When Facial Expressions Dominate Emotion Perception in Groups of Virtual Characters

Robin Palmberg
KTH Royal Institute of Technology, Sweden
Email: robinpa@kth.se

Christopher Peters
KTH Royal Institute of Technology, Sweden
Email: chpeters@kth.se

Adam Qureshi
Edge Hill University
United Kingdom
E-mail: Qureshia@edgehill.ac.uk

***Abstract*—Virtual characters play a central role in populating virtual worlds, whether they act as conduits for human expressions as avatars or are automatically controlled by a machine as agents. In modern game-related scenarios, it is economical to assemble virtual characters from varying sources of appearances and motions. However, doing so may have unintended consequences with respect to how people perceive their expressions. This paper presents an initial study investigating the impact of facial expressions and full body motions from varying sources on the perception of intense positive and negative emotional expressions in small groups of virtual characters. 21 participants views a small group of three virtual characters engaged in intense animated behaviours as their face and body motions were varied between positive, neutral and negative valence expressions. While emotion perception was based on both the bodies and the faces of the characters, we found a strong impact of the valence of facial expressions on the perception of emotions in the group. We discuss these findings in relation to the combination of manually created and automatically defined motion sources, highlighting implications for the animation of virtual characters.**

I. Introduction

Virtual characters are an essential component for creating the compelling populated virtual worlds used in games, cinema, architecture and a host of other domains with applications from entertainment to learning. These characters may be merely graphical representations that are driven and manipulated fully by real human as avatars, or they may be sophisticated non-player characters (NPCs), or agents, that are self-driven based on autonomous controllers. A typical requirement for them is that they must be capable of engaging in social interactions which are realised visually through the precise and synchronised control of a host of non-verbal behaviours occurring across their faces and bodies.

One of many questions that remains to be investigated relates to the contributions of the face and body to the perception of emotional expressions. This paper concerns the modality (face or full-body) that is the most important when a group of characters is to express a specific set of emotions to a human audience. A scenario involving a group of three characters was created in which their faces were deformed according to manually defined blend shapes (or *morph targets*) while their bodies were animated using motion-captured data as they engaged in intense negative and positive expressions relating to fear, anger and happiness. Unlike previous studies involving both human stimuli and virtual characters, we found a strong impact of the valence of facial expressions on the perception of emotions in the group. We explain this result as a potential arising from the use of *Frankenfolk* [1] i.e. characters whose motions have been composed from a variety of different sources and the use of different control methods for their faces and bodies. Such practices are common when creating virtual characters for applications such as serious games.

This paper is organised as follows. Section II presents related work considering the contributions of facial expressions and full-body movements to the perception of emotion from individuals and groups, based on the use of both human stimuli and virtual characters. Section III presents our experiment design, setup and results, including the selection of stimuli through a pre-study for use in the main study in which the facial expressions and body motions of characters were varied. The results are discussed with respect to previous findings and future implications in Section IV. We conclude in Section V.

II. Related Work

While facial expressions of emotion have been heavily studied in humans, a large quantity of that research has focussed on faces in isolation without accounting for the important role of context on emotion perception [2]. In typical natural circumstances, faces are not experienced in isolation but are accompanied by bodies and their motions [3]. Wenzler et al. [4] focussed on intense situations and found that the facial expressions of both adults and children are not diagnostic for the valence of the situation, due to the ambiguity of extreme facial expressions. Significantly, Aviezer et al. [5] show that during intense sports situations faces are less diagnostic when distinguishing between positive and negative valence emotions, while the body maintains its diagnosticity. Overall, results suggest that the face and the body may contextualise each other and that their relative ambiguity fluctuates depending on the emotion at hand.

Studies have also considered perception and groups of virtual characters. Carretero et al. [6] evaluated the impact of varying background emotional expressions of an irrelevant crowd of characters on the perception of emotion of a small group of relevant foreground characters and found that a negative background influenced perception of the foreground. Clavel et al. [7] found that the recognition of surprise and fear were more dependent on the body posture, while sadness

relied more on the face, while Courgeon et al. [8] investigated the impact of viewing distance and camera angle on combinations of face and body expressions of emotions. Ennis et al. [9] investigated whether emotions can be recognized through the body or facial animations alone as they would naturally occur in conversations, rather then extreme emotions. Findings suggested that emotions (anger, fear, happiness and sadness) could be recognised from either modality alone, but the combination of face and body motions was preferable for more expressive characters. Ondrej et al. [1] combined voice, body, face motion and appearance modalities from different sources to create composite characters, or *Frankenfolk*. In a similar manner, this study investigates the impact of different motion source and types, but on the perception of expressions within small groups, a common situation in game scenarios.

III. Experiment

The experiment consisted of a stimuli creation phase (Section III-A) in which a set of character appearances and videos of candidate facial expressions and body motions were assembled using 3D game technologies from a set of virtual character appearance and motion assets. In the prestudy (Section III-B), participants rated the stimuli in terms of their valence for use in the main study. The main study (Section III-C) used a carefully selected subset of the stimuli from the prestudy to assemble small groups of characters with varying facial expressions and body motions.

A. Stimuli Creation

The main task involved in stimuli creation was the use of 3D animated characters and motions to create a set of video recordings of behaviour for use in the pre- and main studies. All body and face animations were created using the animation controller in the *Unity 3D* game engine with the *MCS Female* asset by Morph3D. Body animations were created using *Social Motion Pack Take 1* by PolygonCraft in combination with animations from the *Taichi Character Pack* by Game Asset Studio. These animations are generated from acted/instructed human motions. The set of facial expressions of the virtual character were created manually in Unity using blend shapes (*morph targets*) and all of the appearances had the same underlying geometric and blend shape setup. The body and face expressions fit into categories of fear and anger for the negative valance expressions and happiness for the positive expressions. These emotions were chosen as they are expressed equally well by both the face and body, are arousing (e.g. in contrast to sadness) and have been widely studied [10].

B. Prestudy

The purpose of the pre-study was to gather perceptual data concerning the stimuli in order to make a choice of expressions to be used in the main study. A set of virtual characters positioned close to the camera conducted a set of facial expressions (Figure 1) and a second set, positioned further back in the environment, conducted a set of body motions. Their faces and hands were blurred in order to ensure that they did not interfere with the study. Ratings from 21 participants were used to choose a single negative, neutral and positive facial expression from the initial set of five facial expressions (ratings of final expressions are shown in Figure 1).

Fig. 1: Selected facial expressions from the prestudy (corresponding to Figure 2, left). From left to right: negative (FE-Neg), neutral (FE-Neu) and positive (FE-Pos) valence.

A single negative, neutral and positive body motion was chosen from an initial set of four body motions based on participant ratings (see Figure 2).

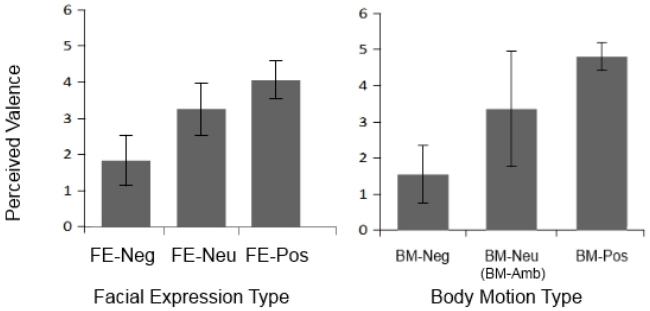

Fig. 2: Perceived valence of (left) facial expressions and (right) body motions (bars = standard deviations).

C. Main Study

1) Design and Stimuli: The stimuli selected from the prestudy were used to compose the scene for the main study. The camera was placed at eye-level in the scene with a virtual character directly facing it and two characters on either side, a common camera viewpoint in games (see Figure 3). This group formation was chosen in order to emphasise the three dimensional nature of the scene and provide additional information concerning the body motions of the characters.

The experiment used a within-subjects 3 (body motions) x 4 (facial expressions) design. Videos were created for each of the experimental conditions mixing the facial expressions *blurred*, *negative*, *neutral* and *positive* (FE-Blr, FE-Neg, FE-Neu, FE-Pos) and body motions *negative*, *ambiguous* and *positive* (BM-Neg, BM-Amb, BM-Pos), resulting in a total of 12 videos. Participants viewed each condition four times e.g. 48 video viewings in total. Each video lasted for an average of 10 seconds. The ordering of the stimuli was randomised over participants according to a Latin Square design. Tobii Studio

software was used to display the stimuli to participants in an ordered manner, to control their exposure time to the videos and to record their answers after each trial.

Fig. 3: The main experiment consisted of a group of three characters aligned so that their body motions are visible, while only the facial expression of the central character is visible.

2) Setup: The main experiment consisted of 21 University students (16M:5F) pursuing computer science courses pursuing computer science courses.15 of the participants in the main study had also taken part in the pre-study. Participants were informed about the purpose of the experiment and provided their consent to participate, before being seated approximately 60cm from a 24″ LED screen at 1920x1080 resolution. Each experiment lasted approximately 30 minutes, including the pre- and post-experiment briefings.

3) Results: The analysis consisted of a 3 x 4 within subjects ANOVA and statistical assumptions relating to distribution, outliers and sphericity were checked. The results of the study are summarised in Figures 4 and 5. There was a main effect of body motion ($F(2, 40) = 72.78, p < 0.01, = 0.78$), as well as a main effect of facial expression ($F(3, 60) = 167.89, p < 0.01, = 0.89$). Post-hoc tests showed that whilst negative body motion was perceived as significantly more negative than both the ambiguous and positive body motions (both $p's < 0.01$), there was no difference between the latter ($p = 0.37$). With respect to facial expression, there was no difference in perceived valence between the blurred and neutral conditions ($p = 0.07$), but the negative condition was rated significantly more negative than all other conditions, whilst the positive condition was rated as significantly more positive than all other conditions (all $p's < 0.01$).

There was an interaction between body motion and facial expression ($F(6, 120) = 9.23, p < 0.01, = 0.32$ (see Figure 4). This was explored further using simple main effects. For all facial expressions, there was no difference between ambiguous or positive body motion ($p's < 0.05$), but negative body motion was perceived as significantly more negative than either ($p's < 0.01$).

When the body motion was negative or ambiguous, whilst there was no difference between the blurred and neutral facial expression ($p's > 0.09$), negative facial expressions were rated as significantly more negative than all other expressions, and

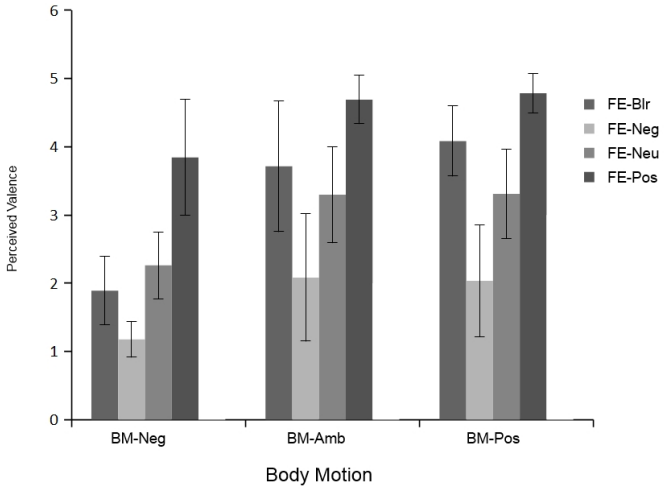

Fig. 4: Results of the perceived valence of facial expression types for each body motion category.

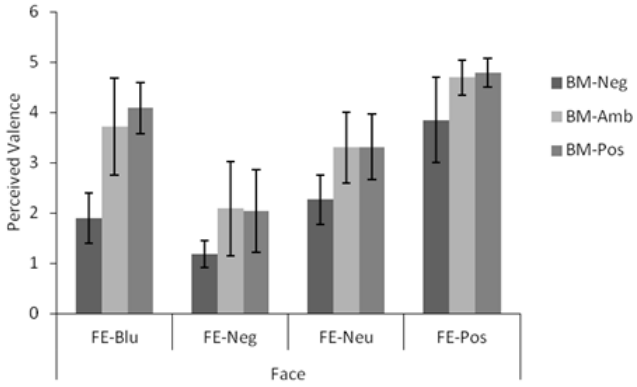

Fig. 5: Results of the perceived valence of body motion types for each of the facial expression categories.

positive facial expressions were rated as significantly more positive than all other expressions. For positive body motion, all facial expressions were perceived as significantly different, with negative facial affect perceived as the most negative, followed by the neutral expression, blurred and finally the positive expression (rated as most positive).

IV. Discussion

Overall, that there was an effect of both body motion and facial affect separately, as well as an interaction, is in agreement with previous studies suggesting people recognise emotions from both body motion and facial expressions of individuals. The valence congruent cases (BM-Neg+FE-Neg and BM-Pos+FE-Pos) were less ambiguous than the valence incongruent cases (BM-Neg+FE-Pos and BM-Pos+FE-Neg).

Comparing body only cases (BM-Neg+FE-Blr, BM-Amb+FE-Blr and BM-Pos+FE-Blr) with the *valence congruent* cases (i.e. BM-Neg+FE-Neg and BM-Pos+FE-Pos), in all three cases the addition of facial expression appears to clarify the valence of the stimuli in comparison to those cases

containing only body motion (i.e. blurred face and hands). See Figure 4. This also occurs in comparison to when a neutral face is used and a *valence incongruent* face is used (i.e. negative body motion coupled with a positive facial expression).

While there is no difference between ambiguous and positive body movement for any facial expressions (i.e. BM-Amb is always similar to BM-Pos in the FE-Blu, FE-Neg, FE-Neu and FE-Pos conditions), negative body motion (BM-Neg) is perceived as significantly more negative for all facial expressions (Figure 5). While facial expression had a similar effect on all body motions (BM-Neg $<$ BM-Amb $<$ BM-Pos), only negative body motion reduced the rating of facial expressions. This could indicate bias towards negative stimuli.

When viewed as a whole, the stimuli used in the study seem to fit well with those used in intense situations of joy, triumph and defeat, e.g. [5]. In these studies, it has been hypothesised that the ability of the different channels (face, body) to allow one to determine the emotion being signalled relates to their relative contributions to the perception of dynamic emotions [11]. A closer look at the stimuli in our study reveals that the level of control in the face and full-body are likely different to real world cases, since they came from two different sources. While the overall expressions generated by the characters seem intense, only the body motion had been captured from the movements of people. In contrast, the facial animations were manually constructed via blend shapes and were held static during the animation sequences. That is, the faces of the characters were perfectly controlled in a manner that one would not expect of real faces under similar circumstances. Therefore, in contrast to intense situations involving real stimuli in which the body dominated emotion perception due to a hypothesised break down in the diagnosticity of the face, the diagnosticity of the faces remained intact and, when viewed in the context of more intense body motions, it became the dominant cue.

A. Implications for Serious Games

While some industry and research efforts can afford to use fully motion captured characters i.e. both the face and body motions of individuals are mapped directly onto virtual characters, in many smaller efforts it is much more typical to combine blend shape controlled facial expressions that have been defined manually with full-body motions that have been recorded separately from real actors. The reasons behind this are cost and ease: manually animating the full-body of characters is difficult and motion tracking equipment is expensive, but there is a wide availability of prerecorded motion capture clips available for use in modern animation packages. On the other hand, facial animation is easier to manually control via blend shapes (*morph targets*) and does not seem to necessitate the need for facial motion capture equipment. Yet, as the results indicate in this study, care needs to be taken when mixing behaviours from different sources together, especially when the actors and/or control methods vary. The results of this study may also suggest new methods of control: For example, during intense episodes of emotion, an animator may explicitly decide not to use an original motion-captured facial expression, but instead employ a more controlled expression to help clarify the communicated emotion through the face.

V. Conclusion

This paper presents an initial study investigating the impact of facial expressions and full body motions on the perception of intense positive and negative emotional expressions in small groups of virtual characters. We found a strong impact of the valence of facial expressions on the overall perception of the emotion of the group, despite only a single face in the group being clearly visible and body motion occupying a far larger amount of the screen space. We hypothesise that our findings are based on the use of varying sources/controls for the face and body motions, an important issue for the developers of game scenarios, in which such methods are commonplace.

VI. Acknowledgements

This work has been funded by the European Commission (EC) Horizon 2020 ICT 644204 project ProsocialLearn. The authors are solely responsible for the content of this publication. It does not represent the opinion of the EC, and the EC is not responsible for any use that might be made of data appearing therein.

References

[1] J. Ondřej, C. Ennis, N. A. Merriman, and C. O'Sullivan, "Frankenfolk: Distinctiveness and attractiveness of voice and motion," *ACM Trans. Appl. Percept.*, vol. 13, no. 4, pp. 20:1–20:13, Jul. 2016. [Online]. Available: http://doi.acm.org/10.1145/2948066

[2] L. Barrett, B. Mesquita, and M. Gendron, "Context in emotion perception," *Current Directions in Psychological Science*, vol. 20, no. 5, pp. 286–290, 2011.

[3] L. Abramson, I. Marom, R. Petranker, and H. Aviezer, "Is fear in your head? a comparison of instructed and real-life expressions of emotion in the face and body," *Emotion*, vol. 17, no. 3, pp. 557–565, 2017.

[4] S. Wenzler, S. Levine, R. van Dick, V. Oertel-Knchel, and H. Aviezer, "Beyond pleasure and pain: Facial expression ambiguity in adults and children during intense situations," *Emotion*, vol. 16, pp. 807–814, Sep. 2016.

[5] H. Aviezer, Y. Trope, and A. Todorov, "Body cues, not facial expressions, discriminate between intense positive and negative emotions," *Science*, vol. 338, no. 6111, pp. 1225–1229, 2012. [Online]. Available: http://science.sciencemag.org/content/338/6111/1225

[6] M. R. Carretero, A. Qureshi, and C. Peters, "Evaluating the perception of group emotion from full body movements in the context of virtual crowds," in *Proceedings of the ACM Symposium on Applied Perception*, ser. SAP '14. New York, NY, USA: ACM, 2014, pp. 7–14. [Online]. Available: http://doi.acm.org/10.1145/2628257.2628266

[7] C. Clavel, J. Plessier, J.-C. Martin, L. Ach, and B. Morel, *Combining Facial and Postural Expressions of Emotions in a Virtual Character*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 287–300. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-04380-2_31

[8] M. Courgeon, C. Clavel, N. Tan, and J.-C. Martin, *Front View vs. Side View of Facial and Postural Expressions of Emotions in a Virtual Character*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 132–143.

[9] C. Ennis, L. Hoyet, A. Egges, and R. McDonnell, "Emotion capture: Emotionally expressive characters for games," in *Proceedings of Motion on Games*, ser. MIG '13. New York, NY, USA: ACM, 2013, pp. 31:53–31:60. [Online]. Available: http://doi.acm.org/10.1145/2522628.2522633

[10] M. Kret, K. Roelofs, J. Stekelenburg, and B. de Gelder, "Emotional signals from faces, bodies and scenes influence observers' face expressions, fixations and pupil-size," *Frontiers in Human Neuroscience*, vol. 7, no. 810, 2013.

[11] B. App, C. Reed, and D. McIntosh, "Relative contributions of face and body configurations: perceiving emotional state and motion intention," *Cogn Emot*, vol. 26, no. 4, pp. 690–698, 2012.