FACING CHARACTERS. A VOLUMETRIC INTELLIGENCE FRAMEWORK

VICTOR PARDINHO, PIA TIKKA

In the enactive virtual reality installation *The State of Darkness* (Tikka et al. 2022) a participant is facing a strange prisoner Adam B. in a stressful interrogation situation without having any clue of what binds them two together. Distant sounds from the prison corridors frame their encounter inside the prison cell abstracted to darkness. The behaviour and facial responses of the virtual character are modified by the real-time feedback from the participant's psychophysiological data, however, in a way the participant has no control of. The experience builds up around the participant's efforts to try to read the facial expressions of Adam B. Is he someone to trust, or does his restless eye-gaze signal deceitfulness. Simulating any stressful human-human encounter, in the *State of Darkness,* the digitally constructed humanlike Adam B. behaves as having a mind of its own.

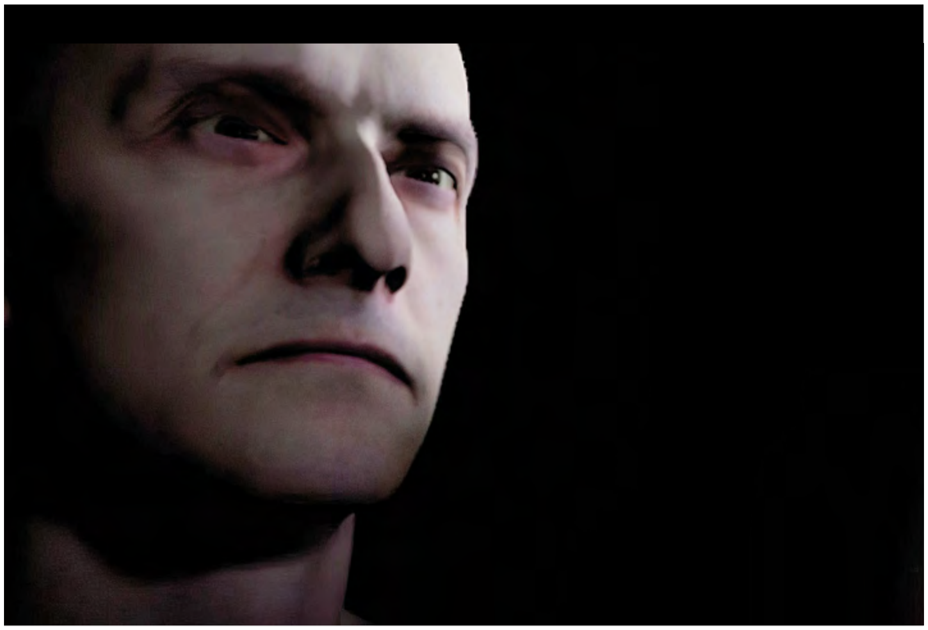

FIG. 1 *STATE OF DARKNESS*

The human-to-human encounters rely on the bodily gestures and verbal utterances of the interacting parties, but above all, they are guided by the facial non-verbal expressiveness of the two persons. The face is "constituted in as well as constitutive of social interaction" between two people (Haugh 2009: 11). In this paper we extrapolate the understanding of the dynamics of human face-to-face encounters in everyday world to the encounters between a human and an artificial character in virtual narrative context. We draw on Erving Goffman's seminal focus on the facial micro-movements and ritual expressions as those events that determine the nature of the facial encounter during copresence (Goffman 1967: 1). The virtual character's face with its eye-gaze, head position and micro expressions are interpreted as one would interpret a human face in a similar social situation. This despite the fact that the artificial humane face is a digital machinery driven by a real-time adaptive intelligent system (AI), or perhaps by a more rigid, rule-based computer program.

In this paper the novel notion of *volumetric intelligence* (Pardinho 2019) refers to the technical constitution of a tridimensional machinery of a humanlike artificial face. The attribute of volumetric means something can be created or measured in terms of three dimensional space. Intelligence, in turn, is originally related to human cognition, today also related to the term "artificial intelligence" (AI), i.e. computerized intelligence that replicates or copies dynamics of human intelligence. In the current connection, the notion of volumetric intelligence suggests that an humanlike 3D-face can be harnessed with dynamical computer-generated behaviors so natural that the face may be interpreted to be a face of another human. This would be the case in the virtual reality installation the *State of Darkness.* A co-presence is constituted between the human and the artificial character that appears having a mind of its own (Tikka et al. 2020). The term co-presence refers to the sense of "being together" in a mediated environment (de Greef/Ijsselsteijn 2000; Ho et al. 1998), involving mutual awareness of the other (Heeter 1992). The way the humane character shows concern and feelings about what happens to the things and people around it makes us care about it (Bates 1994: 122). Similarly, as in real-life, or in movies, also in the encounters in the virtual worlds, the mind and mood of the people should reflect the dynamical changes of the world and people around them.

The following discussion is two-fold: First we describe the process of creating an artificial character Adam B. for our project *The State of Darkness* by means of low-cost photogrammetry techniques, digital software automation, and with a relatively minimal workload. The description will follow the method developed in Pardinho (2019). In the second section we describe making the character alive, as well as the design principles behind the facial interaction between the participant and the humanlike character Adam B. We conclude with the challenges related to facial inter-acting for and with the artificial character.

CREATING THE VIRTUAL CHARACTER

Since *Digital Emily,* the first humanlike avatar of a live actress, was introduced (Alexander et al. 2008), a range of methods have been developed for creating humanlike characters, including the pipelines presented at the *Actor and Avatar* conference, ZHdK, by Derek Bradley from Walt Disney Research Studio Zürich (https://studios.disneyresearch.com/digital-humans/) and Matthias Wittmann from Digital Domain (https://www.digitaldomain.com/digital-humans/). While characters developed by industrial studios highlight the possibilities of the photogrammetry technologies, more time and cost effective methods need to be created. The pipeline described here, is intended mainly for the use of individual media-artists and low-cost content developers, who need to have tools for fast proto-typing of characters for their projects.

The character creation starts from capturing an actor in a tridimensional space. By taking several photographs of the actor from different angles (preferably 360°), it's possible to calculate their position and unique shape in tridimensional space. This *photogrammetry* technique "encompasses methods of image measurement and interpretation in order to derive the shape and location of an object from one or more photographs of that object" (Luhmann et al. 2007: 2). The content generated with photogrammetry techniques allow multiple viewing points to the virtual character in tridimensional environments (e.g. Virtual

Reality, or Augmented Reality). When implemented in the virtual environment, the participant may walk around the character and even interact with the character in a relatively natural manner, if such an option is provided in the design of the experience.

The workflow for capturing an actor in a photogrammetry booth is quite different from capturing an actor's performance for a traditional movie scene. Instead of a moving performance, the actor takes one by one a set of key facial expressions and holds his face still while the synchronized multi-camera system simultaneously records each expression from multiple angles. A library of photographs of key expressions of a specific actor may be also acquired from commercial libraries. For instance, our character for the VR installation *State of Darkness,* depicted in the images of this article, was retrieved from http://www.triplegangers.com. Once the set of the live actor's expressions required for the specific project are completed, they are further processed by the photogrammetry software, which generates 3D models from the 2D images using a series of processes powered with computer vision algorithms. The four stages of the workflow are briefly discussed below. (For more details see Pardinho 2019.)

CAMERA ALIGNMENT

The first step in the post-processing workflow is the alignment of the camera position for each captured photograph. The software *Agisoft* searches for common features in each photograph and matches them pixel by pixel (Agisoft 2016). Using triangulation algorithms the software generates the probable position and orientation of the camera at the moment it took the picture (Warne 2015).

DENSE POINT CLOUD

After the algorithm has generated a sparse point cloud of the actor based on the photographs, it further generates what is called a *dense point cloud* (Fig. 2). This included depth maps of each photo. By analyzing the position of the photographs and the generated sparse point cloud data, the algorithm compares the RGB values, meaning the colour information, between the tridimensional points. By this comparison, it's now possible to generate a dense point cloud, a much larger collection of 3D points in the space representing the captured object. Together with the information from the last step, knowing the photographs position and the sparse point cloud data, it produces a much higher amount of small 3D points in the space with RGB values created by analyzing the colour data of the photos (Koutsoudis et al. 2014).

FIG. 2
DENSE POINT CLOUD

3D MESH GENERATION

With the dense point cloud generated, now containing also RGB values for each 3D point in space, a 3D mesh of our captured actor can be generated. At this stage, the software implements a Structure-From-Motion (SfM) and dense multi-view stereo-matching algorithms (Koutsoudis et al. 2014) to create a mesh triangulation. This enables generation of a 3D mesh, or 3D model, of the scanned actor.

3D MESH DECIMATION

One more crucial step has to be performed before the new 3D asset (Adam B., our digital 3D actor) can be exported. This process is *Mesh Decimation* (Kobbelt et al. 1998), which consists of decimating the numbers of polygons from a tridimensional mesh. This makes the object lighter and easier to manipulate, requiring also less computer processing power.

With the set of 3D models with different expressions, (Fig. 3) it is now possible to move the work totally inside a virtual tridimensional workspace in a real-time game-engine. When the character is inside a real-time engine, it can be prepared to perform facial expressions by morphing between the different facial expressions. At this point it's possible to connect the facial expression system to other real-time workflows, which will be discussed in the following chapters.

FIG. 3
3D READY ASSETS FROM DIFFERENT SCANNED FACIAL EXPRESSIONS.

GIVING THE CHARACTER A MIND

In this section we give an overview of the technical and artistic principles related to the facial interaction between the participant and the humanlike character Adam B. of the *State of Darkness*. The discussion is extended to future work in the field. The artificially intelligent systems based on Neural Networks and Machine Learning applications allow to drive the character's facial expressions in real time over the network in such a way that it seems alive. Furthermore, complex interactions between a human and virtual character can be created to the extent that the character generates new expressions based on the learned, *giving it an appearance of having a mind of its own.*

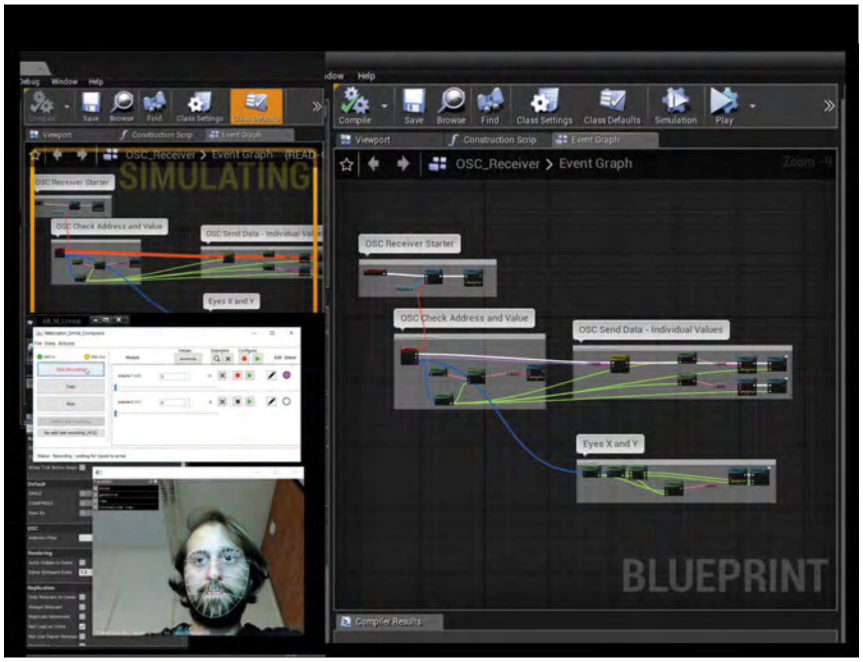

FIG. 4
OPEN SOUND CONTROL SYSTEM INSIDE UNREAL ENGINE ENVIRONMENT.

CONTROLLING CHARACTER'S FACIAL EXPRESSIONS

To control the facial expressions of the volumetric 3D replica of the actor in the Unreal Engine (UE), we used Open Sound Control (OSC) (Schmeder et al. 2009, 2010) and Unreal Engine's Blueprint programming system (plugin; Buisson 2014) (Fig. 4). The application FaceOSC (McDonald 2012) allowed us to match in real-time the facial feature data from the viewer's webcam to the facial feature data of the 3D character in UE, modifying them accordingly.

In this process, we used Machine Learning software *Wekinator* (Fiebrink/Cook 2010) to train a range of 3D character's facial expressions based on the recorded data of human dynamical expressions. In case of neutral facial expressions, for example, Wekinator will gather samples of data values that represent a human neutral expression, producing an estimate of values that can be categorized as a neutral expression. Similarly, recording smiling facial expressions, the accumulated sample data will be categorized to constitute this particular facial expression. By this training method, the application is capable of learning which range of data represents a neutral, smiling, and in principle, any other specific expression. Eventually, the virtual character's facial behaviour can be controlled according to the trained data from Wekinator. The application performs the character's facial expressions in real-time based on any actor's facial data recorded by the computer's webcam.

DESIGNING THE FACIAL INTERACTION

How the participant experiences the mood and mind of the virtual character is not only dependent on the facial or bodily expressions but also of the context within which the encounter takes place. In the VR installation *State of Darkness,* the participant's face-to-face meeting with Adam B. is contextualized within the frames of (i) the participant's prison-related imagery from lived experiences, media, news, books, films, and so on, and, (ii) Adam B.'s behavior. However, these two domains cannot be separated in the real-time recording of the participant's physiological responses, thus, in the following the recordings by the biodata are assumed to be influenced by both (i, ii). While the context-awareness cannot be emphasized enough, in this section, however, we focus on the creation of the facial expressions of an artificial character. The facial expressiveness of Adam B. includes two different layers: (iii) the pre-recorded *default* behavior based on the narrative context (without the participant's feedback) and (iv) the real-time *enactive* behavior (adaptive to the participant's feedback). (Fig. 5)

FIG. 5

PARTICIPANTS EXPERIENCING *THE STATE OF DARKNESS.*

For the *State of Darkness,* the default facial behavior (iii) was acted by a professional stage performer, who listened to the full soundtrack of the VR experience while simulating with facial expressions the imagined dynamical changes in Adam B's mindset. This professional acting-out the default states of the inner mind of Adam B. is considered crucial in the process. The facial motion capture software Face cap in iPhone (© Bannaflak) allowed to export recordings to FBX with mesh, blendshapes and animation data. The motion capture data was then implemented to the timeline of the facial feature data of the character in UE. The feedback from the psychophysiological recordings (iv) was conducted using commercially available biosensors. During the VR experience, the excitement and arousal of the participant triggered a direct response from Adam B. The stronger the emotion the participant was feeling, the stronger the emotional response of Adam B.

MEASURING PHYSIOLOGICAL SIGNALS

For the sake of installation practicalities in an open public space, we used the Empatica 4 wrist sensors, which provided relatively accurate measurements of the changes in heart rate (HR) and electrodermal activity (EDA). HR and EDA changes provide physiological quanta of arousal, aversion, and stress response. The EDA, or skin conductance, is a well-known indicator of arousal but is a very relative measure: levels of skin conductance vary between people and even same people can have different baseline level and range of activity between different recording times due to reasons such as how well the skin was cleaned and what is the temperature in the room. Therefore we needed to build a system that dynamically calibrates to the current level and range of each user. We did this by keeping a running window of recent skin conductance activity from which we constantly recalibrated the system to provide accurate feedback of the current changes in the user's level of arousal. Once the calibrated arousal levels were calculated, they were packaged into OSC (open sound control) protocol and sent over UDP network connection to the Unreal Engine.

REFERENCES

Agisoft (2016), Agisoft PhotoScan user manual: professional edition,1.2. © 2016 Agisoft LLC.

Alexander, O./Rogers, M./Lambeth, W./Chiang, J. Y./Ma, W. C./Wang, C. C./Debevec, P. (2010), The Digital Emily Project: Achieving a Photo-realistic Digital Actor. IEEE Computer Graphics and Applications, 30(4): p. 20–31.

Buisson, G. (2014), UE4-OSC. https://github.com/monsieurgustav/UE4-OSC (10.02.2022)

Ekman, P. / Rosenberg, E. L. (Eds.) (1997), What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS), Oxford University Press.

Fiebrink, R. / Cook, P. R (2010), The Wekinator: A System for Real-Time, Interactive Machine Learning in Music, in: Proceedings of The Eleventh International Society for Music Information Retrieval Conference ISMIR Utrecht.

Goffman, E. (1967), Interaction Ritual, London.

Greef, de H. P. / IJsselsteijn, W. A. (2000), Social Presence in the PhotoShare Tele-application, In: W. A. IJsselsteijn, J. Freeman, & H. Ridder, de (Eds.), Proceedings of the 3rd international workshop on presence, Presence 2000, Delft March 27–28.

Haugh, M. (2009), Face and Interaction, in Francesca Bargiela-Chiappini and Michael Haugh (Eds.), Face, Communication and Social Interaction, London, p. 1–30.

Kobbelt, L./Campagna, S./Seidel, H. P. (1998), A General Framework for Mesh Decimation. Graphics interface 98: p. 43–50.

Koutsoudis, A./Vidmar, B./Ioannakis, G./Arnaoutoglou, F./Pavlidis, G./Chamzas, C. (2014), Multi-Image 3D Reconstruction Data Evaluation. Journal of Cultural Heritage, 15(1): p. 73–79.

Luhmann, T./Robson, S./Kyle, S. A./Harley, I. A. (2006), Close Range Photogrammetry: Principles, Techniques and Applications, Dunbeath.

McDonald, K. (2012), Faceosc; https://github.com/kylemcdonald (10.02.2022)

Pardinho, V. (2019), Volumetric Intelligence: A Framework for the Creation of Interactive Volumetric Captured Characters. M.A. Diss. Aalto University Press. http://urn.fi/URN:NBN:fi:aalto-201902241788 (10.02.2022)

Schmeder, A., Freed, A., & Wessel, D. (2010), Best Practices for Open Sound Control. In: Linux Audio Conference, volume 10.

Tikka, P./Bastamow, T./Gerolin, I./Gerry, L.J./Pardinho, V./Kaipainen, M./Kosunen, I. (2022), The State of Darkness: Non-Human Narratives Embedded in the Encounters With Artificial Agents, in: Lissa Holloway-Attaway and Néill O'Dwyer (Eds.), Non-Human Narratives, Carnegie Mellon University: ETC Press.

Triplegangers (2017), http://www.triplegangers.com (02.02.2022)

Warne, M. (2015), Photogrammetric Software as an Alternative to 3D Laser Scanning in an Amateur Environment (Dissertation). http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-169911 (02.02.2022)

CHALLENGES OF INTERPRETING FACIAL EXPRESSIONS

Considering the complexity of the setting, we assume that the same facial expression in two different moments during the interaction can be interpreted in very different ways, imagine flashing a smile when hearing a painful cry, or when responding to another person's smile. It thus follows that relying merely to simple basic emotion categorizations based on the facial features (e.g. FACS, Ekman and Rosenberg 1997) can be to some extent misleading, when interpreting the situatedness of the participant and/or the character Adam B. Thus, we prefer not to name the emotional states based on pre-categorizations but, instead, use references in terms of the temporally unfolding narrative, both as consequences of the past events and anticipations of the future events. The facial expressions generated by the unreal engine are constellations of multiple feature vectors (numerical data) each assigned with control over particular points in the character' face map and modifiable in real-time. Thus, the momentary expressions of Adam B. merge as a combination of the time locked default facial data (iii) and its dynamical modifications depending on the psychophysiological feedback data from the participant (iv). In the installation, the physiological data (iii) also affected the context, creating specific changes in the audio-visual prison environment.

CONCLUSION

The natural human face as the goal in creation of artificial characters, the live actors' professional expertise remains on the spotlight in the creative processes of each individual media work. An actor gives a character its unique personality, the individual ways the character expresses attention, intentions, motivation, care, love, and hate. Novel technologies, such as 360 degree volumetric video capture of actors in action, will give the actors even more freedom to act their characters in volumetric spaces, a topic of our future research. Whatever the chosen method, the challenges of creating and controlling humanlike characters expand from the technological to the psychological domains, far beyond the challenges of creating other animated entities, such as angry birds or speaking cars. This is due to the fact that humans are evolutionarily equipped to "read" friendliness or hostility in milliseconds from other people's faces or bodily movements. Such emotive-cognitive abilities are directly transferred to the performative settings, be those on the live stage or in audiovisual media, where actors enact human behaviors in different narrative contexts. The emphasis of context-aware reading and interpretation of the facial expressions multiply exponentially the complexity of executing control over the adaptive artificial character's behavior. The challenge is best handled by transferring the control of the facial behavior of artificial characters to the artificially intelligent systems. As this is an emerging and fast advancing field, we envision that the next edition of the *State of Darkness* may already introduce a fully independent Adam B., with a repertoire of emergent self-generated humane responses to challenge the participant facing him on the other side of the table.

ACKNOWLEDGMENTS:

We thank Dr. Ilkka Kosunen for insightful comments. The writing has been funded by Sense of Space Oy and EU Mobilitas Pluss Top Researcher Grant (MOBTT90), Estonian Research Council in association with Tallinn University.