Building a Hybrid Virtual Agent for Testing User Empathy and Arousal in Response to Avatar (Micro-)Expressions

Dooley Murphy
Institute of Visual Design
The Royal Danish Academy of Fine Arts
Copenhagen, Denmark
dooleymurphy@gmail.com

ABSTRACT

This poster paper describes a hybrid (i.e., film and CG) method for capturing and implementing facial expressions for/in VR. A video camera was used to capture an actor's performance. The actor's eyes and mouth were isolated, and footage was processed as movie textures to overlay a static 3D model of a head. Micro-expressions (subtle, rapid movements of muscles in and around the eyes and mouth in particular) are thus captured in a fine-grained, yet low- cost and low-tech alternative to established techniques. A future experiment will compare the emotive efficacy of the hybrid virtual agent with that of a conventional (fully CG) rigged avatar head in a 6DoF scenario that transitions from sympathetic (gauging empathy by self-report) to confrontational (gauging physiological arousal by heart-rate or GSR). The experiment's prospective design is discussed, as well as its significance for the study of the crucial intersection of social plausibility and perceptual realism in VR.

CCS CONCEPTS

• **Human-centered computing → Virtual reality**; **HCI theory, concepts and models**;

KEYWORDS

Virtual reality; social presence; avatar capture; social plausibility

ACM Reference Format:
Dooley Murphy. 2017. Building a Hybrid Virtual Agent for Testing User Empathy and Arousal in Response to Avatar (Micro-)Expressions. In *Proceedings of VRST '17*. ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3139131.3141217

1 INTRODUCTION

Recent academic and industrial research has sought to compare conventional video with more immersive virtual reality (VR) technologies. The present project is an attempt to add to our understanding of how small, non-verbal facial cues can influence VR users' (addressees') appraisal of social plausibility relative to expressive granularity across filmed (see Figure 1) and computer graphics (CG) avatar face conditions.

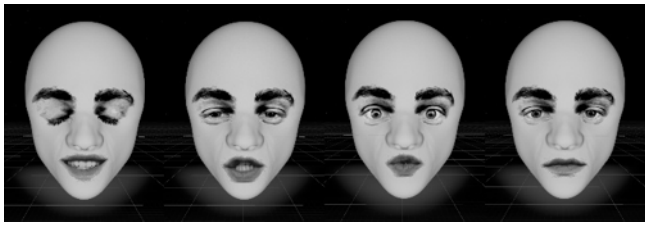

Figure 1: Some basic expressions under crude virtual lighting.

Micro-expressions can be defined as subtle, rapid facial expressions that occur within 1/25 th of a second. The human face's 43 muscles twitch and contort when we communicate, and are used (typically unconsciously) by all parties in exchanges as auxiliary, non-verbal cues.

1.1 Social Presence and Media Richness

Users' sense of social presence with embodied virtual agents in VR is strongly affected by other non-verbal cues such as proxemics and mutual gaze [Bailenson et al. 2004]. Relatedly, the concept of media richness [IJsselsteijn et al. 2003] is in line with the present project's loose hypothesis that communication media allowing for a greater breadth and depth of expressive granularity will more strongly support social presence, here appraised in terms of emotive efficacy (specifically empathy/sympathy; subjective self-report, and anxiety; psychophysiological arousal and/or behavioural measures).

1.2 Social Plausibility vs. Perceptual Realism

Studies by Slater and colleagues (e.g. [Vinayagamoorthy et al. 2005]) have shown that not only is perceptual realism a moving target, but also that user appraisals of it are influenced by a scenario's social plausibility. Naturally, a hypothetical CG avatar face could be photorealistic (perfectly proportioned; well textured; accurately lit) but still evoke the uncanny valley effect if its rigging/facial animations do not fit its voice and/or intended/expressed personality [Isbister and Nass 2000]. Complementarily, this project aims to determine whether a perceptually unrealistic avatar face (with 2D features projected crudely onto a 3D model) can be considered as socially plausible as a resource-intensive, fully-rigged CG avatar face by virtue of exhibiting a comparable degree of expressive granularity.

VRST '17, November 8–10, 2017, Gothenburg, Sweden

ACM ISBN 978-1-4503-5548-3/17/11.
https://doi.org/10.1145/3139131.3141217

2 METHODOLOGY

2.1 Preparation, Processing, Presentation

An actor's face was painted bright white, excluding areas immediately around the eyes and mouth. (Luminous green paint was found not to assist in the later colour-/luma- keying phase.) They were positioned in front of a makeshift teleprompter (a glass-fronted box angled at 45deg) containing a Canon EOS 650D shooting HD video at 50fps, 720p. A dramatic monologue was recorded, with high-quality audio being captured by an external device. Two high-contrast black dots painted between the eyes and at the top of the philtrum (below the nose) respectively were used as reference points for motion stabilisation in Adobe After Effects. Footage was imported into Adobe Premiere Pro, and brightness/contrast levels were adjusted to effectively key-out all parts of the actor's face except the eyes and mouth. Brighter areas (whites of eyes, teeth) obscured in initial processing were restored through the layered application of image masks. Footage was then exported as Apple QuickTime (.mov), as this is the only video container readable by Unity engine at the time of publishing. Footage was imported into Unity as a movie texture asset (legacy) rather than a video clip so that it could be read by a custom shader. Alpha channel transparency is not supported by QuickTime, so footage was given a bright green background which was again keyed-out using the chroma- keying ability of said custom shader within Unity.

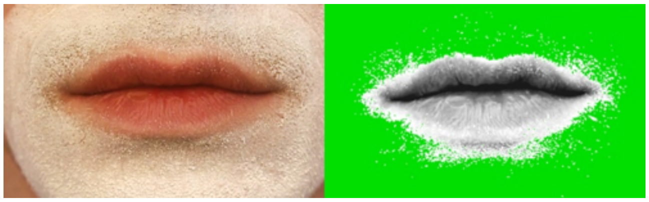

Figure 2: Raw cf. processed footage, ready for application.

Limitations of this technique include the necessity of viewing the avatar face from the front: The greater the viewing angle, the more evident it is that the facial features are projected onto 2D planes. For this reason, a C# script employing Unity's Transform.LookAt function was attached to the avatar head and directed towards the virtual camera (a HTC Vive), ensuring mutual gaze was held for the duration of the exposure condition.

2.2 Future Testing

This subsection outlines considerations of experimental design. In the near future, a rigged avatar head with a high degree of expressive granularity will be built for a second, CG experimental condition. As a dramatic production with obvious entertainment applications, the stimulus material — the totality of the virtual scenario — divided into emotional "phases" in order to build narrative momentum and achieve a sense of affective variety. This gradual shift in tone doubles as a means of separating testable factors so to avoid confounds when interpreting results. In the first few minutes of the encounter, the actor/avatar head recounts a tale of loneliness and discrimination, designed to garner sympathy from the user/viewer by means of empathy. A post-test questionnaire will seek to ascertain whether there is a difference in viewer empathy across "filmed face" and "rigged CG face" conditions, with a possible "no face" control condition. The second portion of the scenario is designed to induce a sense of anxiety, as the actor/avatar head starts to become agitated and confrontational. It moves gradually closer to the user/viewer, encroaching on their personal space. In this phase of the exposure, (psycho)physiological arousal will be measured by means of heart-rate monitor and/or skin conductance, and possibly also avoidance behavior. This is in keeping with Bailenson et al.'s recommendation (2004) that factors pertaining to social presence (among other constructs) are, when possible, best quantified by objective metrics.

3 SUMMARY

In an age and professional domain where perceptual realism is striven for (and assumed to emerge as a function of technical sophistication), "hacky," low-fidelity solutions are easily neglected. With an eye towards historic investigations into the interdependence (or separability) of perceptual realism and social plausibility, this project will test whether an expressive, filmed avatar face can elicit similar levels of affect/emotion as a rigged CG avatar head with facial animations that may or may not be deemed to traverse the uncanny valley.

REFERENCES

Jeremy N. Bailenson, Rosanna Guadagno, Eyal Aharoni, Aleksandar Dimov, Andrew C. Beall, and Jim Blascovich. 2004. Comparing behavioral and self-report measures of embodied agents' social presence in immersive virtual environments. (01 2004).

Wijnand IJsselsteijn, Joy van Baren, and Froukje van Lanen. 2003. Staying in touch: Social presence and connectedness through synchronous and asynchronous communication media. In *Human–Computer Interaction: Theory and Practice (Part 2)*, Julie Jacko and Constantine Stephanidis (Eds.). CRC Press.

Katherine Isbister and Clifford Nass. 2000. Consistency of Personality in Interactive Characters. *Int. J. Hum.-Comput. Stud.* 53, 2 (Aug. 2000), 251–267. https://doi.org/10.1006/ijhc.2000.0368

Vinoba Vinayagamoorthy, Anthony Steed, and Mel Slater. 2005. Building Characters: Lessons Drawn from Virtual Environments. In *Toward Social Mechanisms of Android Science: A CogSci 2005 Workshop*. 119–126.