The Effect of Posture and Dynamics on the Perception of Emotion

Aline Normoyle, Fannie Liu, Mubbasir Kapadia, Norman I. Badler*
University of Pennsylvania

Sophie Jörg†
Clemson University

Figure 1: *Characteristic frame for each emotion from the clips with the best recognition rates.*

Abstract

Motion capture remains a popular and widely-used method for animating virtual characters. However, all practical applications of motion capture rely on motion editing techniques to increase the reusability and flexibility of captured motions. Because humans are proficient in detecting and interpreting subtle details in human motion, understanding the perceptual consequences of motion editing is essential. Thus in this work, we perform three experiments to gain a better understanding of how motion editing might affect the emotional content of a captured performance, particularly changes in posture and dynamics, two factors shown to be important perceptual indicators of bodily emotions. In these studies, we analyse the properties (angles and velocities) and perception (recognition rates and perceived intensities) of a varied set of full-body motion clips representing the six emotions anger, disgust, fear, happiness, sadness, and surprise. We have found that emotions are mostly conveyed through the upper body, that the perceived intensity of an emotion can be reduced by blending with a neutral motion, and that posture changes can alter the perceived emotion but subtle changes in dynamics only alter the intensity.

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

Keywords: emotions, perception, motion editing, virtual characters

1 Introduction

To engage us in a movie or game, virtual characters need the ability to convey emotions in a convincing way. This goal is most often achieved by capturing human performances directly with motion capture, as this technology faithfully records the nuances of an actor's performance. However, one limitation of motion capture is that it does not inherently adapt to new situations. Thus, extensive research has explored how to increase the reusability and flexibility of motion capture clips, leading to numerous techniques in common use today, such as inverse kinematics, interpolation, blending, retargeting, morphing, move trees, motion graphs, overlays, and splicing. Further extensive research, mostly in psychology, has explored how we perceive emotions. However, little research has looked at the perceptual effects of motion editing techniques.

In this paper, we investigate which aspects of body language are important for conveying emotions with two goals in mind. The first goal is to understand how changes introduced by motion editing might alter the emotional content of a motion, so we can ensure that important aspects of a performance are preserved. The second goal is to gain insight on *how* we may edit captured motion to change its emotional content, further increasing its reusability. We study six basic emotions [Ekman 1992] shown to be readily recognized across cultures: anger, disgust, fear, happiness, sadness, and surprise (Figure 1). The motion of a character's body can effectively express all basic emotions [Atkinson et al. 2004], including its context and intensity, and is the focus of our study.

Many motion editing techniques affect part of the body, change poses, or modify the motion dynamics. Thus, we performed three experiments to gain a better understanding of the perception of emotions based on posture and dynamics. We recorded an actor giving ten short performances of each basic emotion, which were mapped to a generic humanoid virtual character. In our first experiment, we analyse this large set of clips to search for differences in posture and dynamics between emotions and to compare the perception of our stimuli to previous work. We also establish a baseline set of twelve animation clips (two for each emotion) having the highest recognition rates among viewers. As some editing techniques affect only specific joints of the body, our second experiment determines what part of the body (either head, upper body and head, or lower body) conveys the emotion most strongly. As previous work has established a relationship between pose, dynamics, and the perception of emotion [Wallbott 1998; Pollick et al. 2001; Atkinson et al. 2004; Coulson 2004; Roether et al. 2009], our third experiment systematically alters the poses and joint velocities. Our experiments yield the following observations:

- We confirm several findings on the perception of emotions based on a large and varied set of clips. For example, we con-

*e-mail:alinen,fannliu,mubbasir,badler@seas.upenn.edu
†e-mail:sjoerg@clemson.edu

firm that happy and angry movements have higher velocities and greater joint amplitudes whereas sadness has slower joint velocities and smaller amplitudes.

- Emotions are mostly conveyed through the upper body.
- The perceived intensity of an emotion can be reduced by blending with a neutral motion.
- We find that posture changes can alter the perceived emotion and its intensity while changes in dynamics only alter the intensity.

2 Related Work

There exists extensive research in psychology [Roether et al. 2009] that aims to understand the perceptual significance of body motion on conveying emotions. Coulson [2004] showed that happiness and sadness are clearly recognizable while disgust is harder to discern. He also found that surprise and fear were harder to discern from purely static poses. Wallbott [1998] observed a relationship between emotion type and posture characteristics, showing that differences in emotion can be partly explained by the dimension of activation, where activation refers to velocities, accelerations, and jerk. Atkinson *et al.* [2004] showed that emotion could be recognized from body motion, with exaggerations in motion increasing the recognition accuracy. Roether *et al.* [2009] observed that elbow and hip flexion were important attributes for anger and fear, while head inclination was important for recognizing sadness in motions. Sawada *et al.* [2003] found that dancers varied the speed, force, and directness of their arm movements when conveying joy, sadness, and anger. Recent work [Ennis and Egges 2012] observed that negative emotions are better recognizable.

The recognition of emotions is mostly consistent across cultures and geometries. Kleinsmith *et al.* [2006] conducted a study to evaluate the cultural differences in the perception of emotion from static body postures, observing moderate similarity across cultures. Pasch and Poppe [2007] evaluated the importance of the realism of the stimuli on the perception of emotion, demonstrating that high realism did not always conform to an increase in agreement of the emotional content. McDonnell *et al.* [2008] investigated the role of body shape on the perception of emotion and found that emotion identification is largely robust to change in body shape.

Several previous studies focus on specific motion categories such as gait [Crane and Gross 2007; Roether et al. 2009], knocking and drinking [Pollick et al. 2001], and dance movements [Sawada et al. 2003], whereas this study uses non-restricted actor portrayals, similarly to previous work [Wallbott 1998; Atkinson et al. 2004; McDonnell et al. 2008]. This choice was made for two reasons. First, we did not want to inadvertently restrict the motions to parts of the body. For example, if we based our experiments on knocking, we might bias the upper body to display most of the emotion. Second, because several related studies have already performed systematic comparisons between motions of the same type, we wanted to investigate how their findings compared to our varied motion set.

Impact of Motion Editing. Avoiding unwanted artifacts during motion editing is an important issue in computer animation. Ren *et al.* [2005] presented a data-driven approach to quantifying naturalness in human motion, which they evaluated against edited and keyframed animations. Ryall *et al.* [2012] and Reitsma and Pollard [2003] observed viewers' sensitivity to timing changes made to walking and ballistic motions and found that participants were more sensitive to time warping when slow motions were made faster than when fast motions were made slower. Safonova and Hodgins [2005] analysed the effect of interpolation on physical correctness and found that such operations can create unrealistic trajectories for the center of mass or create unrealistic contacts.

Emotion Simulation. Algorithms for synthesizing emotional movements is important for both the domains of embodied conversational agents and animation. Existing approaches have looked at how to better parameterize animations for different emotional styles [Unuma et al. 1995; Chi et al. 2000; Hartmann et al. 2006] as well as how to model style (including emotion) and transfer it automatically [Brand and Hertzmann 2000; Hsu et al. 2005; Shapiro et al. 2006; Torresani et al. 2007]. Another approach is to add emotional styles directly to the joint curves. For example, Amaya *et al.* [1996] described a signal processing technique for transforming neutral animations into emotional animations. These techniques investigate more intuitive ways for users to change pose, velocity and acceleration automatically to achieve a desired effect. In this work, we restrict our analysis to the simple operations of occlusions, offsets, blending, and time warping which allows us to systematically test changes to either pose or dynamics.

3 Experiment 1: Emotion Recognition and Movement Analysis

In our first experiment, we determine the recognition rates achievable with our clips. We use the results to select well recognized motions for the subsequent experiments, to compare our stimuli to previous studies, and to validate existing findings against our diverse motion set.

3.1 Stimuli creation

We invited an experienced stage actor to give ten short performances of each of the six emotions: anger, disgust, fear, happiness, sadness, and surprise (1 *actor* $\times$ 6 *emotions* $\times$ 10 *portrayals* = 60 animation clips). The actor was asked to convey each emotion as convincingly as possible using his entire body and no vocal cues. He was also told that his face and hands would be blurred. Given these instructions, the actor improvised ten entirely different portrayals of each emotion, some of which were repeated until the actor was satisfied with his performance. For example when performing anger, the actor's performances ranged from tantrums, to yelling and pointing, and finally to head shaking and staring (to show contained rage, or annoyance). In general, our actor portrayed each emotion using stereotypical movements. To show disgust, he most often turned away, made sweeping away gestures with his hands, or acted sick to his stomach. To show fear, he would back away, place his hands up in front of the body, or freeze in place. To show sadness and grief, he would cross his arms, hang his head, or cover his face with his hands. To show happiness, he cheered, jumped for joy, bowed, and twirled. To show surprise, he performed startled jumps, followed by relief. Although such performances may be considered staged exaggerations of natural emotions, they are typical of the types of movements commonly captured for entertainment. Previous research has shown that exaggerated expressions of emotion are more readily recognizable by viewers [Atkinson et al. 2004]. As a result, the recognition rates for our stimuli might be higher than for more realistic displays of emotion.

Our actor's body motions were recorded with a 12-camera optical Vicon system and post-processed in Vicon Nexus and Autodesk MotionBuilder. A standard skeleton with 22 joints was used. We then created a skinned character adapted to the actor's skeleton.

As we did not record facial or finger motions, we box blur the face and hands of the virtual character similar to McDonnell *et al.* [2008] to ensure that the motionless, unnatural-looking faces or fingers would not distract the viewer. We kept the blurred area as small as possible, hiding the facial features and the finger motions of the character but still showing the general orientation of the head and the hands. The box blur algorithm was implemented as a Maya plugin.

The resulting character was rendered in a neutral environment. A few clips had to be excluded from the experiment because they included too many occluded markers to be cleaned accurately, or resulted in clips that were too short. In total, we obtained 55 animated clips (11 anger, 9 disgust, 7 fear, 10 happiness, 9 sadness, and 9 surprise), each between 2 and 10 seconds long at 24 fps.

3.2 Method

Fifteen participants (8M, 7F) between 17 and 53 (mean 25.6) watched all 55 clips. All participants were naïve to the purpose of the experiment and had normal or corrected to normal vision. After each clip, they were asked to specify which emotion they thought was conveyed in the video with a forced-choice between anger, disgust, fear, happiness, sadness, and surprise. Although forced-choice questionnaires have the drawback of restricting the responses from our participants and could potentially inflate recognition rates [Frank and Stennett 2001], we chose this format to be able to compare our study to related work, which mostly uses forced-choice.

As the goal of this experiment was to find how well our stimuli conveyed the basic emotions, participants could perform the study at their own pace and view each clip as often as they wanted. They were sent a link to the study and were allowed to view the stimuli on their own computers and to take breaks. The clips were presented in a different random order for each participant.

After all 55 clips had been viewed and an emotion selected for each clip, participants were asked to watch them a second time and to rate the intensity and energy of the emotion on a scale from 1 (not intense/low exertion) to 5 (very intense/high exertion). Definitions of intensity – "How deeply the person feels the emotion" – and energy – "The level of exertion and vigor of the person's movement" – were displayed on screen.

The entire experiment took between 30 and 45 minutes to complete. Participants had the option to perform the study in our lab, where they were compensated with food and refreshments, or they could choose to perform the study elsewhere without compensation.

3.3 Results

On average across all clips, 62.4% of the clips were recognized correctly. Figure 2 summarizes our results. In general, we see that happiness, anger, fear, and surprise were recognized best whereas disgust and sadness were recognized least. Anger was generally very well recognized with the exception of two motions which had only one correct response each. The two poorly recognized motions were subtle anger expressions ("contained rage") which did not translate well to the character without facial animation. Many participants mis-categorized these anger motions as sadness. In general, clips with low recognition rates fell into two categories: participants had no agreement on the displayed emotion, suggesting that they were merely guessing, or there was high agreement among participants for the wrong emotion. Most clips with low scores were in the first category. Examples from the second category included a sadness clip where our actor was waving his hands (to gesture "go away"), which was chosen as anger or happiness, and a happiness clip which was predominately chosen as anger.

Disgust was recognized least well in our experiment (although two disgust motions had high recognition rates). Disgust is known to be

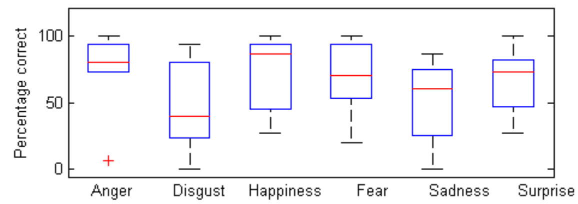


Figure 2: *Emotion recognition from our first experiment. Each column summarizes the percentage of correct responses for each emotion. The red middle band indicates the median value, the box top and bottom shows the first and third percentiles (75% and 25%), and the top and bottom whiskers show the maximum and minimum recognition percentage. Crosshairs show outliers.*

Emotion	Happiness	Sadness	Anger	Disgust	Fear	Surprise
Happiness	**68.0**	3.3	24.7	0.7	0.0	3.3
Sadness	13.3	**50.4**	10.4	8.1	12.6	5.2
Anger	5.6	12.7	**72.1**	6.7	0.0	3.0
Disgust	9.6	20.7	7.4	**46.7**	11.1	4.4
Fear	0.0	6.7	1.0	9.5	**71.4**	11.4
Surprise	5.2	9.6	8.1	3.0	8.9	**65.2**

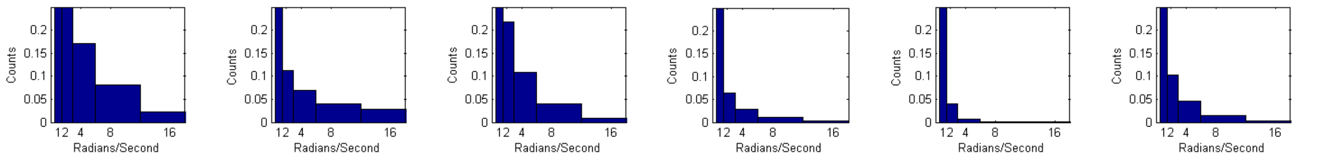
Table 1: *Confusion matrix from our first experiment. Entries show the percentage of times participants chose each emotion in a forced choice experiment. The displayed emotions are listed on the left, the selection of the viewer at the top.*

a less readily recognized emotion and our result is consistent with a large body of work [Ekman 1992; Atkinson et al. 2004]. However, sadness is typically recognized at a higher rate. Our hypothesis regarding this finding is that the actor often tried to show grief and distress, where the body typically moves more than in a depressed individual. Without facial capture, this subtlety was lost. In the confusion matrix in Table 1, we see that disgust was most often confused with sadness. With 12.5% of the total selections, disgust was also selected less often than any other emotion (all emotions were displayed equally often: 16.7% of the time).

From this experiment, we choose the two animated performances with the highest recognition rates to use for all subsequent experiments: two anger motions each with 100% correct recognition; two disgust motions (recognition rates 80% and 93%); two fear motions (93% and 100%); two happy motions (93% and 100%); two sad motions (80% and 86%); and two surprise motions (100% and 87%). In case of a tie, the clips that looked least similar to each other were selected.

We also analyse the pose and velocities of our motion clips and compare our findings to previous studies. Figure 3 shows histograms of the rotational speeds for the major animated joints of our character, namely the root (pelvis), left and right hips, knees, ankles, spine, shoulders, elbows, wrists, and neck. To compute angular velocities, we first compute quaternion rates using 5-point central differencing and then convert the quaternion rate to an angular velocity vector according to [Diebel 2006]. Our motions are consistent with previous research which states that anger and happiness tend to have larger and faster joint movements, whereas fear and sadness tend to have smaller and slower joint movements [Roether et al. 2009].

Figure 4 compares the amplitudes for the head, shoulders, and elbows (amplitude is defined as the difference between the max joint angle and min joint angle for each motion category). Our findings are consistent with previous research which states that happiness



(a) Anger (b) Disgust (c) Happiness (d) Fear (e) Sadness (f) Surprise

Figure 3: *Histograms of rotational speeds (radians/second) for the major joints of our character: root, left and right hips, knees, ankles, spine, shoulders, elbows, wrists, and neck. The x axis shows bins corresponding to speeds of 1, 2, 4, 8, and 16 radians/second respectively. The y axis shows counts for each speed across all motions of each category, normalized to range from 0 to 1 based on the maximum bin size. Our results are consistent with previous published research which states that anger and happiness tend to have larger and faster joint movements, whereas fear and sadness tend to have the least joint movement. For our motions, surprise and disgust lie somewhere in between these two extremes.*

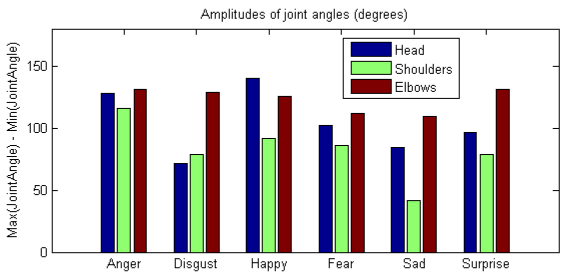
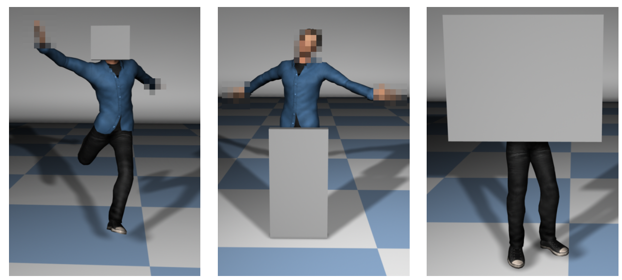


Figure 4: *Joint amplitudes. Our captured motions are consistent with published research which states that happiness and anger have higher amplitudes whereas sadness and fear have lower amplitudes.*

and anger have higher amplitudes whereas sadness and fear have lower amplitudes. Existing research is less clear regarding disgust and surprise. For our dataset, surprise shared amplitude characteristics with fear, but had greater elbow movement. Disgust had high elbow movement but low head and shoulder movement.

Lastly, we looked at modal and average flexion angles, defined as the angle between limbs. Specifically, previous research describes reduced head angle for sad walking and increased elbow angle for fearful and angry walking [Roether et al. 2009]. However, our motion set did not produce convincingly consistent results. Most joint angle distributions were not normally distributed. Based on histograms of joint angle, both sad and disgust motions had modal head angles of 160 (where 180 corresponds to looking straight forward and 90 degrees corresponds to looking straight down) whereas all others had modal head angles of 170 degrees. Elbow angle was greatest for disgust and fear (110 degrees, where 180 corresponds to a fully flexed arm), second largest for sadness and anger (150 degrees), and smallest for surprise and happy (170 degrees). To understand these results, we note that many of our sad clips had the hands at the face, and several of our disgust motions huddled the arms into the body. Many of the anger motions contained punching and swinging gestures.

We averaged the ratings of each participant over all clips of the same emotion and used a repeated measures ANOVA to determine that there were significant differences in intensity (F(5,70)=43.1, $p < 0.001$) and energy (F(5,70)=91.4, $p < 0.001$) for the different emotions. Post-hoc Tukey tests were used to determine that the intensities and energies of happy and angry were significantly higher than the other emotions; that all emotions except fear were significantly higher in intensity than sadness; and that all emotions were significantly higher in energy than sadness. The correlation coefficient between intensity and energy was 0.65.



(a) (b) (c)

Figure 5: *Stimuli examples from our second experiment in which we hid either the head, the lower body, or the upper body.*

4 Experiment 2: Partial Occlusions

Many motion editing operations can be applied to parts of a virtual character's body. For example, inverse kinematics and overlays for reaching usually just affect the upper body whereas ground clamping techniques to adapt walking motions usually just affect the lower body. Therefore, in our second experiment, we determine which parts of the body are important in conveying emotions.

4.1 Stimuli

For each of the six emotions, we chose the two clips with the highest recognition rates from Experiment 1 (see Section 3.3). We then occluded different parts of the body: the head motion (or NH for "No Head motion"), the lower body motion (NL), and the upper body motion (NU). The unaltered motion is labeled OR for "original". We did not alter the root motion for any of the conditions to avoid very unnatural motions that could affect the ratings in unintended ways (see video for examples). To occlude the body parts, we erase all motion from the considered part and cover it with a nondescript flat cuboid that we attached to the character (see Figure 5). We obtain 6 *Emotions* x 2 *Clips* x 4 *Occlusion types* = 48 different clips.

4.2 Method

Sixteen participants who were not involved in the previous experiment watched all of the clips in small groups of 1–3 participants on a large projection screen in a seminar room. As the aim of this experiment is not to determine the highest possible recognition rate of each clip but to investigate differences between several partial occlusions, we chose a faster pace for this experiment. Participants viewed a clip once. Then they had a total of six seconds to specify the perceived emotion in a forced-choice between the six basic emotions and then the perceived intensity of that emotion on a scale from 1 to 5 similar to Experiment 1. After four seconds, a sound was played together with the number of the next clip to alert participants to look at the screen again. Then the next clip started. Participants were asked to watch each clip in its full length.

Although our pilots showed that six seconds was very short, participants were able to follow the instructions after a short training phase. However, we decided not to ask to rate the energy of the clips as we found that participants were not able to effectively distinguish between intensity and energy in such a short time. Our second reason for the very fast pace was that participants watched the same motions with different occlusions. Once a non-occluded animation has been viewed, it is possible for a participant to recognize that animation in subsequent clips and to infer the perceived emotion and intensities. The fast pace of our experiment did not give participants time to think about the motions. Based on questions and conversations in the debriefing, we assume that many participants started to recognize some of the motions towards the end of the experiment.

Before starting each experiment, we showed participants four training clips at the same pace as the experiment. The training clips were chosen from the unused clips in the first experiment. A short break to answer any questions ensured that participants understood the instructions. The participants viewed all 48 clips in random order. After a short break, they viewed all 48 clips again in a different random order. The full experiment took about 25 minutes to complete and participants were rewarded with $5.

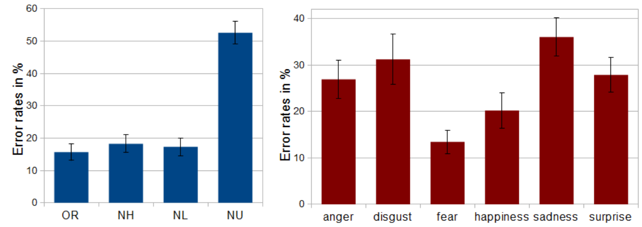


Figure 6: *Error rates for each condition (left) and for each emotion (right). Emotions were recognized significantly less often when the upper body was hidden (NU: no upper body) than when it was visible (OR: original, NH: no head, NL: no lower body). Error bars represent one standard error of the mean in all graphs.*

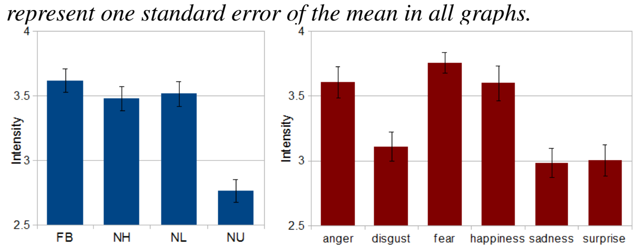


Figure 7: *Intensities for each condition (left) and for each emotion (right). Emotions were rated to have a lower intensity on a scale from 1 to 5 when the upper body was occluded. Also, anger, fear, and happiness were rated to have a significantly higher intensity than disgust, sadness, and surprise.*

4.3 Results

4.3.1 Emotion recognition

Three participants either did not follow the instructions or checked the boxes in an illegible manner. Their answers had to be discarded, leaving 13 participants in our analysis. For the emotion recognition, we computed the error rates for each participant, emotion, and occlusion type by averaging over the two clips and two repetitions. We then performed a repeated measures ANOVA with the within-subject factors *Occlusion type* and *Emotion*. We used Newman-Keuls post-hoc tests to determine the origin of the significant effects.

We found a main effect of Occlusion type with $F(3, 36) = 62.8, p < 0.001$, due to the fact that the condition where the upper body was hidden had significantly higher error rates (lower recognition rates) than the other three occlusion types. This effect was surprisingly distinct as can be seen in Figure 6 (left). There were no significant differences between the other three occlusion conditions.

As expected, we also found a main effect of Emotion ($F(5, 60) = 5.3, p < 0.001$), meaning that the different emotions were not recognized equally well. Fear was recognized best on average and significantly better than all other emotions except happiness (see Figure 6, right). Sadness had the lowest recognition rate (or highest error rate), which differed significantly from fear and happiness and reflects the already lower recognition rate of the original clips.

Furthermore, there is an interaction effect between Occlusion type and Emotion ($F(15, 180) = 7.5, p < 0.001$), which can be traced back to three causes: First, fear was the only emotion where the error rates remained the same in all occlusion conditions. For all other emotions, hiding the upper body resulted in significantly higher recognition errors than all or most (for anger) other conditions. Second, sadness with an occluded head (NH) was recognized least of all OR, NH, and NL motions and significantly less well than eight other motions. Third, anger with occluded lower body (NL) was recognized second least of all OR, NH, and NL motions, leading to significant differences with three of them.

4.3.2 Intensities

To analyse intensity, we had to discard the results of one more participant as those had not been reported correctly, reducing the participant count to 12. Similarly to emotion recognition, we computed the averages for each participant, emotion, and occlusion type over the two clips and two repetitions, performed a repeated measures ANOVA with the within-subject factors *Occlusion type* and *Emotion*, and used Newman-Keuls post-hoc tests to determine the origin of the significant effects.

Here again, we found a main effect of Occlusion type ($F(3, 33) = 29.5, p < 0.001$) based on a significant difference between the clips where the upper body was occluded (NU) and the clips in the three other occlusion conditions. The intensity was rated significantly lower for NU (see Figure 7, left). There was also a main effect of Emotion with $F(5, 55) = 8.3, p < 0.001$. The post-hoc test showed that the intensity ratings were split into two groups: Anger, Fear, and Happiness were rated to have higher intensities than Disgust, Sadness, and Surprise (see Figure 7, right). There were no significant differences within each group but each combination of ratings across groups was significantly different from each other. Finally, there is an interaction effect between Occlusion type and Emotion ($F(15, 165) = 3.5, p < 0.001$), which, however, does not add new insights. The posthoc test reveals that for each individual emotion, the intensities are rated significantly lower when the upper body is occluded than in the three other occlusion types, whereas there are no significant differences between those three.

4.3.3 Discussion

We infer that the upper body is crucial for the perception of emotions. The lower body or the head alone were not relevant in our set of clips to recognize the emotion. The irrelevance of the head for all emotions except sadness could have been due to our blurring of

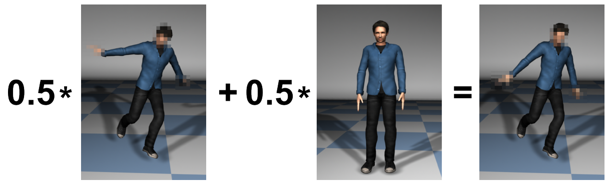


Figure 8: *Body blend (BB) condition. The upper body joints are blended with a neutral motion having the arms at the side. The above example shows the result of blending 50% of the original motion with 50% of the neutral motion. Joint rotations are represented using quaternions and blended with slerp. This condition changes both pose and velocity.*

the head in the baseline clips: when we occluded the head entirely, there was no considerable impact. Alternatively, this finding might have be due to the head being unimportant for recognizing the emotion in nearly all our clips. However, the relatively high error rate for sadness when the head was occluded complies with previous work that head motion is particularly important for displaying sadness.

However for the lower body, which was not blurred, the emotion could be effectively conveyed through lower body motions, for example through the kicking motion for anger, the running away motion for fear, or the jump for surprise. Interestingly, differences between the occluded recognition rates were smallest for two of the emotions that displayed very distinct lower body motions, namely fear and anger. Because we decided to hide the lower body motion but leave the root (pelvis) motion unaltered, viewers may have inferred the lower body motions based on the movements of the upper body. We also considered several other options, such as deleting the root motion or replacing the lower body motion with a neutral lower body motion. However, these options drastically changed the full body motions instead of just hiding parts of the body and were therefore discarded.

The intensity ratings largely mirror the recognition rates: when the error rates were higher, the intensity was judged lower. This is not surprising: participants might recognize an emotion less clearly when its intensity is low or might attribute a low intensity to a motion when they are unsure which emotion it represents.

5 Experiment 3: Posture and Dynamics

Previous research suggests that velocity, accelerations, and jerk (defined as the time derivative of acceleration) are important factors in emotional body language along with pose [Roether et al. 2009]. Because motion editing procedures such as interpolation and blending change both the pose and dynamics, we investigate these effects. We hypothesize that small scale changes might affect the emotion intensity, while large scale changes might affect whether the emotion is recognized correctly.

5.1 Stimuli

We filtered the major joint curves of our best-recognized motions to produce changes to either the poses, the velocities, or both. For this experiment, we created four conditions: two conditions (BB25, BB50) in which we change poses and velocities by blending the upper body with a neutral posture, one condition (DTW) in which we change the timing but not the poses through dynamic time warping, and one condition (OFF) where we change the poses but not the timing by setting constant offsets to either the shoulders, elbows, or head.

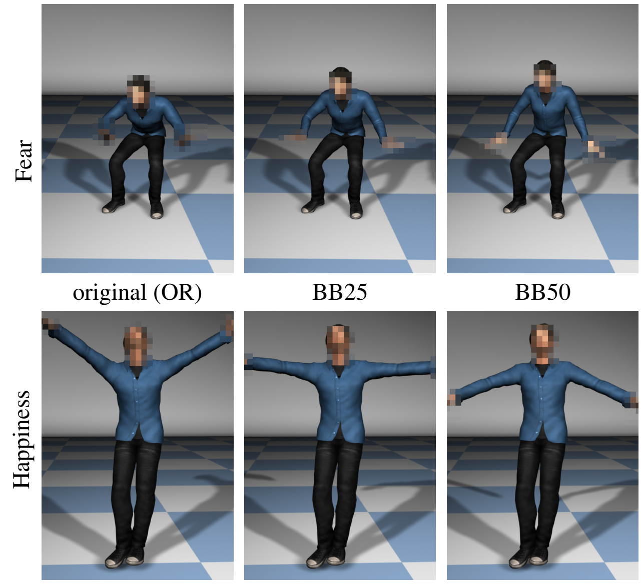


Figure 9: *Stimuli examples from body blend (BB) conditions. Original motions appear in the first column. The second column shows BB25, which retains 75% of the original motion. The third row shows BB50, which retains 50% of the original motion. As the poses moved towards neutral, the perceived intensity of the emotion is decreased.*

Our two body blend conditions (BB25, BB50) blend the joints of the upper body, from the spine and upwards (Figure 8 and 9). The upper body was chosen because the previous experiment showed it to be the most relevant for the perception of emotions. To compute each blend condition, each frame (where a frame consists of the rotations for each joint) of the original motion is blended with a neutral pose having the arms down at the side. BB25 blends 75% of the original motion with 25% of the neutral pose. BB50 blends 50% of the original motion with 50% of the neutral pose.

Our dynamic time warping condition (DTW) modifies the timing of the motion such that no joint velocity is higher than a given maximum. We choose our maximum value separately for each emotion as 200% of the average speed of the fastest moving joint. If the speed of all joints for a set of frames does not exceed the maximum value, those frames remain unchanged. Given a maximum speed, dynamic time warping is performed by computing new times for each frame and then resampling the motion curves at the original framerate. Specifically, if a frame originally occured at time t and had its fastest joint i moving at $v > v_{max}$, we adjust the time for this frame so that it occurs at $t + v/v_{max}\Delta t$, where Δt is 1/framerate. The curve is resampled by interpolating between the original poses.

Our offset condition (OFF) modifies the poses without changing the timing. Offsets were specified manually for either the shoulders, elbows, or spine and neck by specifying an offset pose q^{user} for a single reference frame $q(\hat{t})$ of each original motion (Figure 10). From the offset and reference pose, we compute an offset rotation q^{offset} which is then applied to all frames.

$$q^{offset} = (q(\hat{t}))^{-1} q^{user}$$
$$q_i^{new}(t) = q^{offset} q_i(t)$$

For our offset condition, we added offsets for each of our twelve clips (two per emotion). We created three motions with changed elbows, three motions with altered shoulders, three motions which

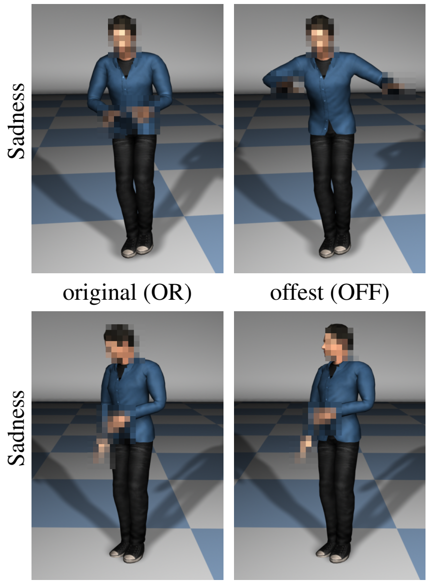


Figure 10: *Stimuli examples from the offset condition. Original poses appear in the first column. Modified poses appear in the second column.*

modified the neck and spine upwards, and three motions where the neck and spine went downwards.

We apply these four posture and velocity conditions — BB25, BB50, DTW, and OFF — to the two clips with the best recognition rates for each emotion (and keeping the original motion OR) to obtain 6 *Emotions* x 2 *Clips* x 5 *Alterations* = 60 different clips.

5.2 Method

We used the same fast paced method as experiment 2 (see Section 4.2). Seventeen naïve participants, who were not involved in any of the previous experiments, took part in experiment 3, which took less than 30 minutes to perform. As before, they were rewarded with $5.

5.3 Results and Discussion

One participant with unclear answers had to be excluded, leaving 16 participants in the analysis. As in Experiment 2, we computed the averages for each participant, emotion, and Alteration type (OR, BB25, BB50, DTW, and OFF) over the two clips and two repetitions, performed a repeated measures ANOVA with the within-subject factors *Alteration* and *Emotion*, and used Newman-Keuls post-hoc tests to determine the origin of the significant effects.

5.3.1 Emotion recognition

As expected, we found a main effect of Alteration with $F(4, 60) = 5.1, p < 0.01$ (see Figure 11, left). The error rates for the motions blended to 50% with the neutral motion (BB50) and the ones with offsets (OFF) were recognized significantly less well than the unmodified ones (OR). There were no significant differences between the recognition rates of BB25, the time-warped motion (DTW), and the original condition. However, the difference between the conditions OFF and DTW was significant. We also found a main effect of Emotion ($F(5, 75) = 6.1, p < 0.001$) due to the sadness motion being recognized at a significantly lower rate than all of the other emotions, which restates a result we found throughout the whole study.

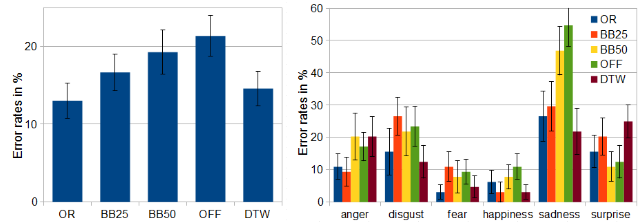


Figure 11: *Error rates for each alteration (left) and for all emotions and alterations (right). Emotions were recognized significantly less often when blended to 50% with a neutral motion or when an offset was added. Error bars represent one standard error of the mean in all graphs.*

Finally, there is an interaction effect between Alteration and Emotion with ($F(20, 300) = 3.8, p < 0.001$), mainly due to the offset and 50% neutral blended sadness clips (OFF and BB50) having significantly higher error rates than any other combination of Emotion and Alteration (see Figure 11, right). We found that those two combinations (sadness OFF, and sadness BB50) are also the origin for the main effects of Emotion and Alteration. The differences between the alterations of the other emotions were not significant.

5.3.2 Intensities

As expected, alterations changed the perceived intensities of our clips. The perceived intensities of the clips with the alterations BB50 and DTW were significantly reduced with a main effect of Alteration ($F(4, 60) = 5.1, p < 0.01$, see Figure 12, left). As before, the clips with different emotions were also rated as having different intensities (main effect of Emotion with $F(5.75) = 13.2, p < 0.001$), with anger and fear having the highest intensities, and sadness the lowest.

Finally, the interaction effect between Emotion and Alteration with $F(20, 300) = 2.5, p < 0.001$ showed that our modifications had a different effect depending on the emotion. For anger, time warping (DTW) significantly reduced the perceived intensity, whereas for sadness, blending (BB50) and adding an offset (OFF) reduced the intensity significantly (see Figure 12, right).

5.4 Discussion

We found that motion editing techniques can affect the recognition of emotions and its perceived intensity. The conditions BB50 and OFF, which both modify the posture, influenced the emotion recognition. The conditions BB50 and DTW, which both modify the timing, lead to a lower perceived intensity. From these results, we might infer that posture is a strong indicator of the type of emotion while timing and dynamics contribute to its perceived intensity. However based on the interaction effects, these effects are not

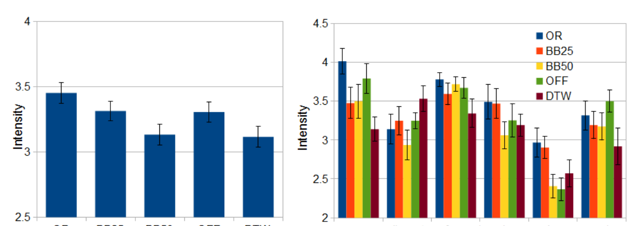


Figure 12: *Intensity for each alteration (left) and for all emotions and alterations (right). Emotions were rated to have a lower intensity on a scale from 1 to 5 when blended to 50% with a neutral motion or when time warped.*

equally strong for all emotions. The decreasing intensity for OR, BB25, and BB50 in Figure 12, left, suggests that the average intensity of a motion can be decreased by blending that motion with a neutral motion.

Participants were not able to determine the emotion sadness in the alterations OFF and BB50 as well as for the other emotions and alterations. This could be due to one of the OFF clips and both BB50 clips changing the orientation of the head, which was shown to be crucial in Experiment 2. Not surprisingly, the perceived intensity also decreased for those two cases where the emotion was not well recognized.

6 Conclusion

We investigated how changes to captured motion clips, such as those which commonly occur through motion editing, might alter the recognition and perceived intensity of an emotional performance. Rather than look at categories of motion, such as gait, we study a varied set of emotion clips. From these, we learn that the upper body motion is most crucial for the recognition of emotions, that changes to posture can change the perceived motion type whereas changes to dynamics can change the perceived intensity, and that the perceived intensity of an emotion can be reduced by blending with a neutral motion. However, these results do not apply equally well to all motions and emotions, and future work will try to understand these differences.

These findings might motivate one to take care when splicing and using IK to control the upper body, since such changes can affect emotion recognition and reduce the motion's perceived intensity. When blending major joints, such as the head, one might use smaller blend weights so that emotional content is not diluted. Future work will try to verify these hypotheses as well as determine whether such heuristics can be used to enhance automated algorithms for motion style.

Acknowledgements

This work was partially supported by NSF grant CNS-0855192. We wish to thank our participants, Parag Acharya for recruiting and running participants, Autodesk for donating licenses for Motion Builder and Maya, and the anonymous reviewers for their insightful comments.

References

AMAYA, K., BRUDERLIN, A., AND CALVERT, T. 1996. Emotion from motion. In *Graphics Interface*, GI '96, 222–229.

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.

BRAND, M., AND HERTZMANN, A. 2000. Style machines. In *SIGGRAPH*, 183–192.

CHI, D., COSTA, M., ZHAO, L., AND BADLER, N. 2000. The emote model for effort and shape. In *SIGGRAPH*, 173–182.

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. In *ACII*, 95–101.

DIEBEL, J. 2006. Representing attitude: Euler angles, unit quaternions, and rotation vectors. Tech. rep., Stanford University.

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

ENNIS, C., AND EGGES, A. 2012. Perception of complex emotional body language of a virtual character. In *MIG*, 112–121.

FRANK, M. G., AND STENNETT, J. 2001. The forced-choice paradigm and the perception of facial expressions of emotion. *Journal of Personality and Social Psychology 80(1)*, 75–85.

HARTMANN, B., MANCINI, M., AND PELACHAUD, C. 2006. Implementing expressive gesture synthesis for embodied conversational agents. In *Gesture in Human-Computer Interaction and Simulation*, GW'05, 188–199.

HSU, E., PULLI, K., AND POPOVIĆ, J. 2005. Style translation for human motion. In *SIGGRAPH*, 1082–1089.

KLEINSMITH, A., DE SILVA, P. R., AND BIANCHI-BERTHOUZE, N. 2006. Cross-cultural differences in recognizing affect from body posture. *Interact. Comput. 18*, 6 (Dec.), 1371–1389.

MCDONNELL, R., JÖRG, S., MCHUGH, J., NEWELL, F., AND O'SULLIVAN, C. 2008. Evaluating the emotional content of human motions on real and virtual characters. In *APGV*, 67–74.

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

POLLICK, F. E., PATERSON, H. M., BRUDERLIN, A., AND SANFORD, A. J. 2001. Perceiving affect from arm movement. *Cognition 82*, 2, B51 – B61.

REITSMA, P. S. A., AND POLLARD, N. S. 2003. Perceptual metrics for character animation: Sensitivity to errors in ballistic motion. In *ACM Transactions on Graphics*, vol. 22, 537–542.

REN, L., PATRICK, A., EFROS, A. A., HODGINS, J. K., AND REHG, J. M. 2005. A data-driven approach to quantifying natural human motion. *ACM Trans. Graph. 24*, 3 (July), 1090–1097.

ROETHER, C. L., OMLOR, L., CHRISTENSEN, A., AND GIESE, M. A. 2009. Critical features for the perception of emotion from gait. *Journal of Vision 9(6)*, 1–32.

RYALL, K., HOYET, L., AND HODGINES, JESSICA K. O'SULLIVAN, C. 2012. Exploring sensitivity to time-warped biological motion. *Perception 41*, 149.

SAFONOVA, A., AND HODGINS, J. K. 2005. Analyzing the physical correctness of interpolated human motion. In *SCA*, 171–180.

SAWADA, M., SUDA, K., AND ISHII, M. 2003. Expression of emotions in dance: Relation between arm movement characteristics and emotion. *Perceptual and Motor Skills 97*, 697–708.

SHAPIRO, A., CAO, Y., AND FALOUTSOS, P. 2006. Style components. In *Graphics Interface*, GI '06, 33–39.

TORRESANI, L., HACKNEY, P., AND BREGLER, C. 2007. Learning motion style synthesis from perceptual observations. In *NIPS*.

UNUMA, M., ANJYO, K., AND TAKEUCHI, R. 1995. Fourier principles for emotion-based human figure animation. In *SIGGRAPH*, 91–96.

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