n=586), respectively).

We then quantified and performed a statistical analysis of body regions on which the eyes fixated. There were significant differences among the eye fixation patterns observed for the different emotions (Figure 3). Most attention during the perception of images showing anger was paid to the hands (7.32 $\pm$ 1.64) and the arms (4.28 $\pm$ 0.83), with similar observations for fear. For images showing joy most of the time was spent gazing at the head (3.55 $\pm$ 0.89), while the hands and the arms also attracted significant attention. The fixation times for images showing sadness were distributed equally between the head, the arms and the hands. The trunk and the legs never attracted significant attention. There was no change in the fixation behavior in the observation versus the recognition task, and while viewing correctly vs. incorrectly recognized images.

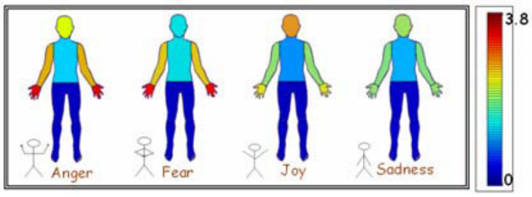

Figure 3. Distribution of fixation behavior on body segments for each emotion during the recognition task. Fixation behavior is represented by color: from blue (marked by 0) to red (3.8 and more).

These fixation patterns were not influenced by the task and were not correlated with recognition performance nor with RT. Our results suggest that fixation behavior is emotion-specific and is automatically triggered by the stimuli. The observed differences in recognition accuracies and RTs could not be explained by the observed fixation behavior. Finally, by analyzing changes in recognition accuracy, RT, and fixation patterns over task repetition we concluded that no perceptual learning could be detected.

4 A COMPUTATIONAL MODEL

We applied a heuristic approach for the modeling of EBL. To that end, we assumed that a set of well defined features with the aid of some combinatorial syntax would be sufficient to represent any EBL. We identified the major feature types, which we then extracted from the pictures as follows.

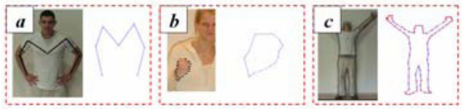

Figure 4. Extracted features: ***a***: Body Joint Angles; ***b***: Gesture; ***c***: Body Silhouette

The photographs were treated with different image processing algorithms, which included body segmentation from the background, labeling body parts, building a multi-joint human body model, estimating head position and hand gestures. Since some of these image processing steps are very complicated to be automatically performed, part of the analysis was performed manually. The feature types that were extracted are: head pose characterized by the direction of the normal vector to the plane of the head, using an idea suggested by Gee&Cipolla [10] , upper arms and trunk joint angles, (as illustrated on the Figure 4a) legs joint angles, right/left hand gestures (Figure 4b) body silhouette (Figure 4c), defined by a polygon and body space. On the whole, there were ***seven different feature types.*** These feature types are commonly used in the literature dealing with recognizing emotions [16] , imitation [11] and affective computing [4]

Then, an algorithm that computes the maximum mutual information between the different features and the various emotions (Equation 1) was developed [15] . Mutual Information (MI) between class C and feature f_i is defined as:

$$\mathrm{I}(\mathrm{C};f_i(\theta_i)) = -\sum_{C}\mathrm{P}(\mathrm{C})\mathrm{Log}(\mathrm{P}(\mathrm{C})) + \sum_{F}P(f_i)\sum_{C}\mathrm{P}(\mathrm{C}\,|\,f_i)\mathrm{Log}(\mathrm{P}(\mathrm{C}|\,f_i))$$

By measuring frequencies of detecting f_i for images within different classes C, we could evaluate the mutual information of the feature with respect to the class. Examples of the most informative features of type "Body Silhouette" for the class Joy are shown on Figure 5.

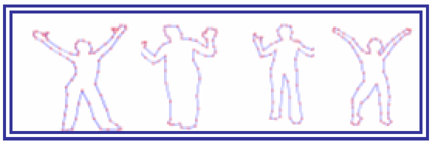

Figure 5. Examples of the most informative features of type "Body Silhouette" for the class of emotion Joy.

Then we re-arranged the lists of most informative features so that each feature added to the list would supply the maximum possible additional MI concerning the emotional class (Equation 2). The selection process was initialized by selecting a feature that provided maximum MI: $f_1 = \arg\max_i I(C; f_i)$

Therefore, each of the following selected features was determined by:

- First, to make sure that information contributed by any candidate feature f_{k+1} is not already represented by any previously selected feature f_i , for each f_{k+1} feature we found the most similar feature f_i (min stage), by computing the conditional MI between the new feature f_{k+1} and the class given the most similar feature f_i : $I(C; f_{k+1} \,|\, f_i)$;
- Then, from all candidate features f_{k+1} the feature with the maximum contribution (i.e. maximum conditional MI) was selected (max stage) as follows:

$$f_{k+1} = \arg\max_{f}\min_{i} I(C; f_{k+1} \,|\, f_i)$$

Then, different feature types were combined into one list, based on calculating the additional MI. The result of the training process consisted of sets of lists which included all possible combinations of the feature types, for each class of emotion. In each list, features were arranged according to their additional *MI*. The results of this stage are illustrated in Figure 6: the six most informative features of type ***Body Joint Angles*** are shown for the class Joy (the first row) in a descending order, whereas the order is changed (Max-Min, second row) when considering the highest additional information score for each of the features.

To recognize a new image *I*, we first measured the function S_I, indicating whether the features from a currently examined list were found in the image or not. Then, to check to which class the image *I* belongs, we need to know the distribution of the functions S of images from the training set and compare it with the computed function S_I. Therefore, for each of four classes of emotions, and for each image in the training set the distribution of S given C_j class of emotion $p(S|C_j)$ was estimated in advance by using the EM algorithm. A mixture of Gaussian model was chosen as an appropriate model for representing the distribution of S. Finally, the image *I* was assigned to a specific class of emotion, j, according to the class for which $p(S_I|C_j)$ is maximal.

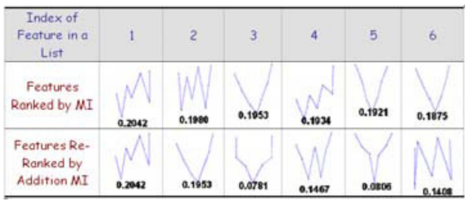

Figure 6. An example showing the first two steps of the algorithm, MI and Max-Min**.** The six most informative ***Body Joint Angles*** are shown for the joy emotion on the first row in a descending order of MI, whereas the sequence of the features is changed (Max-Min, second row) when considering the most additional information provided by each of the features.

We evaluated the recognition rate for the different lists consisting of different feature combinations, and for different sets of features, a few examples of which are presented in Table 1. For each list, we present the average recognition rate (RecRate), taken across all the different classes of emotion, and the required number of features in the combination that maximizes the recognition rate. Using only BODY SILHOUETTE achieves a recognition rate of 46.83% averaged over all emotions. However, using a combination of BODY SILHOUETTE, BODY SPACE, LEFT HAND, and HEAD the best recognition rates: a 70.25% averaged across all emotions was achieved. It should be emphasized that this is a relatively high recognition rate, considerably above chance levels, using only static images as input.

Feature Types Combinations	Rec. Score (%)	Num. Features
Body Silhouette (BS)	46.83	12
Left Hand	43.65	96
BS , Body Space, Left Hand, Head Norm	70.25	396
BS, Body Joint Angles, Left Hand, Head Norm	69.45	285
BS, Body Space, Body Joint Angles, Right Hand	68.25	196

Table 1: Examples of lists with mixed feature types. For each list, we present the average recognition score and the number of features in the list (over all possible number of features) that maximizes the recognition rate.

This work resulted in a model for the recognition of EBL in static images. For a given new image, even if some features cannot be extracted, the model defines the combination of feature sets that has the highest probability of predicting the expressed emotion in the picture.

5 CONCLUSIONS & RELATED WORK

The results of both parts of the presented study, e.g. the computational model and the analyse of fixation behaviour, are consistent with each other. First we have compared the recognition rates achieved by the computational model with the recognition rates of human observers using the same data set.

We have found, on the average that human observers can recognize emotions with an accuracy of 82%, while the recognition scores of the algorithm were around 70%. That is, human performance is not extremely better than the performance by the algorithm. . Similar recognition scores were reported by a number of EBL studies for both human perception [5] and computational models [4, 1]. Nevertheless, human observers may use more information than is available to our model and the ability to extend and improve the model will be discussed below.

Here for the first time eye tracking has been used to examine fixation behavior during perception of EBL When perceiving pictures associated with joy, people tend to fixate on the head, whereas for pictures associated with anger and fear, most attention is devoted to the hands and arms. The feature involving the hands in the eye tracking study encoded information regarding both hand position and configuration. For pictures associated with sadness, people fixate on all body parts: heads, arms, and hands. The legs almost never drew the observers' attention. Also, the effects of hands' and arms' locations were statistically analyzed and it was found that the fixation behavior is emotion rather than location dependent (e.g. the location within the picture: at the center vs. on the sides of the body). For instance, the location of the hands at the center is a feature that is common for both fear and sadness; however, the amount of time that subjects gazed at the hand was considerably longer for fear versus sadness. These results are consistent with features that were found important for the recognition of EBL, both by our computational model and in a number of related studies [1, 4]. It was shown that hand position and configuration, and the orientation of the head are important features for the recognition of EBL.

Finally, it is important to mention perceptuo-motor primitives [9] [11] and their relation to our study. Perceptuo-motor primitives can be defined as a basic set of body segments and postures that can be learned from perceptional behavior (including fixation behavior) and mapped onto the elements of a recognition system (both human and automatic). Therefore, the sets of emotion-specific body segments and postures, identified in our eye-tracking study, were also identified as important features by the computational model. Hence these body segments and the different features characterizing the location and configuration of the body parts which are associated with the different emotions could be defined as emotion-specific perceptuo-motor primitives of EBL.

6 POSSIBLE APPLICATIONS AND FUTURE WORK

The possible applications of the presented computational model to the automatic recognition of EBL are numerous, ranging from the development of robotic systems capable of mimicking and expressing emotions, through having more natural looking characters in computer games (virtual worlds) and animated movies.

In addition, our data suggest that many more experiments are needed in order to further clarify the mechanisms sub-serving the perception of EBL. For example, to unravel the processes involved in the recognition of EBL, the influence of face versus the body on the eye fixation and recognition behavior could be studied by using stimuli which combine incongruent face and body EBL stimuli. Employing this approach in a recent fMRI study, conducted by de Gelder's group [11], it was found that the ability to correctly recognize different facial expressions is strongly influenced by EBL. Another avenue for future research is to investigate the fixation and recognition behavior during perception of dynamic stimuli where information about the expressed emotion may be gathered by focusing on movement kinematics.

We also suggest that several of the widely used paradigms should be reevaluated with respect to EBL, e.g., the influence of affective priming on fixation and recognition behavior [13]; the influence of gender [2] and of the symmetry of different body parts [3] and their effects on the perception and fixation behavior during the observation of EBL.

Acknowledgements: This research was supported in part by the part by grants from the Frontier of Science Program HFSP-RGP0054/2004-C and FP6-2005-NEST-Path Imp 043403,. Tamar Flash is an incumbent of the Dr. Hymie Moross Professorial chair.

REFERENCES

[1] A. Atkinson, W. Dittrich, A. Gemmell, A. Young, Emotion perception from dynamic and static body expressions in point-light and full-light displays, *Perception,* 33: 717-746 (2004).

[2] L. Barrett, L. Robin, P. Pietromonaco K., Eyssell, Are women the "more emotional sex?" Evidence from emotional experiences in social context. *Cognition and Emotion*, 12: 555-578 (1998).

[3] W.Brown, L. Cronk, K. Grochow, A. Jacobson, C.K. Lu, Z. Popovic and R. Trivers. Dance reveals symmetry especially in young men. *Nature,* 438: 1148-1150 (2005).

[4] A. Camurri, B. Mazzarino, M. Ricchetti, R. Timmers, G. Volpe. Multimodal analysis of expressive gesture in music and dance performances. *"Gesture-based Communication in Human-Computer Interaction", LNAI 2915, Springer Verlag* (2004).

[5] C. Darwin, The Expression of the Emotions in Man and Animals ,Univ. Chicago Press (1965).

[6] B. deGelder, Towards the neurobiology of emotional body language, *Nature Reviews: Neuroscience,* 7: 242–249 (2006)

[7] B. de Gelder, N.Hadjikhani. Non-conscious recognition of emotional body language. *Neuroreport, 17*(6): 583-586 (2006).

[8] P. Ekman, Facial Expressions, *In T. Dalgleish and T. Power (Eds.) the Handbook of Cognition and Emotion., 301-320. Sussex, U.K.: John Wiley & Sons, Ltd.,* 1999.

[9] T. Flash, B. Hochner, Motor primitives in vertebrates and invertebrates, *Current Opinions in Neurobiology,* 15: 660-666(2005).

[10] A. Gee , R. Cipolla, Determining the gaze of faces in images. *Image and Vision Computing,* 12(10): 639-647 (1994)

[11] M. Mataric, Sensory-motor primitives as a basis for imitation: linking perception to action and biology to robotics. In: Nehaniv, C. and Dautenhahn, K., Editors, *Imitation in animals and artifacts*, MIT Press (2001).

[12] H. Meeren, C. van Heijnsbergen, B. deGelder, Rapid perceptual integration of facial expression and emotional body language. *Proc. Natl Acad. Sci. USA,* 102:16518–16523 (2005).

[13] S. Murphy, R. Zajonc, Affect, cognition, and awareness: Affective priming with optimal and suboptimal stimulus exposures. Journal of Personality & Social Psychology, 64(5), 723-739 (1993).

[14] Paterson H. M., Pollick F. E., Sanford A. J., The role of velocity in affect discrimination, *Twenty-Third Annual Conference of the Cognitive Science Society.*(2001)

[15] S. Ullman, M. Vidal-Naquet, E. Sali, Visual features of intermediate complexity and their use in classification, *Nature Neuroscience,* 57: 682–687 (2002).

[16] H.Wallbott, Bodily expression of emotion. *European Journal of Social Psychology,* 28: 879-896 (1998).