Developing a Multi-modal Embodied Virtual Agent with Realistic Appearance and Emotional Response for Games

Rachel McDonnell
Trinity College Dublin
ramcdonn@tcd.ie

Edvinas Teiserskis
Trinity College Dublin

Emer Gilmartin
Trinity College Dublin

Cédric Guiard
EISKO
contact@eisko.com

Gerard Lacey
Trinity College Dublin
gjlacey@tcd.ie

ABSTRACT

Cinematic games that require interaction with users are on the rise and realism in AAA games is at an all-time high thanks to advances in real-time rendering, including recently even real-time ray-tracing. Most cinematic sequences in games look for input from the user in the form of text selection on-screen. With the increasing use of conversational agents, we believe that natural language will be important for future interactions. However, to fit with the current realism in games, we postulate that natural dialogue with a high fidelity embodied agent would be an effective delivery. Our aim is to develop realistic interactive avatars that respond not just to voice input but can also react to a range of non-verbal cues and display appropriate reactions. Our agent will have an integrated vision system to sense the users non-verbal communication and respond appropriately. We are investigating the relative roles of body language, gesture, facial expressions and avatar fidelity in the context of dyadic conversations with the human user.

CCS CONCEPTS

• **Computing methodologies → Animation**;

KEYWORDS

embodied conversational agent, affective computing

ACM Reference format:
Rachel McDonnell, Edvinas Teiserskis, Emer Gilmartin, Cédric Guiard, and Gerard Lacey. 2020. Developing a Multi-modal Embodied Virtual Agent with Realistic Appearance and Emotional Response for Games. In *Proceedings of Motion, Interaction and Games, Virtual Event, SC, USA, October 16–18, 2020 (MIG '20),* 2 pages.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 RELATED WORK

Creating a life-like agent is a challenging task and involves combining expertise from a range of fields including psychology, linguistics, computer graphics, and computer vision. The rapid development in real-time rendering technologies has enabled incredibly detailed, high-quality virtual character appearances, often reaching photorealism [6]. However, when studying the use of artificial humans as embodied agents, attempting to design a photorealistic human can sometimes induce an negative response in viewers [11] which has been attributed in part to a mismatch in realism between elements of character design [4, 8, 14]. On the other hand, in some studies, realistic characters were found to be appealing and a positive choice for viewers [2, 10]. Realistic characters might introduce other perceptual effects, such as how trustworthy people find the character [10], which information they disclose to it [12], as well as different emotional response and personality perception [13, 15].

2 SYSTEM DESIGN

We integrate existing technological components into a novel embodied conversational agent system. We use Unity 3D engine as our main platform, with a Python and Google Dialogflow back-end.

2.1 Character

Our character is a custom high-end 3D-scanned model, created by Eisko[1], a leading Digital Double company (see Figure 1). The character has over 200 blendshapes, inspired by Ekman's Facial Action Coding System with additional custom-created shapes for emotion and speech. Because the character's facial expressions have been scanned at a high resolution, it allows for high-fidelity facial expressions and micro-expressions, that would not be possible with most current embodied agents in the literature. The character is rendered using state-of-the-art shaders and advanced lighting and post-processing effects.

2.2 Animation

The synthesized speech is sent from Google's natural language understanding platform Dialogflow[2] and converted to lip movements in Unity in real-time using the Oculus Lipsync[3] Unity Plugin. Natural head and facial movements are motion-captured in advance using Hyprface[4] motion capture solution. The animation system consists of a state-machine in Unity, which contains natural animation for each state to augment the lip movements (e.g., talking, listening, thinking). We ensure multiple clips and random selection to avoid repetition. The state selection is driven by the behavioural system, implemented in the controller script. Our next

MIG '20, October 16–18, 2020, Virtual Event, SC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

[1] https://www.eisko.com/
[2] https://cloud.google.com/dialogflow/docs
[3] https://developer.oculus.com/downloads/package/oculus-lipsync-unity/
[4] https://www.hyprsense.com/hyprface

Figure 1: Our embodied virtual agent with realistic appearance for games.

step is to incorporate rule-based methods for procedural generation of non-verbal behaviours [9].

2.3 Vision System

Our current vision system senses the user's facial expressions using a standard RGB camera and the Hyprface utility which detects and tracks 50 of the user's facial action unit parameters as weight values, and outputs per-frame activation levels for each action unit. From these values, we can detect the user's expression at a higher level of detail than typical emotion-detectors which usually detect just the 6 basic emotions, as proposed by Ekman. We analyse the time series data from Hyprface to detect a number of non-verbal communication gestures such as head nods and shakes. We aim to expand this to include a wider range of body language cues from the torso and arms.

2.4 Dialogue System

We use DialogFlow as our main dialogue system, which processes the voice-based queries and sends back a response in synthesized audio. Our Python controller script sits between Dialogflow and Unity to allow for more complex dialogue functionality such as prompting, interrupting, etc. The controller script is also used for setting the system states (talking, listening, etc.) and for data collection.

In human conversation, the lexical, syntactic, and prosodic form or structure of utterances (how things are said), has long been recognised to be as important as the content (what is said), and clear differences have been shown between written and spoken language. Conversational language uses relatively simple syntax, a limited vocabulary, frequent repetition and backchanneling, and sequential rather than embedded clauses [1, 3, 5]. Utterances tend to be short, averaging less than two seconds in duration [7]. In agent design, a further consideration is text-to-speech performance – long utterances can easily be spoken by a synthesiser but sound unnatural to human ears as they seem to happen on a physically difficult single breath. To ensure that the system's conversation meets the affordances of a human user, utterances were designed to be short, and syntactically and lexically simple.

2.5 Conclusion

We present an early-stage demo of our realistic agent responding to some simple scenarios. In the future, we plan to develop novel user sensing and agent responses. Because our main application is in entertainment, we believe that the characters behaviour should be realistic but also lighthearted. We plan a series of perceptual experiments investigating the appropriateness of our avatar's appearance and dialogue in the context of novel game interactions.

REFERENCES

[1] Douglas Biber, Stig Johansson, Geoffrey Leech, Susan Conrad, Edward Finegan, and Randolph Quirk. 1999. *Longman grammar of spoken and written English.* Vol. 2. Longman London.

[2] E.J. Carter, M. Mahler, and J.K. Hodgins. 2013. Unpleasantness of animated characters increases viewer attention to faces. In *Proceedings of the ACM Symposium in Applied Perception.* 35–40.

[3] Wallace Chafe and Jane Danielewicz. 1987. *Properties of spoken and written language.* Academic Press.

[4] Thierry Chaminade, Jessica Hodgins, and Mitsuo Kawato. 2007. Anthropomorphism influences perception of computer-animated characters' actions. *Social Cognitive and Affective Neuroscience* 2, 3 (2007), 206–216.

[5] S. Eggins and D. Slade. 2004. *Analysing casual conversation.* Equinox Publishing Ltd.

[6] Epic Games, Inc. 2018. SIREN. https://www.3lateral.com/projects/siren.html. (Mar 2018). https://www.3lateral.com/projects/siren.html

[7] Emer Gilmartin, Kätlin Aare, Maria O'Reilly, and Marcin Wlodarczak. 2020. Between and Within Speaker Transitions in Multiparty Conversation. In *Proceedings of Speech Prosody 2020.*

[8] Karl F MacDorman, Robert D Green, Chin-Chang Ho, and Clinton T Koch. 2009. Too real for comfort? Uncanny responses to computer generated faces. *Computers in Human Behavior* 25, 3 (2009), 695–710.

[9] Stacy Marsella, Yuyu Xu, Margaux Lhommet, Andrew Feng, Stefan Scherer, and Ari Shapiro. 2013. Virtual Character Performance from Speech. In *Proceedings of the 12th ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA '13).* Association for Computing Machinery, New York, NY, USA, 25–35. https://doi.org/10.1145/2485895.2485900

[10] Rachel McDonnell, Martin Breidt, and Heinrich Buelthoff. 2012. Render me Real Investigating the Effect of Render Style on the Perception of Animated Virtual Humans. *ACM Transactions on Graphics* 31, 4 (2012), 91:1–91:11.

[11] M. Mori. 1970. The Uncanny Valley. *Energy* 7, 4 (1970), 33–35.

[12] Lazlo Ring, Dina Utami, and Timothy Bickmore. 2014. The Right Agent for the Job?. In *International Conference on Intelligent Virtual Agents.* Springer, 374–384.

[13] Matias Volante, Sabarish V Babu, Himanshu Chaturvedi, Nathan Newsome, Elham Ebrahimi, Tania Roy, Shaundra B Daily, and Tracy Fasolino. 2016. Effects of Virtual Human Appearance Fidelity on Emotion Contagion in Affective Inter-Personal Simulations. *IEEE Transactions on Visualization and Computer Graphics* 22, 4 (2016), 1326–1335.

[14] Eduard Zell, Carlos Aliaga, Adrian Jarabo, Katja Zibrek, Diego Gutierrez, Rachel McDonnell, and Mario Botsch. 2015. To Stylize or Not to Stylize?: The Effect of Shape and Material Stylization on the Perception of Computer-generated Faces. *ACM Trans. Graph.* 34, 6, Article 184 (Oct. 2015), 12 pages.

[15] Katja Zibrek, Elena Kokkinara, and Rachel McDonnell. 2018. The Effect of Realistic Appearance of Virtual Characters in Immersive Environments-Does the Character's Personality Play a Role? *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (2018), 1681–1690.