What You See is What You Get:
Channel Dominance in the Decoding of Affective Nonverbal Behavior Displayed by Avatars

Presentation on the 66th Annual Conference of the International Communication Association (ICA), June 9-13 2016, Fukuoka, Japan.

Daniel Roth[1,2], Carola Bloch[1], Anne-Kathrin Wilbers[1], Marc-Erich Latoschik[2], Kai Kaspar[1], Gary Bente[1]

1 University of Cologne, Cologne, Germany
2 University of Würzburg, Würzburg, Germany

Citation:
Roth, D., Bloch, C., Wilbers, A.-K., Latoschik, M. E., Kaspar, K. & Bente, G. (2016). What You See is What You Get: Channel Dominance in the Decoding of Affective Nonverbal Behavior Displayed by Avatars. *Presentation at the 66th Annual Conference of the International Communication Association (ICA), June 9-13 2016, Fukuoka, Japan*.

Abstract

Nonverbal expressions of emotions play an important role in social interactions. Regarding virtual environments (VEs) and the transmission of nonverbal cues in avatar-mediated communication, knowledge of the contribution of nonverbal channels to emotion recognition is essential. This study analyzed the impact of emotional expressions in faces and body motion on emotion recognition. Motion capture data of expressive body movements from actors portraying either anger or happiness were animated using avatars with congruent and incongruent facial expressions. Participants viewed the resulting animations and rated the perceived emotion. During stimulus presentation, gaze behavior was recorded. The analysis of the rating results and visual attention patterns indicates that humans predominantly judge emotions based on the facial expression and pay higher attention to the head region as an information source to recognize emotions. This implicates that the transmission of facial expression is of importance for the design of social VEs.

Keywords: Avatars, Nonverbal Behavior, Affect, Virtual Environments, Eye Tracking

What You See is What You Get:
Channel Dominance in the Decoding of Affective Nonverbal Behavior Displayed by Avatars

Humans are equipped with the essential skill of decoding emotions from nonverbal cues (Boone & Cunningham, 1998). In classic theory, facial expressions are said to universally classify emotional content (Ekman, 1992). Matsumoto, Frank, and Hwang (2012) reviewed a set of studies and argue that “the estimated amount of information communicated nonverbally ranges between 65% and 95% of the total messages conveyed” (2012, p. 12). In a study by Metallinou, Busso, Lee, and Shrikanth Narayanan (2010) a 75% accuracy rate for happiness, 50-60% for anger, and 35% for neutrality were achieved., indicating the enormous information potential of facial expressions. Researchers also found that body movement is a qualitative indicator for emotions (Atkinson, Dittrich, Gemmell, & Young, 2004; Coulson, 2004). Furthermore, match and mismatch combinations of angry/happy body postures and facial expressions indicated that congruent body motion can improve the recognition rate, while conflicting body motion biases the observer towards the emotion conveyed by the body movement when judging facial expressions (Meeren, van Heijnsbergen, & de Gelder, 2005).

Putting both indicators together, recent findings suggest that “body cues, not facial expressions discriminate between intense positive and negative emotions” (Aviezer, Trope, & Todorov, 2012).

Methodologically, it was shown that virtual faces are comparably recognizable as natural ones (cf. de Borst & de Gelder, 2015). Additionally, McDonnel, Jörg, McHugh, Newell, and O’Sullivan (2008) showed that affective movements are similarly perceived between humans and human as well as non-human avatars. In Virtual Environments (VEs), humans are typically represented as an avatar, a “perceptible digital representation whose behaviors reflect those executed, typically in real time, by a specific human being" (Bailenson & Blascovich, 2004, p. 65). VEs and Virtual Reality (VR) therefore make use of avatar representations to communicate messages between users or between users and agents. However, this rationale can also be inverted 1) to study humans using VEs (Biocca, 1992; Blascovich et al., 2002; Fox, Arena, & Bailenson, 2009) or controllable avatar stimuli from analyzable human motion (Bente, Krämer, Petersen, & de Ruiter, 2001; Bente, Senokozlieva, Pennig, Al-Issa, & Fischer, 2008) and 2), to study the importance of social channels for virtual encounters in VEs and VR (Hoorn, Konijn, & Van der Veer, 2003). The present study investigates the channel dominance in the decoding of emotions from matched and mismatched nonverbal behavior of the body and the face. The results give both guidelines for future VE/VR development as well as transformable results for social and communication psychology.

Method

We used a combination of captured body motions and predefined facial expressions in order to assess dominance in the perception of affective signals of anger and happiness. Male actors displayed emotions based on certain scenario information (e.g. your favorite football team won/lost in a soccer game). Body motion was recorded to animate avatars for stimulus creation. In order to select appropriate body movements and facial expressions, two pre-studies were initially conducted.

In a first pre-study, a sample of 43 participants (M_{age}=24.23) observed 45 different avatars. We used neutral wooden mannequins with no face in order not to bias the impression of the body movements (see Figure 1). For each avatar, participants rated the degree to which it expressed happiness or anger, respectively, using a 7-point scale (1=very happy, 7=very angry). Based on these results, we selected five stimuli perceived as "angry" (M=3.13, *SD*=1.09), five “neutral” stimuli (*M*=3.99, *SD*=0.81), and five “happy” stimuli perceived as “happy” (*M*=4.73, *SD*=1.81). Wilcoxon pairwise test showed that these three emotion categories were significantly different (all $p < .05$). The length of the clips ranged from three to eight seconds. The mean duration of the clips was not significantly different between categories (χ^2 (2)=1.78, *p*=.41).

In a second pre-study (*N*=47, M_{age}=27.23) we identified the most suitable facial expressions based on the same 7-point bipolar rating scale (1=very happy, 7=very angry). The facial expressions (3 per affective state) were pre-animated using blend shapes to come close to the Facial Action Coding System (Ekman, Friesen, & Hager, 2002). The mean rating of the finally selected facial expression was *M*=2.32 (*SD*=0.94) for the angry face, *M*=5.53 (*SD*=0.78) for the happy face, and *M*=4.11 (*SD*=0.6) for the neutral face (see Figure 2). We chose a neutral version with a slightly happily rated face in order to compensate for the negativity bias (Rozin & Royzman, 2001).

In the main study, the pre-selected body movements (five per emotion category) and facial expressions (one per emotion category) were combined, resulting in a 3 (facial expression: happy vs. angry vs. neutral) x 3 (body movement: happy vs. angry vs. neutral) within-subjects design and 45 avatar videos. Sixty-eight participants (M_{age}=24.71) observed all avatar videos while their eye movements were recorded by means of a SMI RED-500 eye-tracker. After each video, they rated their impression on a 7-point scale (1=very happy, 7=very angry). For each avatar, we dynamically coded areas of interest (AOIs) for the head and the body using rectangles that followed the body part movements.

Results

Subjective Ratings

For the sake of facilitated interpretation, we initially transformed the subjective ratings to a scale ranging from -3 (angry) to +3 (happy). We conducted a 3x3 ANOVA (facial expression x body movement) for repeated measures. The ANOVA showed a main effect of facial expression, F (1.59, 106.4)=664.11, η_p^2=.91, p < .001, and a main effect for body movement, F (1.47, 98.71)=4.04, p=.032, η_p^2=.06. Furthermore, an interaction effect was found, F (3.25, 217.87)=23.94, p < .001, η_p^2=.26. Pairwise comparisons showed that all categories of facial expressions were rated significantly different ($MD_{angry-neutral}$=-.72, *SE*=.05, p < .001; $MD_{angry-happy}$=-2.23, *SE*=.08, p < .001; $MD_{neutral-happy}$=-1.50, *SE*=.06 p < .001). Pairwise comparisons for body movement showed significant differences between angry body movements and neutral body movements, $MD_{angry-neutral}$=-.18, *SE*=.04, p < .001. However, neither did we find a difference between angry body movements and happy body movements, $MD_{angry-happy}$=-.09, *SE*=.08 p=.72, nor between neutral body movements and happy body movements, $MD_{neutral-happy}$=.09, *SE*=.06, p=.50 (see Figures 3 and 4).

Eye Tracking Analysis

To analyze gaze data, we used dynamic AOIs for both the head region and body region. A first preliminary analysis and visual inspection revealed that gaze focus was not picked up by exact polygonal AOI definitions. We therefore recoded rectangular AOIs with minor tolerance for both head and body region (See Figures 5 and 6). To account for minor differences in video length, we calculated relative dwell times (dwell time AOI $video_x$ /$video_x$ duration). We excluded cases in which either the pre- or the post-experimental validation error was greater than 1.5°, leading to a reduced sample size of N=60.

Dwell Time. A 2 (AOI: head versus body) x 3 (body movement) x 3 (facial expression) ANOVA showed a main effect of AOI, $F(1, 59)=13.01$, $p=.001$, $\eta^2_p=.18$, and of body movement, $F(2, 118)=78.95$, $p < .001$, $\eta_p^2=.57$. Pairwise comparisons showed that the observers' attention focused longer on the head area ($M_{DThead=}$.55 equaling 55% of stimulus time, $SE=.03$) in comparison to the body area ($M_{DTbody=}$.37 equaling 37% of stimulus time, $SE=.03$), $MD_{DTbody\text{-}DThead}=-.18$, $SE=.05$, $p=.001$. Mean relative dwell time was highest for happy movements ($M_{DThappyBM=}$.49, $SE=.01$), and angry movements still had a higher dwell time ($M_{DTangryBM}=.46$, $SE=.01$) than neutral movements ($M_{DTneutralBM}=.42$, $SE=.01$). All movements differed significantly from each other in relative dwell times (all ps <.001). Dwell times for facial expression showed no significant difference, $F(1.69, 99.82)=1.19$, $p=.308$, $\eta_p^2=.020$. No significant interaction effects were found for dwell time analysis.

Number of Fixations. We calculated the same 2x3x3 ANOVA for the (non-relative) number of fixations. The analysis showed a main effect for movement, $F(2,118)=385.9$, $p < .001$, $\eta_p^2=.87$. A happy movement led to the highest fixation rate ($M_{FhappyBM=}$7.9, $SE=.20$), followed by the angry body movement's fixation rate ($M_{FangryBM=}$5.4, $SE=.22$) and the neutral body movement's fixation rate ($M_{FneutralBM=}$4.42, $SE=.16$), all ps <.001. Additionally, the analysis showed an AOI by body movement interaction, $F(1.79,105.36)=4.33$, $p=.02$, $\eta_p^2=.07$ (see Figure 7). In both AOIs, all movements differed significantly from each other (all ps <.001). There were significantly more fixations when a happy body movement was present (Head: $M_{FhappyBM=}$8.43, $SE=.36$, Body: $M_{FhappyBM=}$7.38, $SE=.33$) compared to angry movements (Head: $M_{FangryBM=}$5.9, $SE=.39$, Body: $M_{FangryBM=}$4.92, $SE=.35$). The least fixations were present in both AOIs with neutral movements (Head: $M_{FneutralBM=}$4.57, $SE=.32$, Body: $M_{FneutralBM=}$4.3, $SE=.32$). The head AOI tended to have an overall higher amount of fixations ($M_{Fhead=}$6.3, $SE=.34$) in comparison to the body AOI ($M_{Fbody}=5.52$, $SE=.32$), but the difference was not significant, *p*=.16. Means and standard deviations of each condition are depicted in Table 1.

Discussion

The present study aimed at investigating channel dominance of facial expression and body movement. The results suggest that humans judge affective nonverbal behavior of face and body primarily with a higher focus on the facial expression and head area. Within all conditions, facial expression was the predominant cue, as indicated by rating results, longer dwell times as well as more fixations on the head AOI.

In our results, body movement was not a qualitative factor in judging emotional displays. The interaction effect between body motion and facial expression however shows that observers judged more extreme, with a tendency that relied mostly on the facial expression, with congruent body motion*facial expression combinations, but also incongruent emotional displays (happy*angry, angry*happy). Our results show, that when the face displayed happiness, avatars were rated the most positive, irrespective of the body movement, and when the face displayed anger, the stimuli were judged more negatively. This could be interpreted in a way that facial expressions are perceived as more clear identifiers of emotions, whereas the body movements used in this study contained a certain ambiguity. This being said, relying more on facial expressions, an incongruent but emotional body movement may be misinterpreted/irritating due to high emotional intensity or in informativeness. Moreover, happy and angry body movements are characterized by higher movement complexity and motion energy (e.g. shaking arms and more extensive locomotion). This interpretation may be supported by the gaze data, as mean relative dwell time for both AOIs was by trend lowest in neutral movements. Similarly, the amount of fixations was higher in both defined emotional displays, than in the neutral condition.

Limitations

First, the facial expression display used in the study was a static and artificially pre-defined expression. Second, although anger and happiness can be classified to lay contrary on the valence dimension (Fontaine, Scherer, Roesch, & Ellsworth, 2007), it is unclear how these results can be generalized for other emotions. Third, the actors who displayed the emotions were male, which means that results are limited to male body movements and to instructed contexts rather than spontaneous emotions. Furthermore, our sample was a typical student sample and gender was not normally distributed which may had impact on the results. For example, Hall and Matsumoto (2004) showed that female participants were more accurate in judging nonverbal communicated emotions and Krüger and others (2013) found gender specific differences when decoding emotional biased walking from male and female actors.

Conclusion

The present study showed that emotional displays of the face and body of avatars are primarily judged by the facial expression. The results therefore implicate that within VEs and VR, displaying facial expression may have higher importance than body movement to transmit emotions and therefore may increase affective understanding in avatar-mediated communication, which should be considered in future social VE/VR designs. As a transfer to traditional social psychology, the results are mainly in line with the presented literature. However, we did not find body motion to be a qualitative factor when discriminating affective displays, nor did it clearly bias the perceived affect to the displayed (body) emotion.

References

Atkinson, A. P., Dittrich, W. H., Gemmell, A. J., & Young, A. W. (2004). Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception-London*, *33*(6), 717–746.

Aviezer, H., Trope, Y., & Todorov, A. (2012). Body Cues, Not Facial Expressions, Discriminate Between Intense Positive and Negative Emotions. *Science*, *338*(6111), 1225–1229. http://doi.org/10.1126/science.1224313

Bailenson, J., & Blascovich, J. (2004). Avatars. In \iEncyclopedia of Human-Computer Interaction (pp. 64–68). Berkshire Publishing Group.

Bente, G., Krämer, N. C., Petersen, A., & de Ruiter, J. P. (2001). Computer animated movement and person perception: Methodological advances in nonverbal behavior research. *Journal of Nonverbal Behavior*, *25*(3), 151–166.

Bente, G., Senokozlieva, M., Pennig, S., Al-Issa, A., & Fischer, O. (2008). Deciphering the secret code: A new methodology for the cross-cultural analysis of nonverbal behavior. *Behavior Research Methods*, *40*(1), 269–277. http://doi.org/10.3758/BRM.40.1.269

Biocca, F. (1992). Communication within virtual reality: Creating a space for research. *Journal of Communication*, *42*(4), 5–22.

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, *13*(2), 103–124.

Boone, R. T., & Cunningham, J. G. (1998). Children's decoding of emotion in expressive body movement: The development of cue attunement. *Developmental Psychology*, *34*(5), 1007–1016. http://doi.org/10.1037/0012-1649.34.5.1007

Coulson, M. (2004). Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence. *Journal of Nonverbal Behavior*, *28*(2), 117–139.

de Borst, A. W., & de Gelder, B. (2015). Is it the real deal? Perception of virtual characters versus humans: an affective cognitive neuroscience perspective. \iFront. Psychol. 6: 576. Doi: 10.3389/fpsyg.

Ekman, P. (1992). An argument for basic emotions. *Cognition & Emotion*, *6*(3-4), 169–200.

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). FACS manual. *A Human Face*.

Fontaine, J. R., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The world of emotions is not two-dimensional. *Psychological Science*, *18*(12), 1050–1057.

Fox, J., Arena, D., & Bailenson, J. N. (2009). Virtual Reality: A Survival Guide for the Social Scientist. *Journal of Media Psychology: Theories, Methods, and Applications*, *21*(3), 95–113. http://doi.org/10.1027/1864-1105.21.3.95

Hall, J. A., & Matsumoto, D. (2004). Gender differences in judgments of multiple emotions from facial expressions. *Emotion*, *4*(2), 201.

Hoorn, J. F., Konijn, E. A., & Van der Veer, G. C. (2003). Virtual reality: Do not augment realism, augment relevance. *Upgrade-Human-Computer Interaction: Overcoming Barriers*, *4*(1), 18–26.

Krüger, S., Sokolov, A. N., Enck, P., Krägeloh-Mann, I., Pavlova, M. A., & Lappe, M. (2013). Emotion through Locomotion: Gender Impact. *PLoS ONE, 8*(11), e81716.

Matsumoto, D., Frank, M. G., & Hwang, H. S. (2012). Reading People. *Nonverbal Communication: Science and Applications: Science and Applications*, 1.

McDonnell, R., Jörg, S., McHugh, J., Newell, F., & O'Sullivan, C. (2008). Evaluating the emotional content of human motions on real and virtual characters. In *Proceedings of the 5th symposium on Applied perception in graphics and visualization* (pp. 67–74). ACM.

Meeren, H. K., van Heijnsbergen, C. C., & de Gelder, B. (2005). Rapid perceptual integration of facial expression and emotional body language. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(45), 16518–16523.

Metallinou, A., Busso, C., Lee, S., & Narayanan, S. (2010). Visual emotion recognition using compact facial representations and viseme information. In \iAcoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on (pp. 2474–2477). IEEE.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*(4), 296–320.

Figures

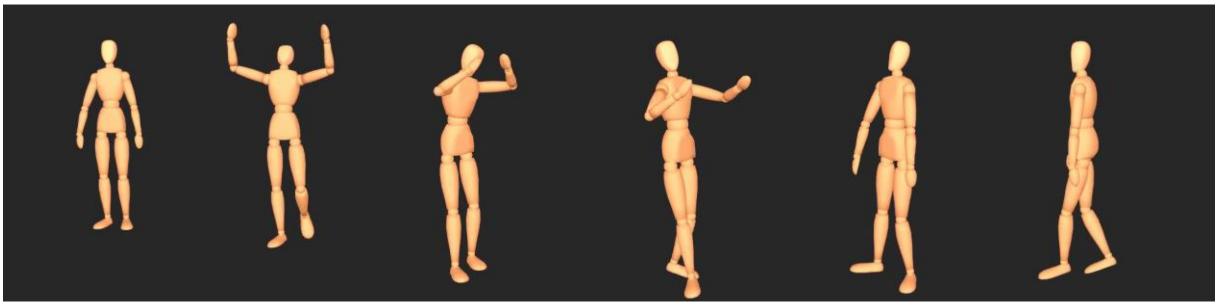

Figure 1. *Exemplary example of the behavioral animations used for pretesting body movement.*

Figure 2. *The selected male avatar and the selected facial expressions for anger (left), neutral (center), happy (right).*

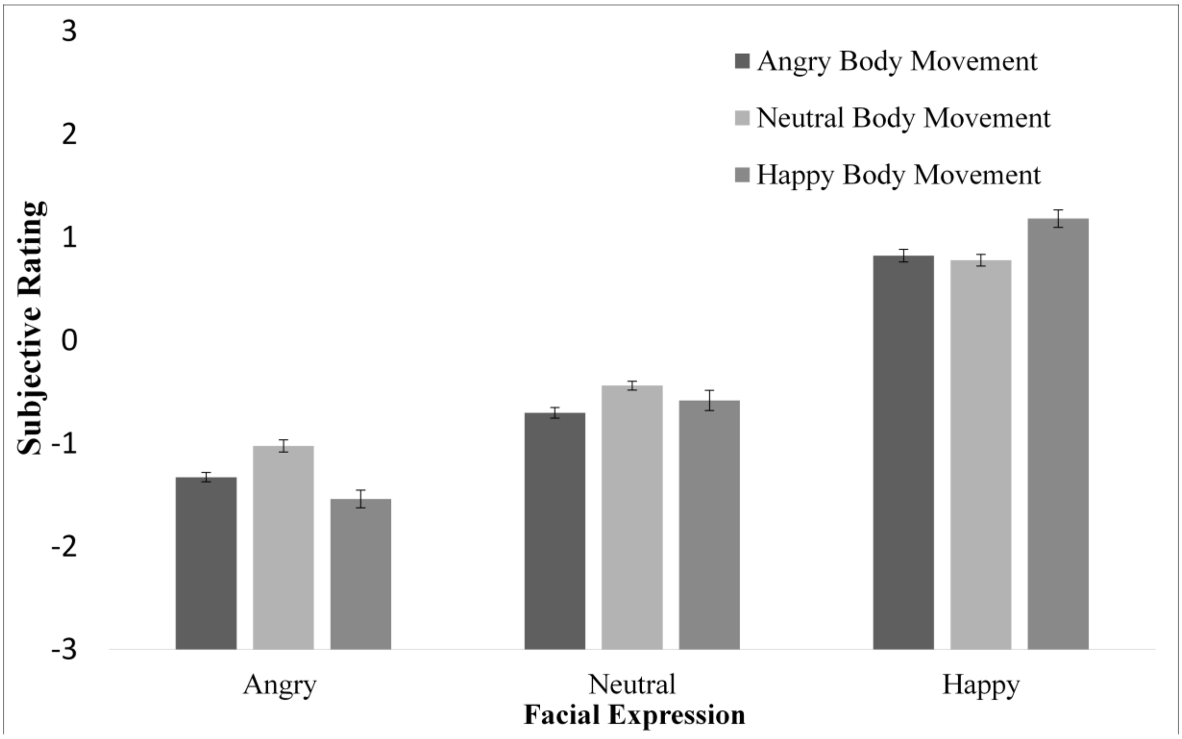

Figure 3. *Mean subjective rating results of the main study for the nine stimulus conditions (body movement x facial expression). The vertical scale ranged from -3=very angry to 3=very happy. Error bars denote the standard error around the mean.*

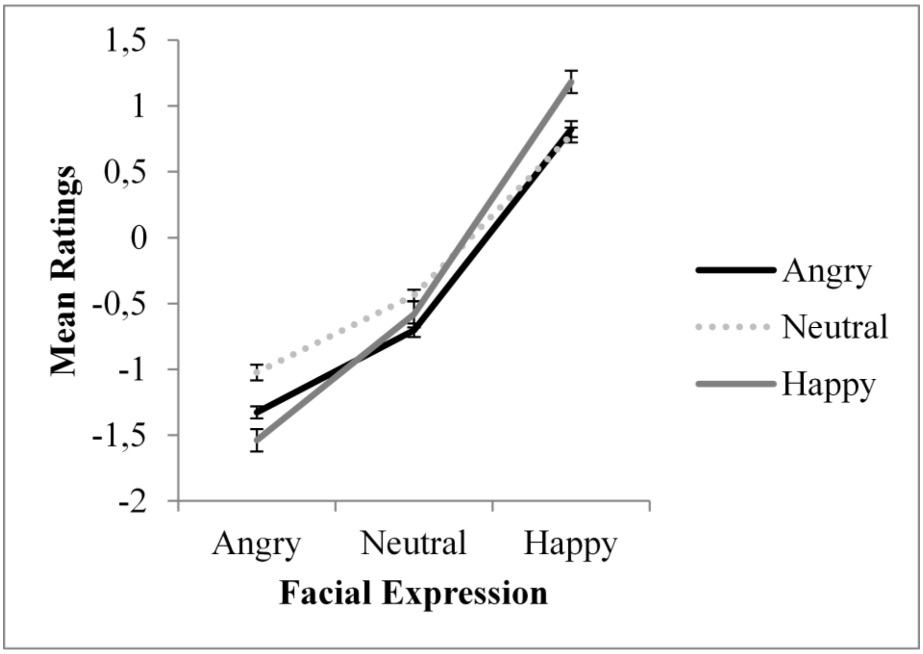

Figure 4. *Interaction between facial expression and body movement. The vertical scale ranged from -3=very angry to 3=very happy. Error bars denote the standard error around the mean.*

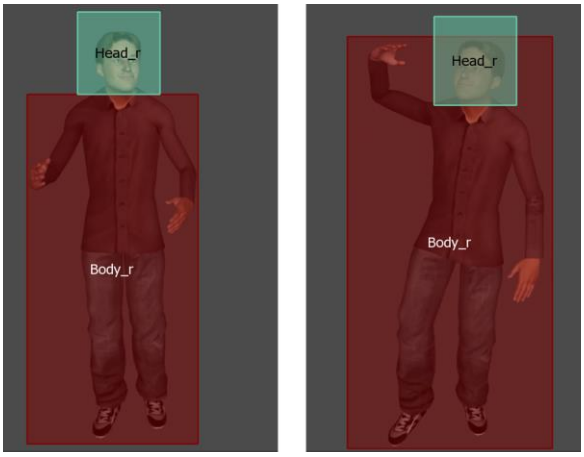

Figure 5. *Two exemplary frames showing the defined dynamical head and body AOIs used for eyet-tracking analysis.*

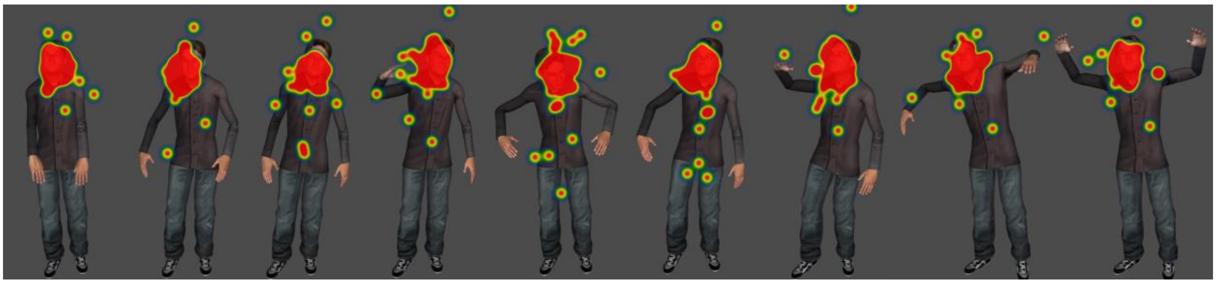

Figure 6. *Heat map of the users' focus in multiple frames of an exemplary stimulus.*

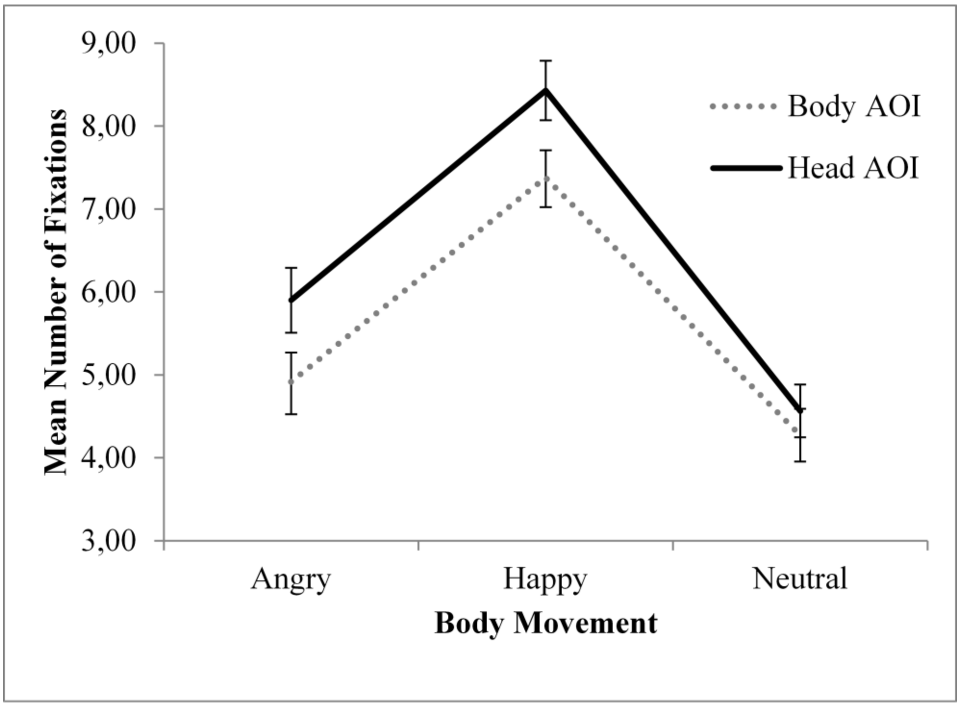

Figure 7. *Interaction between body movement and AOI. Error bars denote the standard error around the mean.*

Displayed Emotion		**Gaze Variables**			
Body Movement	**Facial Expression**	**Number of Fixations**		**Relative Dwell Time**	
		Head	Body	Head	Body
		Mean (SD)	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*
Anger	*Anger*	5.73 (3.07)	4.83 (2.77)	.56 (.25)	.37 (.20)
	Happiness	5.85 (3.45)	5.04 (2.92)	.55 (.27)	.36 (.21)
	Neutral	6.11 (3.33)	4.88 (2.96)	.55 (.26)	.36 (.22)
Happiness	*Anger*	8.21 (2.98)	7.52 (2.99)	.58 (.23)	.40 (.17)
	Happiness	8.34 (3.12)	7.17 (2.87)	.58 (.25)	.39 (.18)
	Neutral	8.73 (3.08)	7.44 (2.62)	.59 (.22)	.40 (.17)
Neutral	*Anger*	4.79 (2.80)	4.27 (2.46)	.50 (.26)	.33 (.23)
	Happiness	4.46 (2.65)	4.29 (2.87)	.50 (.27)	.33 (.23)
	Neutral	4.45 (2.42)	4.26 (2.53)	.50 (.26)	.37 (.24)

Table 1. *Means and standard deviations of gaze variables for each condition.*