Evaluating the emotional content of human motions on real and virtual characters

Rachel McDonnell[1] Sophie Jörg[1] Joanna McHugh[2] Fiona Newell[2] Carol O'Sullivan[1]

[1*] Graphics Vision & Visualisation Group and [2†] Institute of Neuroscience, Trinity College Dublin.

Figure 1: *Image taken from: real video, high resolution virtual male, low resolution virtual male, wooden mannequin, toon, zombie.*

Abstract

In order to analyze the emotional content of motions portrayed by different characters, we created real and virtual replicas of an actor exhibiting six basic emotions: sadness, happiness, surprise, fear, anger and disgust. In addition to the video of the real actor, his actions were applied to five virtual body shapes: a low and high resolution virtual counterpart, a cartoon-like character, a wooden mannequin, and a zombie-like character (Figure 1). Participants were asked to rate the actions based on a list of 41 more complex emotions. We found that the perception of emotional actions is highly robust and to the most part independent of the character's body.

CR Categories: I.3.7 [Computer Graphics]: Three Dimensional Graphics and Realism—Animation;

Keywords: perception, motion-capture

1 Introduction

Virtual characters in animated movies and games can be very expressive and have the ability to convey complex emotions. However, it has been suggested that the more anthropomorphic a humanoid robot or virtual character becomes, the more eerie it is perceived to be [Mori 1970]. In this paper, we focus on one potential factor that could affect this 'uncanny valley' theory, by examining how emotional actions on real and virtual characters are perceived. We ask the following question: are real and virtual characters moving with the same biological motion perceived to be exhibiting the same emotions?

*e-mail:{Rachel.McDonnell, Sophie.Joerg, Carol.OSullivan}@cs.tcd.ie
†e-mail:{mchughje, fnewell}@tcd.ie

Evidence from neuroscience studies suggest that different neural networks are activated when viewing real or virtual stimuli [Perani et al. 2001]. Therefore, we chose as our stimuli both real and virtual versions of humanoids acting out six basic emotions: sadness, happiness, surprise, fear, anger and disgust. Our 'real' stimulus is a video of an actor whose motion was captured while acting out the six emotions. We created a highly realistic virtual counterpart along with a lower resolution version (by reducing the number of polygons in the model), which were both animated using the original captured motion as depicted in the video. Furthermore, we explored how emotions were perceived when the same actions were applied to anthropomorphic but non-human characters. A cartoon-like figure was chosen in order to determine if his friendly appearance would lead him to be perceived as exhibiting more positive emotions, whereas we felt that a zombie-like character would be perceived in a more negative light. A wooden mannequin model was chosen as a completely neutral character that we felt may be perceived to be exhibiting less intense emotions than the others. In all cases, we wished to examine the effects of body representation and motion, and did not want to confound this with expressive facial or hand movements. Hence, the faces and hands were blurred in all videos.

We showed the six videos to six groups of participants, who were asked to rate them according to a list of complex emotions. We found that body representation had little or no effect on people's perception of the emotional content of the scene, which leads us to conclude that: a) relative to real footage, captured body motion is as effective at depicting emotional body language and b) people's perception of emotional motion is very robust and unaffected by a character's physical appearance.

2 Background

It has been shown that body movements and static postures are sufficient to convey at least the basic emotions. The responses to experiments which aim to recognize emotions from body motions are highly consistent [Wallbott 1998; Coulson 2004; Atkinson et al. 2004; Crane and Gross 2007].

Much research has been conducted on the perception of human motion and emotions. It is widely accepted that people can recognize human motion from even a very small set of cues [Johansson

1973]. Atkinson et al. [2004] showed that a small number of cues are enough to recognize every emotion with an above chance rate for point-light animations. They confirmed that emotions can be recognized from body motion even when static form is minimized by use of point-lights. Moreover, they found that the rated intensity of the motion depends more on the motion than on the shape, since point-light motions and real footage were equally rated.

Pasch and Poppe [2007] compare the interpretation of static body postures for a realistic virtual character and a mannequin similar to the wooden mannequin in our study. They show that realism has an effect on the perception of body postures representing the basic emotions, with very different results for each emotion. However, there is no consistent result that would show which character is more qualified to convey emotions.

Evidence from the fields of psychology and neuroscience has shown that different neural networks are activated when presented with real and virtual stimuli. For example, Perani et al. [2001] demonstrated this when participants viewed sequences of a real hand and virtual reproduction. They also found a limited effect due to the degree of realism of the reproduction. Han et al. [2005] analyzed the differences in brain activity when participants viewed cartoons or movie clips. Findings from this study suggest that the human brain functions in a different way when interacting with real people in everyday life than with artificial characters or static social stimuli. More recently, Mar et al. [2007] conducted a study using a set of stimuli derived from the film *Waking Life*. Participants saw identical biological motion for animated and live-action footage. Using BOLD (blood oxygenation level dependent) signals as a measure of neural activation, a higher level of activity was found in visual processing areas when presented with real stimuli. Chaminade et al. [2007] investigated how the appearance of computer animated characters influenced the perception of their actions. They found that the perceived biological nature of a motion decreased with characters' anthropomorphism.

Other studies have shown that people can respond socially to human and nonhuman entities [Reeves and Naas 1996; Slater and Steed 2002] and that they can engage with virtual humans whether or not they look human [Nowak and Biocca 2003].

Computer animation researchers have also analyzed the interactions between motion and the model it is applied to. For example, Hodgins et al. [1998] and McDonnell et al. [2005] performed perceptual experiments, the results of which indicated that a viewer's perception of motion characteristics is affected by the geometric model used for rendering. In [McDonnell et al. 2007], we showed that sex perception of virtual characters was influenced by the body shape and also by the motion that was applied to it.

3 Video Stimuli

Following a pilot video capture shoot using non-actors, it was deemed necessary to use an actor due to the lack of expressiveness of the emotions portrayed by the former. Three naïve male actor volunteers participated in a motion capture session, which was conducted on a 10 camera Vicon optical system, using 41 markers placed according to the Vicon human template. The capture environment consisted of a blue backdrop 3m high x 2m wide, and a blue floor area, 4.5m long x 2m wide, placed inside the capture volume of the motion capture system. The actor was centered within this area for the duration of the recordings. He was instructed to act out each of the 6 Ekman basic emotions [Ekman 1992] in turn: sadness, happiness, disgust, surprise, fear, and anger. We recorded up to ten examples of each emotion.

The video stimuli were recorded during the motion capture session, to ensure that identical video and virtual stimuli could be achieved. The video camera used was a JVC Digital Mini-DV camera (Model Number GR-DVL-167EK, with 520 lines of resolution, using PAL color encoding system). It was mounted on a tripod stand (at a height of 1.1m) at a consistent distance of 5.5m between the camera and the backdrop, and 3.75m between the camera and the actor. These measurements were recorded in order that virtual replicas of the video stimuli with identical dynamic information could be obtained.

Using a commercial video editor, the background was removed from the video stimuli, and brightness and contrast were optimized. Three examples of each emotion were chosen as candidates for the final set of emotions.

3.1 Choosing suitable actor and emotions

A basic categorization task was used in order to determine which of the examples of each emotion performed by each actor were most recognizable. Ten participants (3M-7F) took part in the study. A six alternative forced choice paradigm was employed, where participant's viewed the stimuli and indicated which emotion they thought was being expressed; sadness, happiness, anger, fear, surprise or disgust. Using an ANOVA, first we chose the actor with the best overall performance. We then chose one example from each of his acted emotions with the highest rate of recognition. Rates ranged from 87 to 100% identification accuracy. The emotions selected ranged in duration from 3 to 5 seconds, so we selected the most expressive 3 seconds from each emotion as our final set.

3.2 Blurring faces and hands

Since we were only concerned with, and therefore only captured the body motion of the actor, the face and hands of the actor in the videos were blocked out. We chose to blur the pixels of the face and hands, since this did not substantially interfere with the content of the movies. The motion of the face and hand blockers was obtained in 3D Studio Max by scripting an algorithm that automatically followed the motions of the virtual character's face and hands. The blockers obtained in 3D Studio Max were then applied to the real videos in a commercial video editor which blurred the correct areas of the video sequence.

4 Virtual Stimuli

Six body representations were used to display the motion in this experiment. The video clips of the real actor were used since we wished to test if participants would be more responsive to emotions displayed by a real human. Two different virtual human models were chosen to display the captured emotions. The first model was a high detailed virtual male model (*HiMale*), with a similar build and height as the captured actor. The model wore a black suit with sensors attached and a white face mask, similar to the real actor. This model had 10,000 polygons. The second model was identical to the first, but at a much lower resolution of just 1000 polygons (*LoMale*). Two different rendering styles were employed in order to increase the difference in realism between the models. The *HiMale* was rendered using a rendering engine for generating photorealistic images. Soft shadows and motion blur were incorporated to increase the realism. The *LoMale* was rendered in a style more commonly used in real-time applications such as games, with Gouraud shading and no motion blur or shadows.

In order to examine the effect of non-human body representations on our ability to perceive emotion, a further three models were chosen. We first chose a *Zombie* model that we felt might appear eerie when human motion was applied to it. This model had a decomposing humanoid body with bones, flesh, and organs visible (Figure 1 (far right)). We then chose a *Toon* model which could be perceived as cute and non-threatening. Finally, a neutral *Wooden* mannequin was chosen as we wanted a model that had neither negative nor positive connotations. We hypothesized that the zombie would increase participant's ratings of the negative emotions such as rage and horror, and that the toon would increase ratings of the positive emotions, such as affection and calmness. Finally, we hypothesized that the neutral character would have no influence on ratings. All three models were rendered to the same quality as *HiMale*.

Using the measurements taken at the time of capture, we recreated the environment in 3D Studio Max. The six motion-captured emotions corresponding to the 3 second clips chosen in Section 3.1 were applied to all five virtual characters. The video stimuli were captured using PAL at 25 frames per second, so we insured that the virtual models were also rendered at this rate. All virtual stimuli also had the face and hands blurred, in order to match the real videos and to ensure that their static facial expressions would not influence the perception of their emotion. (See color plate for examples of video and virtual stimuli).

5 Method and Participants

Rating scales are commonly used in psychology in order to gather information on subjective conditions such as the attitudes and emotions of participants, which could not be collected using direct measures. We chose this technique as we wished to collect the subjective judgements of participants with respect to the emotions that were being conveyed by the different bodies. We compiled a list of 41 simple and complex emotions and asked participants to rate the intensity of each of these emotions that the character was displaying. The list included the 6 basic Ekman emotions so that we could test the intensity ratings for the actual emotions as well as for more complex emotions. (The full list of emotions is available online at *http://gv2.cs.tcd.ie/mcdonner/APGV08.htm*).

We used a between-groups design for this experiment, where each group viewed a different body shape. This paradigm was used as we felt that, if allowed to view all of the body shapes, participants would have been more conscious of body shape and altered their ratings accordingly. Sixty-six participants (11 per group) took part in this experiment (44M, 22F), ranging in age from 19 to 45 and from different educational backgrounds. All participants were naïve as to the purpose of the experiment and were given book vouchers as a reward for participation.

The experiment was displayed on a 24-inch flat screen LCD monitor. Participants first viewed a static image of the body shape that they would be viewing for the experiment and were asked to describe the appearance of the character in as much detail as possible. They then viewed the character displaying one of the 6 basic emotions and were allowed to repeat that motion on the movie player as often as they liked. They make judgements about the level of intensity of each of the 41 listed emotions that they felt the character was displaying. The scale ranged from 'not at all' to 'extremely' on a 10 point scale. The list of 41 emotions was randomly ordered for each participant to avoid ordering effects and the order in which they viewed the 6 actions was also random.

6 Results

Descriptions of the static bodies were first assessed, but minimal insight was gained. The three male characters were simply described as being male, with little reference to the motion capture sensors attached. We found that the zombie was often described as: disgusting, rotting or decomposing. The toon character was often described as cartoon-like but few people described him as cute. The mannequin was mainly described as a wooden person.

We first analyzed the data separately for each of the 6 displayed basic emotions. A two factor ANOVA was conducted on the data where conditions were *body type* (6) and *rated emotion* (41). As expected, for all displayed emotions, there was a main effect of rated emotion where the list of emotions were rated differently to each other ($p < 0.0001$ in all cases). For displayed emotion **Anger**, there was no main effect of body type, which implies that the average ratings for each of the bodies were the same. Also, there was no interaction between body type and rated emotion, which implies that ratings for each of the 41 emotions were not statistically different across body types. Similar effects were found for displayed emotions **Surprise**, **Disgust** and **Fear**. No main effect of body type was found for displayed emotions **Happiness** or **Sadness**. However, interactions between body type and rated emotion did occur for both **Happiness** ($F_{200,2240} = 1.235, p < 0.02$) and **Sadness** ($F_{200,2240} = 1.29, p < 0.004$). Figure 2 illustrates how closely the average intensity ratings for all 5 virtual bodies were to those of the real. Post-hoc analysis using Newman-Keuls pairwise comparisons showed that some of the bodies were rated differently for **Happiness** (Table 1). However, these differences only account for less that 1% of all the one-to-one comparisons possible. Therefore, our main finding is that participants were very robust in their ratings, regardless of the appearance of the character.

Rated Emotion	Rated Higher	Rated Lower
Gloating	Toon	Real
	Toon	Lo
Contentment	Lo	Real
	Hi	Zombie
Ecstasy	Real	Zombie
Pride	Toon	Low
Affection	Zombie	Real

Table 1: *Statistically Significant ANOVA results for Happiness displayed motion ($p < 0.01$ in all cases).*

In order to test whether or not there was a difference between the real and virtual representations, we collapsed all of the ratings for the five virtual characters and compared them to the real data using two-factor ANOVAs for each of the displayed emotions, as before. We found no main effect of body type for any of the displayed basic emotions, nor any interaction between body type and rated emotion for **Anger, Happiness, Sadness** or **Fear**. However, we did find an interaction between rated emotion and body for **Surprise** ($F_{40,760} = 2.22, p < 0.00003$). Post-hoc analysis showed that there were differences in two out of the forty-one rated emotions. Firstly, the average intensity rating for *surprise* was higher for the real character than for the virtual characters ($p < 0.0009$), where the average intensities were 8.7 and 6.8 out of 10 respectively. Also, for rated emotion *interest* ($p < 0.0025$), where the average intensities were 4.9 and 7 respectively. This implied that the real actor conveyed these emotions more intensely than the virtual. A common side-effect of using motion capture data is that limb contacts are difficult to maintain, due to the fact that the original actor may have been a different size to the virtual. For the **Surprise** emotion, the real actor touched his face with both hands to indicate being shocked. In the virtual versions, the hands did not quite reach the face at times, and this may have accounted for the difference in ratings here.

An interaction between body type and rated emotion was also found for **Disgust**, where differences occurred for two of the forty-one

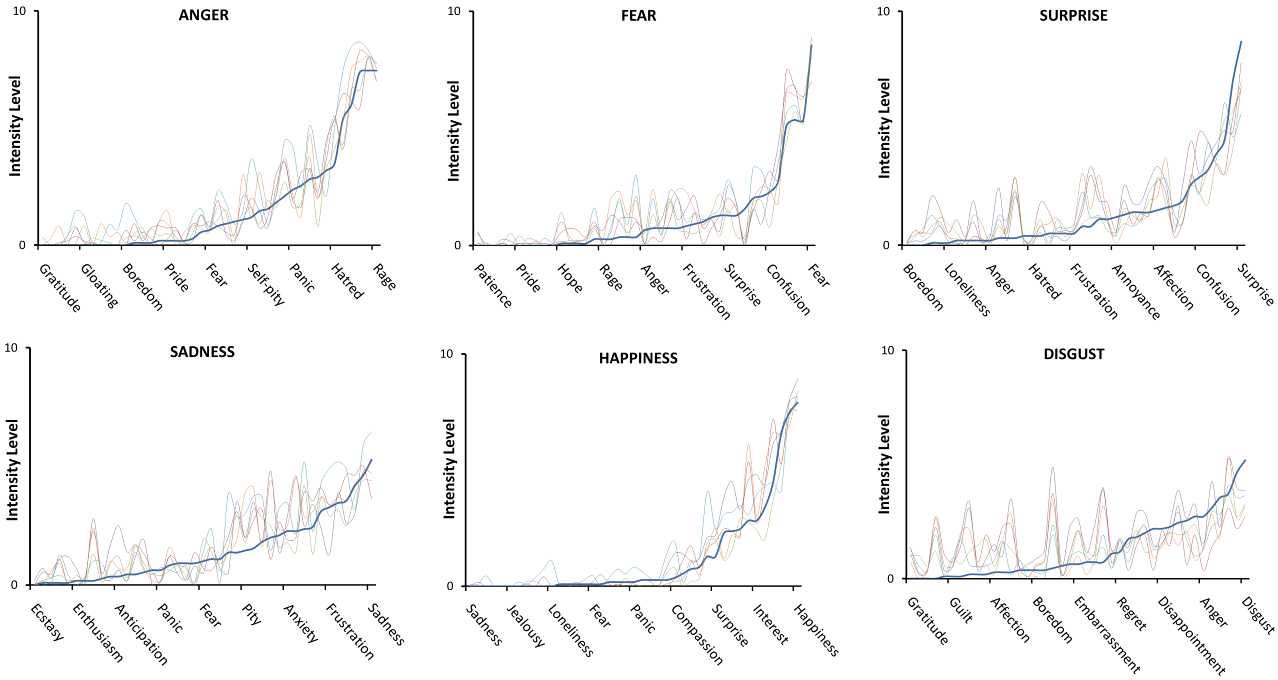

Figure 2: *Average ratings for all bodies for* ***Anger, Fear, Surprise, Sadness, Happiness*** *and* ***Disgust****. The 41 emotions are ordered according to the ratings for the real character, indicated with a dark line. Labels for 9 of the 41 are displayed for ease of viewing.*

rated emotions. Participants rated the real as having a higher intensity of *disgust* than the virtual ($p < 0.04$), with average ratings of 5.1 and 3.2. There was also a difference for *happiness* where the virtual was rated as having a higher intensity than the real ($p < 0.01$), with average values of 0.45 and 3.2. For this motion, there was some hand-hand contact but these were maintained for each of the virtual characters, so this small difference in ratings for real and virtual may have been due to some other factor such as different neural responses for real and virtual characters (as in [Perani et al. 2001]). It is also interesting to note that the shape of the **Disgust** curve in Figure 2 shows the most divergence. Disgust is widely regarded as the most difficult emotion to accurately identify [Atkinson et al. 2004].

7 Discussion and Future Work

Our results indicate that when realistic human body motion is used, it is the motion and not the body representation that dominates our perception of portrayed emotion.

We found no difference between the emotion ratings for high and low resolution virtual models, which would imply that the movements were categorized in the same way regardless of model resolution. There was no difference between the emotion ratings for the zombie and the real human or the other characters. This would imply that, even though descriptions of the static body suggested that participants found him unappealing, an uncanny eeriness reaction did not occur when realistic motion was applied. This indicates that perhaps it could be the face and hand motions that cause the reactions of eeriness and not the body motions, as tested here. We would be interested to investigate this theory further in future work. Also, the motions tested in the study were all high quality motion-captured. It would be interesting to see if lower quality or de-synchronized motion would result in differences between the anthropomorphized bodies.

When we compared the real to all of the virtual characters we did find some differences in intensity ratings for two of the basic emotions. It would be interesting to test if these differences were due to the motions themselves or due to the differences in neural activation for the real and virtual characters. Future eye-tracking and fMRI studies would allow us to further examine the factors involved.

One limitation of these results is that just one actor was used to create the stimuli. It would be interesting to test if the results were replicated for other actors and non-actors. Also, since our results indicate that it is the motion that drives emotion recognition rather than shape, conducting the experiment using a point-light stimulus could allow us to gain further insight. We would also like to explore other experimental paradigms.

References

ATKINSON, A. P., DITTRICH, W. H., GEMMELL, A. J., AND YOUNG, A. W. 2004. Emotion perception from dynamic and static body expressions in point-light and full-light displays. *Perception 33*, 6, 717–746.

CHAMINADE, T., HODGINS, J., AND KAWATO, M. 2007. Anthropomorphism influences perception of computer-animated characters' actions. *Social Cognitive and Affective Neuroscience 2*, 206–216.

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

CRANE, E., AND GROSS, M. 2007. Motion capture and emotion: Affect detection in whole body movement. *Affective Computing and Intelligent Interaction 28*, 95–101.

EKMAN, P. 1992. Are there basic emotions? *Psychological Review 99*, 3, 550–553.

HODGINS, J., O'BRIEN, J., AND TUMBLIN, J. 1998. Perception of human motion with different geometric models. *IEEE Transactions on Visualization and Computer Graphics 4*, 4, 307–316.

JOHANSSON, G. 1973. Visual perception of biological motion and a model for its analysis. *Perception and Psychophysics 14*, 2, 201–211.

MAR, R. A., KELLEY, W. M., HEATHERTON, T. F., AND MACRAE, C. N. 2007. Detecting agency from the biological motion of veridical vs animated agents. *Social Cognitive and Affective Neuroscience Adcance Access 2*, 3, 199–205.

MCDONNELL, R., DOBBYN, S., AND O'SULLIVAN, C. 2005. LOD human representations: A comparative study. In *Proceedings of VCrowds*, 101–115.

MCDONNELL, R., JÖRG, S., HODGINS, J. K., NEWELL, F., AND O'SULLIVAN, C. 2007. Virtual shapers & movers: form and motion affect sex perception. In *Applied Perception in Graphics and Visualization*, 7–10.

MORI, M. 1970. The uncanny valley. *Energy 7*, 4, 33–35.

NOWAK, K. L., AND BIOCCA, F. 2003. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence 12*, 5, 481–494.

PASCH, M., AND POPPE, R. 2007. Person or puppet? the role of stimulus realism in attributing emotion to static body postures. In *Proceedings of the Conference on Affective Computing and Intelligent Interaction (ACII)*, vol. 4738, 83–94.

PERANI, D., FAZIO, F., BORGHESE, N. A., TETTAMANTI, M., FERRARI, S., DECETY, J., AND GILARDI, M. C. 2001. Different brain correlates for watching real and virtual hand actions. *NeuroImage 14*, 749–758.

REEVES, B., AND NAAS, C. 1996. The media equation: How people treat computers, television, and new media like real people and places. In *Stanford, CA*, CSLI Publications.

S. HAN, Y. JIANG, G. W. H. T. Z., AND CAI, P. 2005. Distinct neural substrates for the perception of real and virtual visual worlds. *NeuroImage 24*, 928–935.

SLATER, M., AND STEED, A. 2002. Meeting people virtually: Experiments in shared virtual environments. In *The social life of avatars: Presence and interaction in shared virtual environments*, 146–171.

WALLBOTT, H. G. 1998. Bodily expression of emotion. *European Journal of Social Psychology 28*, 6, 879–896.