Benchmark Framework for Virtual Students' Behaviours

Socially Interactive Agents Track

Jean-Luc Lugrin
HCI Group, University of Würzburg
Jean-Luc.Lugrin@uni-wuerzburg.de

Fred Charles
Faculty of Science and Technology, Bournemouth University

Michael Habel
HCI Group, University of Würzburg

Jamie Matthews
School of Computing, Teesside University

Henrik Dudaczy, Sebastian Oberdörfer
HCI Group, University of Würzburg

Alice Wittmann, Christian Seufert
School of Pedagogy, Würzburg

Julie Porteous
School of Computing, Teesside

Silke Grafe
School of Pedagogy, Würzburg

Marc Erich Latoschik
HCI Group, University of Würzburg

ABSTRACT

This paper demonstrates the integration and evaluation of different atmosphere models into Virtual Reality (VR) training for teacher education. We developed three behaviour models to simulate different levels of class discipline. We evaluated their performances using a combination of objective and subjective measurements. Our initial results suggest that the more believable and distinguishable classroom atmospheres are produced by creating more consistent behaviours across virtual students. Our results confirm the importance of similar behaviours to elicit a particular atmosphere.

KEYWORDS

Virtual Training, Agents Behaviour Generation, Benchmarking

ACM Reference Format:
Jean-Luc Lugrin, Fred Charles, Michael Habel, Jamie Matthews, Henrik Dudaczy, Sebastian Oberdörfer, Alice Wittmann, Christian Seufert, Julie Porteous, Silke Grafe, and Marc Erich Latoschik. 2018. Benchmark Framework for Virtual Students' Behaviours. In *Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), Stockholm, Sweden, July 10–15, 2018,* IFAAMAS, 3 pages.

1 INTRODUCTION

Nowadays, numerous VR training systems are using groups of virtual humans automatically reacting to a user's performance. For instance, VR trainings reducing public speaking anxiety are simulating different types of virtual audiences (e.g., attentive or bored) [3, 7, 13]. The term *atmosphere* is used here to describe how different types of audience may be perceived and corresponds to the collective impression generated by the whole virtual audience [6].

However, creating realistic and diverse atmospheres from the portrayal of virtual humans' behaviours presents several problems. The main one is how to evaluate their quality, their believability, and how to compare different Artificial Intelligence (AI) approaches. A fundamental problem is that no clear method exists to measure *virtual humans' believability* [2]. Research on believability evaluation mostly relied on agent perception studies from videos of virtual agents interacting, followed by application domain specific questionnaires, such as understanding agent's body gesture, personality, emotions or culture recognition, or even if human or machine-controlled [1, 2, 4, 5].

Other approaches have participants building a particular atmosphere by manually assigning behaviours to a group of agents. Fukuda et al. [6] use a panel of experts to compose different virtual classroom atmospheres by assigning the students behaviours and pairing them with the intensity of 6 basic emotional states: happiness, fear, anger, surprise, disgust, and sadness. Kang et al.[8] conducted experiments using a virtual audience and its behavioural models to investigate the variations in audience characteristics resulting in perceivable audience behaviour variations.

However, no standard process nor appropriate metrics have been widely accepted. Our contribution is a novel method to develop and systematically evaluate groups of virtual humans in VR and their capacity to simulate specific atmospheres. Our objective is to produce *believable* behaviours of virtual students creating an atmosphere in line with a set level of classroom discipline (e.g., creating a difficult or easy teaching atmosphere).

2 SYSTEM OVERVIEW

We modified an existing VR training system for teacher education. This system is called: *Breaking Bad Behaviours* [10] : a collaborative VR system for classroom management (CM) skills based on low-cost portable hardware and software. It is capable of simulating VR classroom with up to 40 virtual students. They are semi-autonomous agents which can be controlled at any time by an instructor via a simple desktop Graphical User Interface (GUI). The teacher is immersed within the virtual classroom using a VR headset and 3D controllers. The instructor is an expert in CM, and evaluate the teacher's reactions to the bad or good behaviours of students. This system is integrated into seminars for teachers of primary and secondary schools, where it has been proved significantly better than traditional methods relying on video and role-play game [11].

We integrated a novel set of tools on the top of this system to improve the behaviours of the students reacting to the ones controlled by the instructor. The objective is to make them more representative of a certain classroom atmosphere, for instance very disciplined or very agitated classroom. Therefore, our tools facilitate the process of

Proc. of the 17th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2018), M. Dastani, G. Sukthankar, E. André, S. Koenig (eds.), July 10–15, 2018, Stockholm, Sweden.

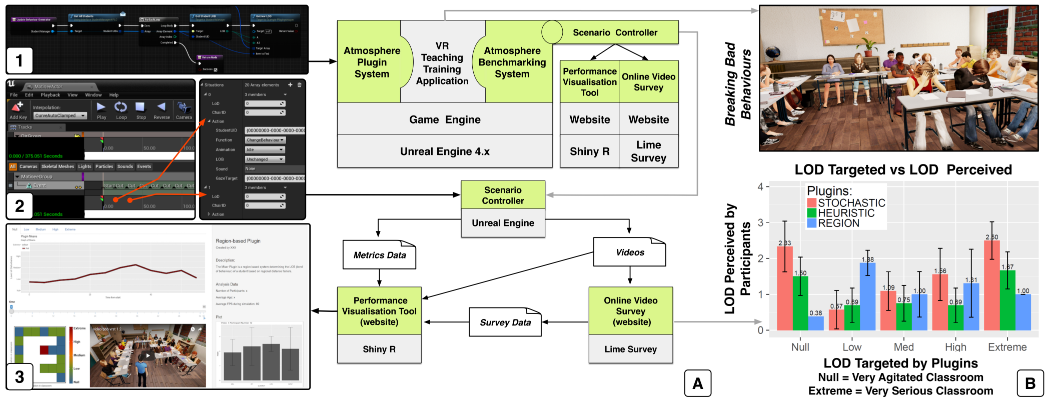

Figure 1: (A) System Architecture with new tools and sub-systems (*green*) (B) Results based on Perceived Level of Discipline.

both programming and comparing classroom atmosphere believability without having to modify the system or to run time-consuming user studies in VR.

Figure 1-A illustrates the two new subsystems:

M_1 **- *Atmosphere Plugin System***: allowing the fast integration of novel virtual student's controllers in the current training system. Their main role is to define the students' level of bad behaviour (LOB) in response to the disruptive behaviours of student controlled by a human and to modulate this by specifying a level of discipline (LOD). We defined a high-level virtual classroom library accessible via the *Blueprint* visual scripting language (Figure 1-A-1). It allows to easily access virtual students' properties, perception state, and assigning behaviours.

M_2 **- *Atmosphere Benchmarking System***: providing a way to quickly program virtual classroom scenario and compare different *Atmosphere plugin* performances on them. It automatically produces videos and collects objective and subjective data, which can be simultaneously visualized and manipulated via a website. Our benchmarking system is composed of a *Scenario Controller* (Figure 1-A-2) and a *Visualisation Tool* (Figure 1-A-3).

The *Scenario Controller* is a simple visual tool allowing the programming of a classroom benchmark scenario with different *situations* over time. Each *situation* is specifying a LOD for the classroom and a LOB for a targeted student. The system also simulates a virtual teacher (from recorded motion-capture data) to create a comparable baseline for all plugins. Videos as well as the plugin performance and behaviour metrics are collected during the execution of the benchmark scenario: *time*, *framerate*, *each student's current LOB* and *class LOD*.

The *Visualisation Tool* is an interactive web application allowing to manipulate and visualise the generated videos as well as the objective and subjective data collected for each plugin. Each plugin's video is enhanced by an interactive and animated *heatmap*, representing the students'LOB evolution and propagation on a top-down view of the classroom (as suggested by [6]). To quickly evaluate the *atmosphere* progression over time, a curve is displaying the evolution of the classroom's LOB average. Additional charts are displaying the responses collected from an online video survey using an *Atmosphere Believability* questionnaire inspired by [9] with additional questions on the perceived level of classroom discipline.

3 PRELIMINARY RESULTS

As a *proof of concept*, we implemented and evaluated three *Atmosphere* plugins. A total of 148 participants were involved in an online video survey evaluation (μ_{age}=21.21, σ_{age}=4.43, 34 males). We used a 3 X 5 between-subjects factorial design: 3 plugins (i.e. *Stochastic, Heuristic and Region-based*) and 5 targeted LOD (i.e. *Null, Low, Med, High, Extreme*). Each participant watched a single video showing the classroom behaviour produced by one plugin configured in one targeted LOD for a particular scenario. One student, called *the troublemaker*, gradually increases his level of bad behaviour in 4 main phases i) quietly listening and taking notes (LOB=*Null*), ii) daydreaming (LOB=*Low*), talking to a neighbor (LOB=*Medium*), and iv) making a 'fart' noise (LOB=*High*). The participants were asked to focus on the reaction of the other students and to evaluate the LOD (i.e., *Not all, Lowly, Highly, Medium, Extremely Disciplined*).

Overall, the *Stochastic* plugin appears as the more appropriate model to create different atmospheres matching the targeted LOD (Figure 1-B). One reason for such a result is that it is producing more consistent behaviours among students. By design, it is giving a LOB equal to the troublemaker's LOB on a random number of students in the class, while the others choose different LOB based on their distances to the troublemaker (*Region-based*) and the teacher (*Heuristic*). Our first results tend to confirm the observations made by [6, 12] regarding the correlation between the number of virtual agents with similar behaviours and the group atmosphere perception.

4 CONCLUSION

This paper presented a method and system to program and compare virtual classroom atmospheres reflecting a particular level of discipline. Our system combines benchmarking and plugin components. We demonstrated our system by integrating and evaluating three *Atmosphere* plug-ins. Our early results confirm the importance of similar behaviours to elicit a particular atmosphere. We are exploring more comparison metrics and automatic analysis. We also hope to build an interdisciplinary community of users and developers, via open-source access and benchmark publishing.

REFERENCES

[1] Elisabetta Bevacqua, Romain Richard, and Pierre De Loor. 2017. Believability and Co-presence in Human-Virtual Character Interaction. *IEEE Computer Graphics and Applications* 37, 4 (2017), 17–29.

[2] Anton Bogdanovych, Tomas Trescak, and Simeon Simoff. 2016. What makes virtual agents believable? *Connection Science* 28, 1 (2016), 83–108.

[3] Mathieu Chollet, Giota Sratou, Ari Shapiro, Louis-Philippe Morency, and Stefan Scherer. 2014. An interactive virtual audience platform for public speaking training. In *International Conference on Autonomous Agents and Multi-Agent Systems*. 1657–1658.

[4] Ionut Damian, Birgit Endrass, Peter Huber, Nikolaus Bee, and Elisabeth André. 2011. Individualized agent interactions. *Motion in Games* (2011), 15–26.

[5] Birgit Endrass, Elisabeth André, Matthias Rehm, and Yukiko Nakano. 2013. Investigating culture-related aspects of behavior for virtual characters. *Autonomous Agents and Multi-Agent Systems* 27, 2 (2013), 277–304.

[6] Masato Fukuda, Hung-Hsuan Huang, Naoki Ohta, and Kazuhiro Kuwabara. 2017. Proposal of a Parameterized Atmosphere Generation Model in a Virtual Classroom. In *International Conference on Human Agent Interaction*. 11–16.

[7] Sandra R Harris, Robert L Kemmerling, and Max M North. 2002. Brief virtual reality therapy for public speaking anxiety. *Cyberpsychology & Behavior* 5, 6 (2002), 543–550.

[8] Ni Kang, Willem-Paul Brinkman, M Birna van Riemsdijk, and Mark Neerincx. 2016. The design of virtual audiences: noticeable and recognizable behavioral styles. *Computers in Human Behavior* 55 (2016), 680–694.

[9] Patrick Kenny, Arno Hartholt, Jonathan Gratch, William Swartout, David Traum, Stacy Marsella, and Diane Piepol. 2007. Building interactive virtual humans for training environments. In *Proceedings of i/itsec*, Vol. 174.

[10] Jean-Luc Lugrin, Marc Erich Latoschik, Michael Habel, Daniel Roth, Christian Seufert, and Silke Grafe. 2016. Breaking Bad Behaviours: A New Tool for Learning Classroom Management using Virtual Reality. *Frontiers in ICT* 3 (2016), 26.

[11] Jean-Luc Lugrin, Sebastian Oberdorfer, Marc Erich Latoschik, Alice Wittmann, Christian Seufert, and Silke Grafe. 2018. VR-Assisted vs Video-Assisted Teacher Training. In *Proceedings of the 25th IEEE Virtual Reality (VR) conference*. http://hci.uni-wuerzburg.de/download/2018-ieeevr-lugrin-vr-teacher-training-poster-preprint.pdf

[12] Louis-Philippe Morency and Stefan Scherer. 2016. Manipulating the Perception of Virtual Audiences Using Crowdsourced Behaviors. In *Intelligent Virtual Agents*.

[13] David-Paul Pertaub, Mel Slater, and Chris Barker. 2002. An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence* 11, 1 (2002), 68–78.