2013

Emotion Capture: Emotionally Expressive Characters for Games

Cathy Ennis
Technological University Dublin, cathy.ennis@tudublin.ie

Ludovic Hoyet
Trinity College Dublin

Arjan Egges
Utrecht University

See next page for additional authors

Recommended Citation

Ennis, C., Hoyet, L., Egges, A. & McConnell, R. (2013). Emotion capture: emotionally expressive characters for games. *Proceedings of MIG'13: Motion in Games*, pp.53-60. doi:10.1145/2522628.2522633

Emotion Capture: Emotionally Expressive Characters for Games

Cathy Ennis* Ludovic Hoyet† Arjan Egges* Rachel McDonnell†

*Virtual Human Technology Lab, Utrecht University
†Graphics, Vision and Visualisation Group, Trinity College Dublin

Figure 1: *Examples of combined facial and body motions for each of our emotions. From left to right on both our male and female characters: Anger, Fear, Happiness and Sadness.*

Abstract

It has been shown that humans are sensitive to the portrayal of emotions for virtual characters. However, previous work in this area has often examined this sensitivity using extreme examples of facial or body animation. Less is known about how attuned people are at recognizing emotions as they are expressed during conversational communication. In order to determine whether body or facial motion is a better indicator for emotional expression for game characters, we conduct a perceptual experiment using synchronized full-body and facial motion-capture data. We find that people can recognize emotions from either modality alone, but combining facial and body motion is preferable in order to create more expressive characters.

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

Keywords: perception, emotion, animation

Links: DL PDF

*email: [c.ennis,j.egges]@uu.nl
†email: [hoyetl,ramcdonn]@tcd.ie

1 Introduction

Virtual characters are used for many different applications, with a range of purposes from entertainment to education. Regardless of their function, it is usually a requisite that these virtual characters are as engaging as possible. In order to give the impression of empathy, a virtual character should mimic human characteristics. Apart from visual aspects such as realistic rendering, natural human movement and the demonstration of emotions are two aspects of human nature that can bring a virtual character to life.

In the entertainment industry, virtual characters are required to be expressive and emotive. For many years, animators have focused on exaggeration of specific motions or body parts to create characters that viewers can become more engaged with. More recently, human-like detail is becoming expected from real-time virtual characters. One such example is the latest Crystal Dynamics Tomb-Raider™game, where the character's face is animated during gameplay for increased character depth and dramatic effect. Furthermore, with the advance of real-time facial motion-capture systems (e.g., [Weise et al. 2011]), it is now possible to interact with virtual characters displaying both facial and body emotion.

Previous research has investigated how emotion is portrayed on the body or face in the absence of speech. Here, we attempt to quantify where emotional information is most present, when a character is expressing emotion through speech. The recorded voice of the performing actor usually conveys both semantic and tonal information with regard to emotion, but in order to produce plausible virtual characters, it is important to understand how such emotion is conveyed through the motions of the character. We aim to investigate a number of questions, such as: during speech, does the body convey information about emotion or does it only convey conversational gestures [Krahmer and Swerts 2007]? Does lip synchrony with speech override the emotion in the face? Are there differences in how emotion is portrayed by males and females? Does the screen-size occupied by characters affect how we perceive emotion on the body or face? Our results will help to guide game developers in understanding where emotional information should be presented for characters, and whether or not special consideration should be taken into account for conveying emotion for male and female characters.

2 Related Work

Perception of human motion and emotion is an area that has been widely researched. It has been shown that humans are extremely capable of recognizing the motions of others, even with very little information [Johansson 1973]. They are also capable of deducting higher-level context, such as the identification of the talker in a dyadic conversation, from point-light data only, where a few key joints are represented by moving dots [Rose and Clarke 2009].

It has also been shown that humans have a high sensitivity to the emotions conveyed by other humans. Early research into the recognition of human emotion resulted in the classification of 6 *basic* emotions. These were emotions that were universally recognized by facial expressions, across different cultures: anger, disgust, fear, happiness (joy), sadness and surprise [Ekman 1992].

As the basic emotions have been investigated for facial expressions, so has their recognizability in the body poses and motions of humans. Atkinson et al. [2004] tested 5 of the 6 basic emotions (minus surprise) using edited video clips and point-light displays of human poses and motions. They found participants capable of recognizing each emotion, apart from disgust, for both static images and moving stimuli. Coulson [2004] found similar results when investigating body posture of a manually posed virtual mannequin.

Clarke et al. [2005] used point-light displays to represent the body motion of two actors during scripted emotional conversations (sadness, anger, joy, disgust, fear and romantic love). They found that participants were able to identify each emotion except disgust from stimuli showing both actors. For conditions where only one actor's motion was visible, the lack of a partner's body motion affected participants' performance. This implies a dependency on the social context when reading body motion for displays of non-extreme emotion.

Virtual characters can provide further insight into the perception of emotion and personality, with the additional opportunity to separate, combine and modify character features and modalities. Neff et al. [2010] combined different linguistic utterances, gesture rates and gesture performance styles on a virtual character, and asked participants to rate how extravert the character appeared. They found that linguistic utterances had the biggest effect on perceived extraversion. However, combining all three modalities resulted in a higher perception of extraversion than any single modality alone. Niewiadomski et al. [2011] investigated how important upper body multi-modal expressions were for the perception of emotional states. They investigated the sequential combination of head movement, gaze, facial expressions, gestures and torso position. Similar to Neff et al., they found that using these modalities in combination improved the recognition of many emotions compared to static image representations or animation of a single signal (such as a smile, or a bow).

Clavel et al. [2009] investigated how participants perceived 5 basic emotions (minus disgust) for blends between neutral and static poses when shown the face only, body only and both face and body of a virtual character. They found that participants were better able to recognize emotions when shown the face and body together for joy. Surprise and fear were more dependent on body posture, while sadness required face poses more for recognition. Anger was equally recognizable across modalities. We investigate whether the presence of natural motion will produce similar results to those found using blended poses.

Courgeon et al. [2011] investigated the effect of camera viewpoint on congruent and incongruent emotional poses and facial expressions for 4 emotions (anger, fear, joy and relaxation). They found that participants depended on facial expressions for incongruent conditions, but were less confident of their answer than for congruent conditions. This is somewhat contradicted by findings from Aviezer et al. [2012] who investigated the importance of body versus face cues for the recognition from still images of people experiencing intense positive or negative emotions. In this study, participants were asked to rate the valance of photographs depicting reactions to winning and losing. While they associated appropriate valance ratings with winning and losing photographs of the body only, they did not perceive any significant difference between the two for photographs depicting the face only.

Results from these studies of portrayal and perception of human emotion are relevant for animators or developers who wish to create engaging and believable characters for movies or games. However, stimuli used for these experiments are often depictions of extreme examples of the tested emotions. As a result, little is known about how emotions are perceived through natural, less contrived motions. We wish to determine whether emotions can be recognized through the body or facial animations alone as they would naturally occur in conversations. Additionally, previous studies consider the linear morphing between set poses for facial and body animation, or use only static images, neglecting to investigate the effect motion has on portrayal of emotion. For the purpose of this paper, we use fully captured face and body motions, as this is a type of animation often used in games. This method of animating gives us a rich data set, continuously capturing subtle movements. Previous studies found differences in how males and females perceive and express emotions [Biele and Grabowska 2006; Borod et al. 1983], so we expect that female and male characters might convey information differently through their body and facial motions. Therefore, our data set contains motions from a number of male and female actors, allowing us to make deductions across both sexes, and avoiding bias effects due to the subjectivity of emotion expression from person to person.

3 Stimuli Creation

We opted to use motion-capture to animate our virtual characters, since we wanted to use natural motion with plausible timing and intensity of expressions. Motion-capture was conducted using a 21 camera Vicon optical system, where 52 markers were placed on the body and 36 markers on the face. We did not capture finger or eye movements of our actors as our results are intended for typical game-standard characters. Typical motion-capture artifacts were avoided with our optimized marker set, as well as the number and appropriate placement of capture cameras. Four female and four male trained actors were recruited. During the capture session they were asked to act out a series of short sentences portraying four of the basic emotions (Anger, Fear, Happiness and Sadness). We chose to focus on these four basic emotions for this experiment due to the fact that they have been identified as highly recognizable, while previous work has found inconsistencies in the recognition of disgust and surprise [Coulson 2004; Clarke et al. 2005]. For each emotion, three sentences were chosen from a validated list of affective sentences for spoken emotion identification [Ben-David et al. 2011]. These sentences have been proven to convey emotional content appropriate to the emotion being portrayed (e.g., "Stop what you are doing and listen to me" for Anger). Each actor acted the same set of sentences for each emotion. Since we wanted to examine the effects of body motion alone, the voices of the actors, though captured, were not used for these experiments.

Two virtual characters were used to display the motions of our 8 actors, one for the male and one for the female actors (Figure 1). The sex of the actor always matched that of the virtual character to avoid motion and form ambiguity [McDonnell et al. 2007], and to avoid introducing any gender bias for certain emotions [Zibrek et al. 2013]. A bone-based approach that used linear blend skinning was used to drive the body and facial geometry. We chose this method of facial deformation over morph-based methods since this approach is most often used for game characters [Simantov 2013].

The motion-captured body markers were used to compute the joint angle animations and mapped onto the virtual characters in Autodesk 3ds Max 2012. The facial motion was directly exported as 3D marker motion, and stabilized by removing the movement of the head. The markers were first aligned to the head and then automatically adjusted to the positions of the bones on the face of the character. The characters' facial bones were then constrained to their corresponding optical markers to produce the animation.

We created three different types of motion stimuli with varying motion *Cues* (Figure 2). The first was *full motion*, where the virtual character was animated using the captured synchronized facial and body motion. The second was *body only* where the face of the character displayed a neutral static pose, and the body alone was animated. The third motion type was *facial only* where the body of the character was posed in a neutral static pose and only the intrinsic facial motion was presented. We chose not to animate the head in the facial only condition as it has previously been shown that rigid head movement does not aid recognition of emotions from dynamic facial expressions [de la Rosa et al. 2013], while recognition of emotions from virtual faces can depend on the viewpoint [Courgeon et al. 2011]. Therefore, we chose to maintain the orientation of the character's head to ensure any errors in emotion categorization were due to the expressions themselves rather than viewpoint irregularities. The character was displayed in the center of the screen, facing forward at the beginning of each clip. Movies were presented at 30 frames per second, at 1240 × 900 resolution on a 19-inch LCD screen. Two hundred and eighty-eight 3-second movies were generated in total (3 motion types × 8 actors × 4 emotions × 3 sentences).

Figure 2: *Example of our three different motion types for Fear. From (l) to (r): Full motion (combined facial and body), body only and facial only.*

4 Experiment

In our first experiment, we wished to determine whether basic emotions conveyed through naturally spoken sentences could be recognized through the body motion or the facial expressions of an actor. Additionally, we wished to investigate if the combination of both of these together resulted in higher recognition or judgments of expressiveness for a virtual character.

We presented participants with video clips of 3s duration, depicting each sentence for our four emotions, for each of our actors. We presented stimuli in 3 blocks; body only, facial only and full motion. The blocks were semi-randomized, with full motions always shown in the final block.

We had 96 trials per block: 4 Emotions (Anger, Fear, Happiness, Sadness) × 8 Actors (4M, 4F) × 3 Repetitions (3 different sentences for each emotion). Participants viewed each video clip in a random order, and after each clip were asked to answer two questions:

- *"Which of the 4 listed emotions do you think the character is conveying?"* Participants were asked to indicate their choice by pressing corresponding keys on the keyboard, marked with the acronyms ANG, FEA, HAP and SAD. They were informed at the beginning of the experiment that one of the listed emotions would be portrayed in each clip.
- *"How expressive are the motions of the character?"* Participants were asked to rate expressiveness on a scale of 1-5 using the number keys on a keyboard, with 1 representing a rating of "Not expressive at all" and 5 representing "Extremely expressive". They were instructed to base their decision on how strong an impression of the indicated emotion they saw in the motions of the character.

Fourteen participants took part in this experiment (7M, 7F; mean age: 32±13). The experiment took approximately 30 minutes in total to complete, and participants were encouraged to take breaks between each block to avoid fatigue. Participants came from various educational backgrounds and were all naïve to the purpose of the experiment. We also ensured an equal number of male and female participants to avoid any possible difference in perception [Biele and Grabowska 2006]. They were recruited via university email lists and were given a book token to compensate for their time.

4.1 Results

Results for both emotion categorization accuracy and expressiveness were analyzed using Repeated-Measure ANalyses Of VAriance (ANOVAs). We had a between-subject factor of *sex of participant* and within-subject factors were *Emotion* and *Cue* (face only, body only and full). We found no effect of participant sex on categorization accuracy or expressiveness, so we average over this in future analysis. Paired T-Tests were used for all post-hoc analysis.

4.1.1 Accuracy

The ability of participants to categorize emotions correctly was of most interest to us, so we calculated for every trial, whether or not

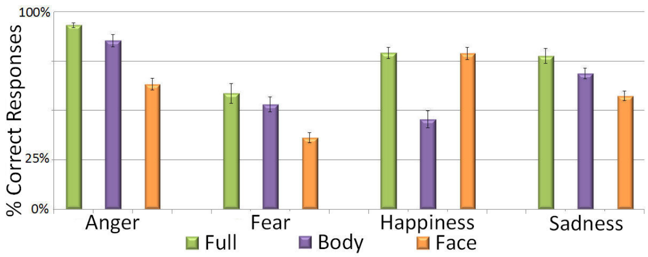

Figure 3: *Participant accuracy for our experiment, showing percentage of correct categorizations averaged over participants for each emotion. Values for full, body only and facial only are shown in green, purple and orange respectively, and chance level is indicated at 25%.*

the correct emotion was selected. This allowed us to analyze the emotion categorization accuracy of participants using ANOVA, as mentioned above. A main effect of Emotion was found ($F_{3,27} = 27.344$, $p < 0.00005$). A pairwise comparison of the means showed that participants, overall, were accurate in emotion categorization, with means ranging from 49% to 81%. *Anger, Happiness, and Sadness were categorized equally accurately, with Fear being more difficult to categorize than any other emotion* ($p < 0.005$ in all cases). This is consistent with previous research which showed Fear to be more difficult to categorize in general [Clarke et al. 2005].

We also found a main effect of Cue ($F_{2,18} = 28.396$, $p < 0.00005$), where pairwise comparisons showed that participants were equally accurate at categorizing emotion when viewing the body or face motion alone, and more accurate when presented with both cues together ($p < 0.00005$). This implies that *overall, facial and body motion cues were equally effective at conveying emotion, but were more effective when combined*.

We also found an interaction between Emotion and Cue ($F_{6,54} = 18.399$, $p < 0.00005$) (Figure 3). This interaction was due to the fact that Anger and Sadness were found to be significantly most easily identified by full, then by body only, and lastly by facial only. Fear also showed a lower level of accuracy for the face alone, but no difference between full and body alone. The opposite was true for Happiness which was categorized equally effectively when full and face only were presented, and less accurately when the body alone was presented ($p < 0.005$ in all cases).

A confusion matrix was created in order to identify where mis-categorization occurred (Table 1). From observation of this matrix, we found that confusion did not occur between dominant (Happiness, Anger) and submissive emotions (Fear, Sadness). However, there were cases where the two dominant emotions were mis-categorized, or the two submissive emotions were mis-categorized. One such example was where confusion occurred between Fear and Sadness in the face when the actor was expressing Fear. This implies that *body motion is important in the portrayal of Fear, as the facial motion alone can be misinterpreted as Sadness*. We also observed that *facial motion is particularly important in the portrayal of Happiness, as the body motion alone can be misinterpreted as Anger*. This result is similar to the findings of [Aviezer et al. 2012], where for highly intense emotions, elation and extreme disappointment were mis-categorized. However, in our case it was for the body, while they found this effect only for the face, using static poses of emotion.

4.1.2 Expressiveness

We conducted ANOVAs with the same within-subject factors and between-subject conditions as before for ratings of expressiveness for correct responses (for each clip, if the answer was incorrect, expressiveness was marked as 1, otherwise it was the recorded expressiveness rating) and found similar trends for the main effects of Cue and Emotion, as well as a significant interaction between them. Looking at the effect of Cue ($F_{2,26} = 77.467, p < 0.00005$), post-hoc analysis showed that participants found full motion to be more expressive (53%, calculated from likert ratings from 1 to 5) followed by body only (39%), with facial only conditions found to be least expressive (35%). We investigated two possible explanations for this unexpected result in a set of follow-up experiments.

We also found a main effect of Emotion and an interaction between Emotion and Cue, both showing the same trends as for participant accuracy, so we determined the correlation between participant Accuracy and accurate Expressiveness using Pearson's correlation coefficient across averages for Cue and Emotion. We found a significant positive correlation for each condition apart from two (full face and body for Anger, and body only for Sadness). The mean significant correlation was 0.786, with $p < 0.05$ in all cases. This high correlation implies that *the more expressive a motion is, the easier it is to categorize correctly*. However, correct expressiveness ratings were not as high as we had expected for these responses overall. Possible reasons for this could be the length of the clips participants saw, the lack of finger motion or audio, or the fact that they were rating less intense portrayals of emotions, as opposed to extreme demonstrations. We also investigated low expressiveness ratings in one of our post-hoc experiments.

4.1.3 Actor Sex

Previous research has shown differences in the portrayal of emotion for males and females [Borod et al. 1983]. However, this difference has rarely been considered for the accurate portrayal of emotion on virtual characters. Therefore, we averaged our results for the 4 male and 4 female actors in order to determine the effect of Actor Sex. An ANOVA was conducted on this data, as before, and we found a main effect of Actor Sex ($F_{1,13} = 29.995$, $p < 0.00005$ on accuracy and $F_{1,13} = 16.133$, $p < 0.005$ on expressiveness ratings). This effect was due to the fact that *overall, our female actors were considered more expressive and were better able to convey emotions than our male actors*. We also found an interaction between Actor Sex and Cue for accuracy of motion categorization ($F_{2,26} = 5.848$, $p < 0.05$) and for expressiveness ($F_{2,26} = 10.792$, $p < 0.00005$). As seen in Figure 4, this was due to the fact that *female bodies were more expressive and better at conveying emotion than male bodies, but there was no difference between the sexes for conveying emotion through facial expression*.

An interaction between Actor Sex and Emotion for participants' ratings of expressiveness ($F_{3,39} = 19.328$, $p < 0.005$) was also found. This was due to the fact that *female actors were more expressive when conveying Happiness, Sadness, and Fear, but males were equally expressive when conveying Anger*.

5 Follow-up Experiments

Our experiment (Section 4) revealed that, in general, facial motion was considered less expressive than body motion. We found this result surprising and further investigated through two follow-up experiments. Firstly, we investigated the possibility that our bone-based facial animation was at fault for the low expressivity. Secondly, we observed that the face occupied a much smaller area on

Displayed Emotion	Rated Anger	Rated Fear	Rated Happiness	Rated Sadness
Anger Full	**93.2%**	3.0%	3.3%	0.3%
Anger Body	**85.4%**	6.8%	6.8%	1.5%
Anger Face	**63.4%**	11.0%	11.9%	13.1%
Fear Full	19.3%	**58.6%**	1.2%	20.8%
Fear Body	27.4%	**53.0%**	4.5%	15.2%
Fear Face	14.0%	36.3%	9.8%	39.6%
Happiness Full	11.9%	2.1%	**78.9%**	7.4%
Happiness Body	32.7%	10.1%	45.5%	11.6%
Happiness Face	8.3%	5.1%	**77.5%**	7.5%
Sadness Full	10.7%	7.7%	3.9%	**77.7%**
Sadness Body	14.9%	8.9%	7.4%	**68.8%**
Sadness Face	6.5%	19.6%	16.1%	**57.4%**

Table 1: *Confusion matrix for participants' emotion selection for each condition in our experiment. The emotion selected significantly most often is highlighted where applicable.*

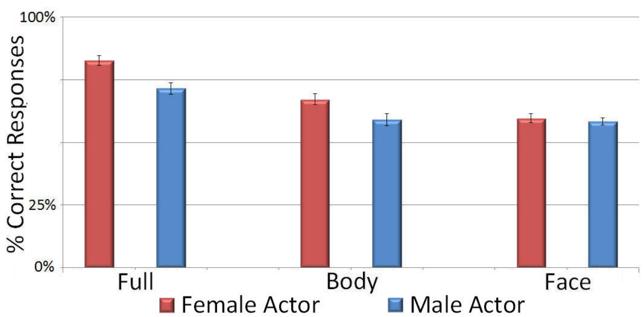

Figure 4: *Interaction between Actor Sex and Cue for participant accuracy for each cue: full, body only and facial only.*

the screen than the body, which could have affected the perception of expressiveness.

5.1 Facial Animation Validation

Previous experiments [McDonnell et al. 2009] have shown that virtual humans are equally expressive at conveying emotion as real humans, for body motion using linear blend skinning methods. However, it has often been stated that the facial animation techniques used in the games industry often lack the expressiveness and emotion of real humans. Therefore, we conducted a short experiment to test if our low ratings of expressiveness for the face were due to a lack of detail in our facial animation method. We used video footage recorded at the time of capture to compare against, in order to get an indication of the difference in expressiveness between the real and virtual faces.

We conducted this experiment over two blocks; one containing video clips of the actor recorded during the motion-capture session, and the other containing the full (face and body) virtual animations as shown in our previous experiment. This was in order to ensure as much similarity as possible across conditions.

Fourteen new participants took part in this experiment (7M, 7F; mean age: 32±10). They viewed stimuli in two blocks, viewing the real videos or virtual animations separately. The blocks were presented in a random order. The conditions within each experiment block were as before. Participants were asked to focus on the facial movements of the character/actor only, and rate how expressive they found each clip to be, on a 5-point likert scale as before. We asked participants to focus solely on the characters' facial motions as it was only the deformations within the face we wished to investigate, not the body animation.

5.1.1 Results

We averaged over participants' responses for expressiveness for each block of the experiment. A repeated-measures ANOVA was then conducted with within-subject factors of *Representation* (real video or virtual animations), *Emotion* and *Actor*. Between-subjects conditions of *Order* (whether they saw the real or virtual stimuli first) and *Sex of participant* did not have any significant effect on results.

We found an effect of Representation on participant responses ($F_{1,10} = 6.69$, $p < 0.05$), where the real video stimuli were found to be more expressive overall. However, there was not a large difference between the mean ratings (68.0% for real and 62.6% for virtual). Furthermore, we found an interaction between Actor and Representation ($F_{7,70} = 4.815$, $p < 0.0005$), where post-hoc analysis showed that participants found the facial expressions of three actors to be more expressive in the real condition. For the other five actors, they were found to be equally as expressive across both real and virtual stimuli. So, while real actors exhibit significantly more expressive facial motions than virtual characters, *the impression of expressiveness does not vary by much between the real and virtual faces*. Any loss of expression when animating the virtual characters could be explained by the fact that we did not capture any eye or finger movement during these sessions, since the animations were captured for real-time applications. Future more rigorous testing would be needed in order to identify the differences between real and virtual faces more formally.

We also found a main effect of Emotion, ($F_{3,30} = 46.444$, $p < 0.0005$). Participants found the Anger conditions to be most expressive, followed by Happiness, then Fear, with Sadness found to be least expressive. This is slightly different to our previous results, but since this refers to overall expressiveness, as opposed to expressiveness of correctly identified stimuli (as we had before), this is not altogether surprising. Previously, if there was confusion between certain emotions (such as Anger/Happiness and Fear/Sadness), then incorrect responses would have received a low expressiveness rating. However, for the purpose of this experiment, judgments of emotion expression, not identification, was the goal and expressiveness would most likely correspond to the level of activation of the actor.

5.2 Screen-size Validation

Our second follow-up experiment aimed to determine if the screen-size difference between body and face contributed to the differences

Figure 5: *Close-up views of facial expressions of Anger (l) and Happiness (r) in our second validation experiment.*

in perceived expressiveness.

For this experiment, we tested the body only and facial only blocks of our first experiment with slight modifications. The facial only condition was zoomed in on the face of the character so that the character's face occupied approximately the same number of pixels as the body of the character in our previous experiment (Figure 5). Similarly, for the body only condition, we zoomed out until the character's body occupied approximately the same number of pixels as the character's face in our previous experiment. The conditions for each experiment block remained the same as in our main experiment.

Eleven new participants took part in this experiment (8M, 3F; mean age: 31±3). As we did not find any effect of participant sex on either accuracy or perception of expressiveness in any of our previous experiments, we did not deem it necessary to ensure equal male to female participant matching for this experiment. Participants were informed as in our first experiment, and asked to answer the same questions categorizing the portrayed emotion and rating the expressiveness on a 5-point likert scale.

5.2.1 Results

We analyzed participant responses for accuracy and expressiveness separately for body only and facial only blocks, comparing with the respective responses from our first experiment. We conducted a repeated measures ANOVA, with a within-subject factor of *Emotion* and a between-subject condition of *Screen-size* for each analysis.

Firstly, we tested the body motion, and found no effect of screen-size on either participants' accuracy in categorizing emotions, or an effect on ratings of expressiveness. Therefore, for the two sizes tested, we found that *regardless of the size of the virtual body on screen, participants were equally able to recognize body emotions, and found them equally expressive*. For facial motion, we also found no main effect of screen-size on participants' accuracy at categorizing emotions. However, we did find a main effect of screen-size on ratings of expressiveness ($F_{1,23} = 6.948, p < 0.05$), where participants rated emotions viewed on a larger screen-size more expressive than those on smaller screen-sizes. This implies that *facial emotions are easily identified, regardless of the size on screen, however, they appear more expressive when viewed up-close*. So, it appears that there are subtle features of facial movements (that contribute to expressiveness) that only emerge in close-up views. Since we do not find this replicated in body motions, it is possible that these features relate the smaller range of movement of motions within the face compared to those of the body.

6 Discussion and Future Work

We have conducted perceptual experiments to investigate how emotion is perceived during expressive speech. Our main finding is the fact that, even when talking, people express emotions in both their faces and bodies, and these emotions can be perceived and identified accurately on virtual humans. This result implies that in order to plausibly convey emotion through the motion of a conversing character, both their body and facial motion should convey emotional information. Future work will investigate which particular body and facial features are used to identify emotion during speech, and whether neutral conversations can be automatically enhanced to convey the desired emotion by altering these features.

An interesting finding was the fact that motions with high activation (Anger and Happiness) were confused in the body motion but not in the face. It may be the case that any highly active body motion could appear happy when paired with a happy face, or angry when paired with an angry face. Future experiments will determine if this would in fact be possible.

Conversely, emotions with low levels of activation (Sadness and Fear) were confused in the face but not in the body motion. This suggests that appropriate body motion is important to avoid confusion in the portrayal of emotions with low activation.

With regard to male and female characters, we found differences with respect to expressiveness in body motion. Our male actors were less accurate at portraying emotion, and less expressive in their bodies than our female actors. Higher expressiveness might have resulted in higher accuracy for emotion categorization, but the motions may not be plausible for male characters. Future work will investigate these differences with respect to the sex of the virtual character displaying emotion.

We found facial and body motion combined to be most effective at conveying emotion and expressiveness. Even when displayed at small screen spaces, the relevant information was perceived and identified. It is likely that there exists a distance at which the facial motion will no longer be visible and body motion will dominate. This distance could be used for level of detail control of bone based animation for real-time crowd scenarios. Future work will investigate the effect of emotion perception in crowds using different level of detail approaches.

We also investigated a possible limitation of our approach i.e., the way in which we captured and drove the facial animation using 36 motion-capture markers on bone-based rigs. We found that this method did result in less expressive facial expressions, compared with video data. However, the mean difference in perceived expressiveness (6%) was not large enough to affect emotion identification. Another possible reason for this reduced expression from our characters could be the fact that we did not include finger or eye-motion for the purposes of this experiment. Using higher quality facial capture (e.g., [Beeler et al. 2011]) and adding finger and eye motion could result in more expressive virtual characters, perhaps disambiguating emotions that are often miscategorized, or facilitating distinctions between more subtle emotions. Approaches such as this are at present not suitable for characters displayed in video games during in-game play, and were considered outside the scope of this research. However, game cut-scenes or cinematics are beginning to employ high-quality facial motion techniques (e.g., LA Noire™) and such methods will be an important consideration for future studies. One possible option to increase perceived expressiveness would be the addition of wrinkle simulation, as previous work has shown that wrinkles, while not necessary for recognition, can improve perceived impression of expressiveness [Courgeon et al. 2009].

In the future, we plan to further investigate the combination of motion-captured facial and body motions for the expression of natural emotions. It would be interesting to determine whether synchrony between the face and body is necessary, or whether highly expressive independent face and body motions can be combined to elicit stronger ratings of expressiveness. We also plan to investigate the effect of incongruent emotions between face and body motions to determine if a richer range of emotions could be depicted in that way. It would also be interesting to investigate the creation of complex emotions through asymmetric combinations of facial motions (similar to [Ahn et al. 2012]) and assess their plausibility through expressive motions as opposed to non-dynamic expressions. Results from these experiments will provide further insight into the perception of emotion for virtual characters, and will help guide performance capture sessions for games and movies.

7 Acknowledgements

This research is funded by the European Community Seventh Framework Program, under the TARDIS project (FP7-ICT-2011-7), and by Science Foundation Ireland as part of the Cartoon Motion and Captavatar projects.

References

AHN, J., GOBRON, S., THALMANN, D., AND BOULIC, R. 2012. Conveying real-time ambivalent feelings through asymmetric facial expressions. In *Motion in Games*, vol. 7660 of *Lecture Notes in Computer Science*. 122–133.

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.

AVIEZER, H., TROPE, Y., AND TODOROV, A. 2012. Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science 338*, 6111, 1225–1229.

BEELER, T., HAHN, F., BRADLEY, D., BICKEL, B., BEARDSLEY, P., GOTSMAN, C., SUMNER, R. W., AND GROSS, M. 2011. High-quality passive facial performance capture using anchor frames. *ACM Transactions on Graphics (SIGGRAPH 2011) 30*, 75:1–75:10.

BEN-DAVID, B., VAN LIESHOUT, P., AND LESZCZ, T. 2011. A resource of validated affective and neutral sentences to assess identification of emotion in spoken language after a brain injury. *Journal of Brain Injury 25*, 2, 206–220.

BIELE, C., AND GRABOWSKA, A. 2006. Sex differences in perception of emotion intensity in dynamic and static facial expressions. *Experimental Brain Research 171*, 1, 1–6.

BOROD, J., KOFF, E., AND WHITE, B. 1983. Facial asymmetry in posed and spontaneous expressions of emotion. *Brain and Cognition 2*, 2, 165–175.

CLARKE, T., BRADSHAW, M., FIELD, D., HAMPSON, S., AND ROSE, D. 2005. The perception of emotion from body movement in point-light displays of interpersonal dialogue. *Perception 34*, 10, 1171–1180.

CLAVEL, C., PLESSIER, J., MARTIN, J.-C., ACH, L., AND MOREL, B. 2009. Combining facial and postural expressions of emotions in a virtual character. In *Proceedings of the 9th International Conference on Intelligent Virtual Agents*, IVA '09, 287–300.

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

COURGEON, M., BUISINE, S., AND MARTIN, J.-C. 2009. Impact of expressive wrinkles on perception of a virtual character's facial expressions of emotions. In *Intelligent Virtual Agents*, vol. 5773 of *Lecture Notes in Computer Science*. 201–214.

COURGEON, M., CLAVEL, C., TAN, N., AND MARTIN, J.-C. 2011. Front view vs. side view of facial and postural expressions of emotions in a virtual character. In *Transactions on Edutainment VI*, vol. 6758 of *Lecture Notes in Computer Science*. 132–143.

DE LA ROSA, S., GIESE, M., BÜLTHOFF, H., AND C, C. C. 2013. The contribution of different cues of facial movement to the emotional facial expression adaptation aftereffect. *Journal of Vision 13*, 1, 1–15.

EKMAN, P. 1992. An argument for basic emotions. *Cognition and Emotion 6*, 3, 169–200.

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

KRAHMER, E., AND SWERTS, M. 2007. The effects of visual beats on prosodic prominence: Acoustic analyses, auditory perception and visual perception. *Journal of Memory and Language 57*, 3, 396–414.

MCDONNELL, R., JÖRG, S., HODGINS, J., NEWELL, F., AND O'SULLIVAN, C. 2007. Virtual shapers & movers: form and motion affect sex perception. In *Proceedings of the 4th symposium on Applied perception in graphics and visualization*, 7–10.

MCDONNELL, R., JÖRG, S., MCHUGH, J., NEWELL, F., AND O'SULLIVAN, C. 2009. Investigating the role of body shape on the perception of emotion. *ACM Transactions on Applied Perception 6*, 3, 1–11.

NEFF, M., WANG, Y., ABBOTT, R., AND WALKER, M. 2010. Evaluating the effect of gesture and language on personality perception in conversational agents. In *Proceedings of the 10th international conference on Intelligent virtual agents*, IVA'10, 222–235.

NIEWIADOMSKI, R., HYNIEWSKA, S. J., AND PELACHAUD, C. 2011. Constraint-based model for synthesis of multimodal sequential expressions of emotions. *IEEE Transactions on Affective Computing 2*, 3, 134 –146.

ROSE, D., AND CLARKE, T. 2009. Look who's talking: Visual detection of speech from whole-body biological motion cues during emotive interpersonal conversation. *Perception 38*, 1, 153–156.

SIMANTOV, J., 2013. The last of us: Character rigs, autodesk booth sessions. Presented at Game Developers Conference (San Francisco, CA, Mar 25–29).

WEISE, T., BOUAZIZ, S., LI, H., AND PAULY, M. 2011. Realtime performance-based facial animation. *ACM Transactions on Graphics (SIGGRAPH 2011) 30*, 4, 77:1 – 77:10.

ZIBREK, K., HOYET, L., RUHLAND, K., AND MCDONNELL, R. 2013. Evaluating the effect of emotion on gender recognition in virtual humans. In *Proceedings of 2013 Symposium on Applied Perception*, to appear.