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