Perception of Emotion in Body Expressions from Gaze Behavior

Andrea Kleinsmith
University of Maryland Baltimore County
Baltimore, MD, USA
andreak@umbc.edu

Azin Semsar
University of Maryland Baltimore County
Baltimore, MD, USA
asemsar1@umbc.edu

ABSTRACT

Developing affectively aware technologies is a growing industry. To build them effectively, it is important to understand the features involved in discriminating between emotions. While many technologies focus on facial expressions, studies have highlighted the influence of body expressions over other modalities for perceiving some emotions. Eye tracking studies have evaluated the combination of face and body to investigate the influence of each modality, however, few to none have investigated the perception of emotion from body expressions alone. This exploratory study aimed to evaluate the discriminative importance of dynamic body features for decoding emotion. Eye tracking was used to monitor participants' eye gaze behavior while viewing clips of non-acted body movements to which they associated an emotion. Preliminary results indicate that the two primary regions attended to most often and longest were the torso and the arms. Further analysis is ongoing, however initial results independently confirm prior studies without eye tracking.

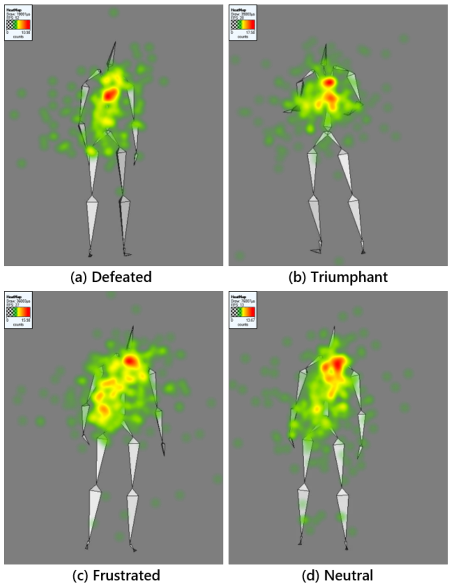

Figure 1: Heat maps of the body expression clips that achieved the highest percentage of observer agreement for each emotion category: (a) defeated; (b) triumphant; (c) frustrated; (d) neutral.

KEYWORDS
Eye tracking; Body expressions; Emotion.

INTRODUCTION

The last several years have seen a boom in emotion-aware technologies. Many aim to recognize the emotion displayed by the user from nonverbal behaviors such as facial expressions [12] [15] and body

CHI'19 Extended Abstracts, May 4–9, 2019, Glasgow, Scotland Uk

ACM ISBN 978-1-4503-5971-9/19/05…$15.00
https://doi.org/10.1145/3290607.3313062

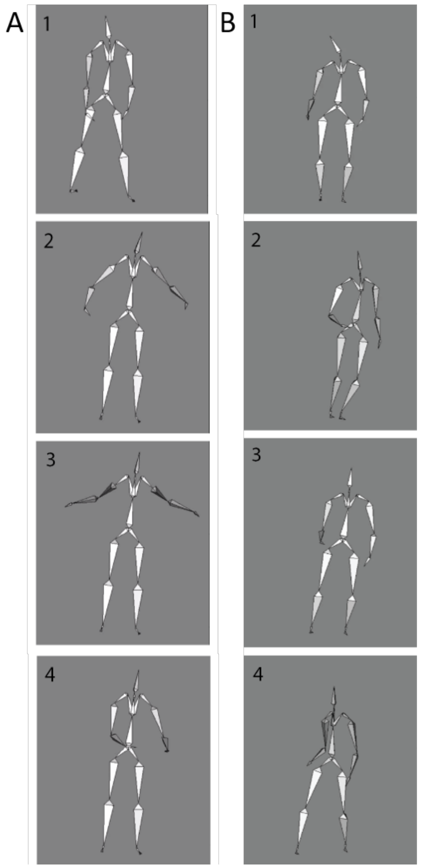

Figure 2: Two examples of the body expression clips used in the study illustrated by a sequence of four frames.

expressions [7] [8]. To do so requires an understanding of what should be recognized, i.e., which features or regions of the face or body are most important for discriminating between emotions. Eye tracking can be an effective method as it can provide an index of overt, yet unconscious attention. The goal of this exploratory study was to investigate the feasibility of using eye tracking to assess features relevant for perceiving emotion from non-acted, dynamic body expressions.

The body of literature using eye tracking to investigate emotion perceived from facial expressions alone is extensive, e.g., [6] [18] [14]. The majority of studies aimed to determine which facial regions are attended to by a general population (e.g., [6]), while other studies focused on clinical populations, such as those suffering from affective disorders (see [2] for a review). Eye tracking studies that include the body in combination with the face have been conducted by one research group from an affective neuroscience perspective. Their primary focus is understanding how each modality is processed when the two are either congruent or incongruent [10], when context information (i.e., a natural scene) is included [11], and when there is a threat [3]. To our knowledge, ours is the first eye tracking study to examine only body expressions.

The body alone is an important modality for perceiving emotion [8] with some emotions better perceived from body information than face information [4]. For instance, when there is uncertainty about the emotion displayed in facial expressions, it is the body that helps resolve the conflict [13] [19]. Our previous work modeled and evaluated the power of a set of low-level features for discriminating between four non-basic affective states [9]. The results revealed that two of the most important features were the torso and the lateral and vertical extension of the shoulders. Using eye tracking and the same body expressions, we attempt to confirm the previous findings.

MATERIALS AND METHODS

Participants: Thirty-two students (14 female) ranging in age from 18-32 participated in the study. It was approved by the Institutional Review Board at the University of Maryland, Baltimore County. All participants provided informed consent and were compensated for participation.

Body Expression Dataset: The naturalistic corpus of the UCLIC Database of Affective Postures and Body Movements was used [9]. This corpus comprises full body motion capture data recorded from 11 participants while they played Nintendo Wii video games and viewed replays of each point just played (refer to [9] for a complete description). To reduce potential bias, the motion capture data was mapped to a simplistic 3D avatar (see Figure 2). A total of 62 video clips were created. Each body expression ranged from 2 to 4 seconds. Given the relatively short duration, each video automatically played twice (i.e., each video was 4 to 8 seconds long). The body expressions were viewed at 628x840 and presentation order was counterbalanced.

Eye tracking: Participants' binocular eye gaze behavior was recorded with a Tobii X3-120 screen-based eye tracker (Tobii Pro, Sweden); sampling at 120 Hz and attached to the bottom of a 24-inch

Table 1: Fixation duration means (± standard deviations) for each emotion and each area of interest (AOI). D=defeated; F=frustrated, N=neutral; T=triumphant.

AOI	D	F	N	T
Head/ Shoulders	1.42 ±0.41	0.99 ±0.5	1.51 ±0.74	1.26 ±0.57
Arms	0.52 ±0.24	0.79 ±0.26	1.87 ±0.34	1.02 ±0.44
Legs	0.515 ±0.11	0.13 ±0.11	0.24 ±0.25	0.16 ±0.18
Torso	2.53 ±0.36	2.4 ±1.01	2.37 ±0.33	2.67 ±0.93

Table 2: Fixation count means (± standard deviations) for each emotion and each area of interest (AOI). D=defeated; F=frustrated, N=neutral; T=triumphant.

AOI	D	F	N	T
Head/ Shoulders	2.76 ±0.73	2.44 ±0.8	2.92 ±0.94	2.72 ±0.9
Arms	2.11 ±1	3.59 ±2.24	2.56 ±0.9	4.14 ±1.33
Legs	0.52 ±0.36	1.09 ±1.51	0.74 ±0.71	0.57 ±0.35
Torso	5.14 ±1.07	5.55 ±2.66	5.33 ±0.84	6.3 ±1.68

monitor. Tobii reports an average accuracy of 0.5°. Time, duration and location of all fixations were recorded. Four different areas of interest (AOI) were defined: head, torso, arms, and legs. We did not consider the left and right arm and leg as separate AOIs because we were not concerned with the side of the body on which participants' gaze was focused, thus one AOI was defined for both arms and one for both legs. All AOIs dynamically adjusted to the movement of that body region, meaning that as the leg moved, for instance, so did the AOI. Refer to Figure 1 to see a representative example of the heat map generated from all participants' eye gaze for each emotion.

Procedure: After providing consent, participants filled out a short demographic questionnaire and the Schutte Self-Report Emotional Intelligence Test (SSEIT) [16]. The eye tracker was calibrated and participants viewed the body expression clips. After each body expression, participants were asked to choose an emotion label from a list of four response alternatives (defeated, frustrated, triumphant, and neutral). The first three emotions are the same as those used in our prior work [9]. As suggested by anecdotal evidence from that study, neutral replaced concentrating as the fourth category. Using Likert scales, participants were also asked to rate the level of arousal (i.e., (1) calm to (5) excited) and valence (i.e., (1) very negative to (5) very positive). This paper discusses the analysis performed on the emotion categories only.

RESULTS AND DISCUSSION

Observer agreement

Observer agreement was computed and the most frequent emotion associated to each body expression was assigned as the ground truth label. The SSEIT was used to assess the participants' ability to discriminate between emotions. This was important given that non-acted expressions were used which are generally subtler and more complex than acted expressions. The average SSEIT score, 128.88 (±7.56; range 112-141), was within the normal range. A moderately high level of internal consistency was confirmed with Cronbach's alpha ($\alpha = 0.714$) [5]. All participants were included in the analyses. Above chance level agreement (i.e., $\geq$ 25%) was achieved for all body expressions, yielding 9 defeated, 9 frustrated, 16 triumphant, and 28 neutral. The highest average percentage of agreement was obtained for neutral (56% ± 12.7%) followed by triumphant (46.5% ± 10.7%), defeated (41% ± 6%) and frustrated (38.5% ± 5%).

It is not surprising that neutral obtained the largest number of body expressions and the highest observer agreement. This is similar to our previous findings for concentrating which observers used when the body postures did not seem to correspond to the other affective states [9]. The majority of disagreement occurred between triumphant and frustrated (8 body expressions) which could be due to the similarity in perceived arousal. Body expressions with larger and faster motions were often perceived as triumphant or frustrated. Slower body expressions with less movement were perceived

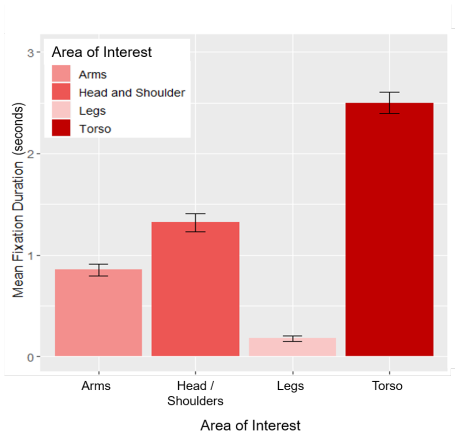

Figure 3: Mean fixation duration (seconds) for each area of interest and standard deviation of the mean for the emotional body expressions.

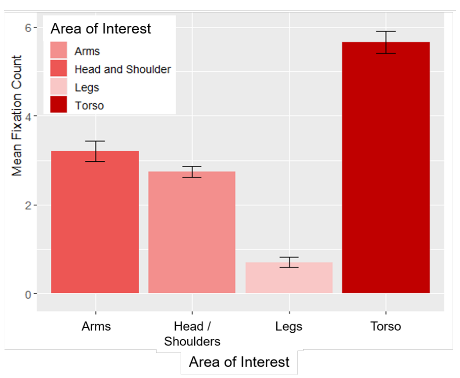

Figure 4: Mean fixation count for each area of interest and standard deviation of the mean for the emotional body expressions.

as defeated or neutral, however less disagreement occurred between these two emotions (5 body expressions). No disagreement occurred between triumphant and defeated.

Body expression clips with low observer agreement were removed prior to further analysis for two reasons. One, low agreement suggests that it was difficult to judge the emotion displayed, thus it would be difficult to draw reliable conclusions from the eye gaze behavior for these expressions. Two, the distribution of body expressions in each emotion was unbalanced, e.g., there are more neutral body expressions than any other category. Considering the primary emotions only, i.e., defeated, frustrated, and triumphant, the body expressions that achieved above chance level agreement ($\geq 33\%$) were retained. This resulted in 7 defeated, 8 frustrated, and 15 triumphant body expression clips. 15 neutral body expressions were also retained to match the highest number of expressions in the emotion categories.

Fixation behavior

Two-way ANOVAs were carried out for fixation duration and fixation count, separately, with a factor for AOI (head, torso, arms, legs) and a factor for emotion (defeated, frustrated, triumphant, neutral). Significant main effects were followed up with Bonferroni-corrected pairwise comparisons and interactions were followed up with Tukey's HSD post hoc tests. Statistical analysis was performed using R version 3.2.0 (R foundation for Statistical Computing, Austria). Refer to Tables 1 and 2 for mean fixation durations and mean fixation counts, respectively.

Fixation duration: There was a significant main effect for AOI, $F(3, 164) = 163.28, p < 0.0001$. There was no main effect for emotion $F(3, 164) = 0.89, p = 0.45$, and no interaction between AOI and emotion $F(9, 164) = 1.09, p = 0.37$. Follow-up analyses on AOI revealed significantly longer durations on the head and shoulders region compared to the legs ($p < 0.0000$) or arms ($p = 0.00016$) regions, while fixation durations on the arms were significantly longer than the legs ($p < 0.0000$). However, the longest fixation durations were on the torso region compared to all other regions (arms: $p < 0.0000$; head/shoulders: $p < 0.0000$; legs: $p < 0.0000$). Refer to Figure 3 for an overview of the results.

Fixation count: The interaction effect between AOI and emotion on fixation count was statistically significant $F(9, 164) = 1.98, p = 0.045$. There were main effects for AOI $F(3, 164) = 127.11, p < 0.0001$ (refer to Figure 4 for an overview of the results) and emotion $F(3, 164) = 3.51, p = 0.017$. Follow-up analyses for emotion revealed triumphant body expressions were attended to more often than defeated ($p = 0.002$), and significant effects were found for all emotions in the mean fixation count of the AOIs. For defeated, participants attended to the torso more often than any other AOI (arms: $p = 0.0007$; head/shoulders: $p = 0.027$; legs: $p < 0.0000$). The same result was found for triumphant, with the torso attended to most often (arms: $p = 0.0003$; head/shoulders: $p = 0.027$; legs: $p < 0.0000$). Arms ($p < 0.0000$) and head/shoulders ($p = 0.0003$) were both attended to more often than the legs. For frustrated, the torso was attended to more often than the head and shoulders ($p = 0.0001$) or legs

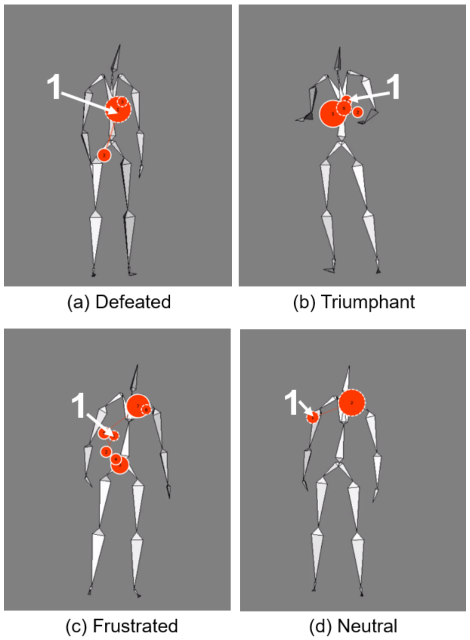

Figure 5: The scan path of one participant for each emotion. The first fixation location is highlighted. Fixation circle size corresponds to the fixation duration.

($p < 0.0000$). Arms were also attended to more often than legs ($p = 0.005$). Overall, the arms were attended to more often in triumphant ($M = 4.14 SD = \pm 1.33$) and frustrated ($M = 3.59 SD = \pm 2.24$) than in defeated ($M = 2.11 SD = \pm 1$) or neutral ($M = 2.56 SD = \pm 0.9$).

Results of the current study confirm our previous results [9] as well as findings from psychology, detailed in [8]. Overall, the results demonstrate that the frequency of fixations on head and shoulders and arms regions differed between 'active' (i.e., high arousal) and 'passive' (i.e., low arousal) emotions, with arms receiving more attention in triumphant and frustrated expressions, and the head and shoulders receiving more attention in defeated expressions.

The torso region was attended to most often and for the longest amount of time, regardless of emotion, and the legs were attended to the least. It is interesting to note that the torso was also an important discriminative feature in our previous study. Although the finding confirms the previous finding, we posit additional interpretations from an eye tracking perspective. First, as the torso was in the center of the screen, roughly, it is likely that participants' eyes returned to this position between body expressions while waiting for the next one to appear. Although we intentionally chose not to include a fixation cross prior to each presentation in an effort to keep viewing conditions natural and unconstrained [6], it is possible that the centered body expression affected fixations [1] [17]. Arizpe et al [1] found that start position greatly affected fixation patterns during a face recognition task. A center start position resulted in longer initial fixations, indicating more peripheral information may be processed. Analyzing the scan path, i.e., location of each subsequent fixation, may help us understand if our results are due to the same central viewing bias. While this analysis is still in the initial stages, the scan paths depicted in Figure 5 show that the first fixation location for each expression is near the center of the torso region, except in the case of neutral. Note that these are the same expressions shown in Figure 1. This leads to our second conclusion that the size of the body expression stimuli also may have affected fixation. The height of the body expression (and therefore, overall size) on the screen was limited by the height of the body with the arms fully extended above the head because some expressions contain this motion. Increasing stimulus size would require turning the monitor to a portrait orientation.

CONCLUSIONS AND FUTURE WORK

The present study confirmed that upper body features are important for discriminating subtle expressions of emotion; however, further investigation is necessary. For instance, the list of emotions may need to be expanded to include an 'other' category. In doing so, the number of expressions most frequently labeled neutral is likely to decrease. However, considering affective dimensions instead of discrete emotion categories may be a better solution. As suggested by [6], [8] and [9], arousal and valence ratings of the expressions should also be included. It is possible that the additional ratings could resolve ambiguities between some emotion expressions. For instance, frustrated and triumphant suggest similar levels of arousal, but different levels of valence, i.e., frustrated at the negative end of the dimension and triumphant at the positive end. Analysis on these ratings is currently underway.

REFERENCES

[1] Joseph Arizpe, Dwight J Kravitz, Galit Yovel, and Chris I Baker. 2012. Start position strongly influences fixation patterns during face processing: Difficulties with eye movements as a measure of information use. *PloS one* 7, 2 (2012), e31106.

[2] Thomas Armstrong and Bunmi O Olatunji. 2012. Eye tracking of attention in the affective disorders: A meta-analytic review and synthesis. *Clinical Psychology Review* 32, 8 (2012), 704–723.

[3] Rachel L Bannerman, Maarten Milders, Beatrice De Gelder, and Arash Sahraie. 2009. Orienting to threat: faster localization of fearful facial expressions and body postures revealed by saccadic eye movements. *Proceedings of the Royal Society of London B: Biological Sciences* (2009), rspb–2008.

[4] Beatrice De Gelder. 2006. Towards the neurobiology of emotional body language. *Nature Reviews Neurosci* 7, 3 (2006), 242.

[5] Robert F DeVellis. 2016. *Scale development: Theory and applications*. Vol. 26. Sage publications.

[6] Hedwig Eisenbarth and Georg W Alpers. 2011. Happy mouth and sad eyes: scanning emotional facial expressions. *Emotion* 11, 4 (2011), 860.

[7] Michelle Karg, Ali-Akbar Samadani, Rob Gorbet, Kolja Kühnlenz, Jesse Hoey, and Dana Kulić. 2013. Body movements for affective expression: A survey of automatic recognition and generation. *IEEE Trans on Affective Comp* 4, 4 (2013), 341–359.

[8] Andrea Kleinsmith and Nadia Bianchi-Berthouze. 2013. Affective body expression perception and recognition: A survey. *IEEE Transactions on Affective Computing* 4, 1 (2013), 15–33.

[9] Andrea Kleinsmith, Nadia Bianchi-Berthouze, and Anthony Steed. 2011. Automatic recognition of non-acted affective postures. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 41, 4 (2011), 1027–1038.

[10] Mariska Kret, Jeroen Stekelenburg, Karin Roelofs, and Beatrice De Gelder. 2013. Perception of face and body expressions using electromyography, pupillometry and gaze measures. *Frontiers in Psychology* 4 (2013), 28.

[11] Mariska Esther Kret, Karin Roelofs, Jeroen Stekelenburg, and Beatrice de Gelder. 2013. Emotional signals from faces, bodies and scenes influence observers' face expressions, fixations and pupil-size. *Frontiers in Hum Neurosci* 7 (2013), 810.

[12] Daniel McDuff, Abdelrahman Mahmoud, Mohammad Mavadati, May Amr, Jay Turcot, and Rana el Kaliouby. 2016. AFFDEX SDK: a cross-platform real-time multi-face expression recognition toolkit. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 3723–3726.

[13] Hanneke KM Meeren, Corné CRJ van Heijnsbergen, and Beatrice de Gelder. 2005. Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences* 102, 45 (2005), 16518–16523.

[14] Matthew F Peterson and Miguel P Eckstein. 2012. Looking just below the eyes is optimal across face recognition tasks. *Proceedings of the National Academy of Sciences* 109, 48 (2012), E3314–E3323.

[15] Evangelos Sariyanidi, Hatice Gunes, and Andrea Cavallaro. 2015. Automatic analysis of facial affect: A survey of registration, representation, and recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37, 6 (2015), 1113–1133.

[16] Nicola S Schutte, John M Malouff, Lena E Hall, Donald J Haggerty, Joan T Cooper, Charles J Golden, and Liane Dornheim. 1998. Development and validation of a measure of emotional intelligence. *Pers and Ind Diffs* 25, 2 (1998), 167–177.

[17] Benjamin W Tatler. 2007. The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of vision* 7, 14 (2007), 4–4.

[18] Avinash R Vaidya, Chenshuo Jin, and Lesley K Fellows. 2014. Eye spy: The predictive value of fixation patterns in detecting subtle and extreme emotions from faces. *Cognition* 133, 2 (2014), 443–456.

[19] Jan Van den Stock, Ruthger Righart, and Beatrice De Gelder. 2007. Body expressions influence recognition of emotions in the face and voice. *Emotion* 7, 3 (2007), 487.