Virtual Agent Interaction Framework (VAIF): A Tool for Rapid Development of Social Agents

Socially Interactive Agents Track

Ivan Gris
The University of Texas at El Paso
El Paso, Texas
ivangris4@gmail.com

David Novick
The University of Texas at El Paso
El Paso, Texas
novick@utep.edu

KEYWORDS

Virtual Reality; Agent Interaction Manager; Social Agents; Tools and Prototyping

ACM Reference Format:
Ivan Gris and David Novick. 2018. Virtual Agent Interaction Framework (VAIF): A Tool for Rapid Development of Social Agents . 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

Creating an embodied virtual agent is often a complex process. It involves 3D modeling and animation skills, advanced programming knowledge, and in some cases artificial intelligence or the integration of complex interaction models. Features like lip-syncing to an audio file, recognizing the users' speech, or having the character move at certain times in certain ways, are inaccessible to researchers that want to build and use these agents for education, research, or industrial uses. VAIF, the Virtual Agent Interaction Framework, is an extensively documented system that attempts to bridge that gap and provide inexperienced researchers the tools and means to develop their own agents in a centralized, lightweight platform that provides all these features through a simple interface within the Unity game engine.

2 BACKGROUND

The complexity of creating embodied conversational agents (ECAs) [1], and systems with ECAs, has long been recognized (e.g., [2]). To date, the most comprehensive support for developers of ECAs has been the Virtual Human Toolkit (VHT) [3], an extensive collection of modular tools. Among these tools, SmartBody [8] supports animating ECAs in real time, and NPCEditor [4] supports building question-answering agents. The entire collection of VHT tools is comprehensive, sophisticated, and evolving. Indeed, VHT is explicitly aimed at researchers—in contrast to developers or practitioners—who want to create embodied conversational agents. Our own experience with VHT was that setting up and using VHT can be difficult for non-researchers.

Some tools for creating ECAs have been aimed more squarely at developers. For example, the UTEP AGENT system [6] enabled creators of ECAs to write scenes, with dialog and gestures, in XML, which was then interpreted at run-time and executed via Unity. Our experience with the AGENT system was that (1) writing in XML led to run-time errors and was difficult for non-coders and (2) the system's jury-rigged connection to Unity was prone to annoying failures.

What is needed then, from the perspective of both developers and newer researchers, is an authoring tool that (a) enables developers to build human-ECA dialogs quickly and relatively error-free, and (b) that runs inside Unity, to promote efficiency and reliability. Toward this end, we developed the Virtual Agent Interaction Framework (VAIF), a platform for rapid development and prototyping of human-ECA interaction. In a nutshell, VAIF is both an authoring system for human-ECA interaction and a run-time system for executing these interactions, all built within Unity.

3 VAIF

VAIF and its applications were developed to enable research into rapport between ECA and human, with particular emphasis on multimodal, multiparty, virtual or augmented reality interaction. It is suitable for use by developers and researchers and has been tested with non-computer-science users. VAIF provides an extensive array of features that enable ECAs to listen, speak, gesture, move, wait, and remember previous interactions. The system provides a library of sample characters with animations, facial expressions, phonemes for lip-syncing, and a configured test scene. In addition, characters can be assigned a personality and an emotional state to enable the creation of more realistic and responsive characters. Each character can appear in one or multiple timelines of events, and multiple characters can appear on the same timeline.

VAIF is open source (https://github.com/isguser/VAIF), with a collection of short video tutorials (available at http://bit.ly/2FaL4bW). It can potentially be integrated with external interaction models or automated systems. VAIF requires Unity 2017.1 or later (it is backwards compatible with other versions, but functionality is not guaranteed), and Windows 10 with Cortana enabled for speech recognition. It is also possible to create an interface to other speech-recognition platforms such as Google Speech. VAIF includes two new agent functions: LookAt, which can detect where the user is looking and InRange, which serves as an invisible range that can be configured to represent how close the user is to the characters. Developers can use these component to have the user interact with characters while approaching and looking at them.

Agents' emotions in VAIF are based on Plutchik's wheel of emotions [7] but can be modified to fit other models of emotion. Likewise, personalities are based on the Myers-Briggs Type Indicator [5] but can be swapped out for other personality models.

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.

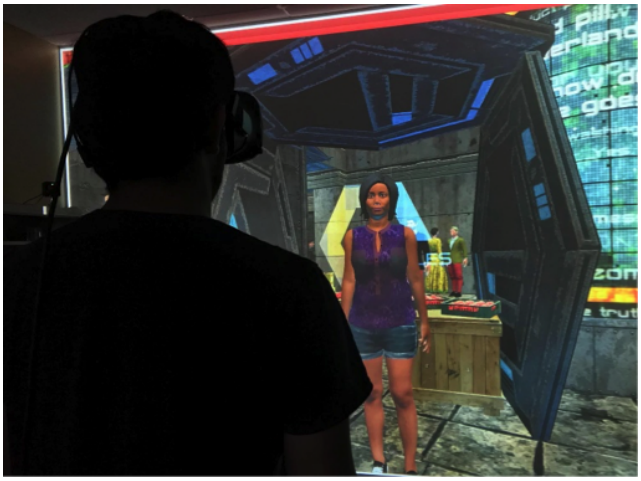

Figure 1: Room-scale VR project using VAIF, where the user walks up to the character to initiate the interaction.

Developing a fully working human-ECA interactive system in VAIF involves five steps:

(1) Choosing one or more agent models from VAIF's character gallery. If needed, as explained in Section 3.1, developers can use tools outside VAIF, such as Fuse and Mixamo from Adobe, to create characters that can be used in VAIF.
(2) Integrating and configuring the agents, using VAIF's configuration screens.
(3) Writing the interaction in VAIF's interaction manager.
(4) Creating lip-synch for the agents. This is done through an external tool such as Oculus OVRLP.
(5) Putting the agents in a virtual reality world. This can be done with an external tool such as VRTK, a Unity asset.

To port the agents into VR systems, we recommend VRTK, a Unity asset that enables developers to build scenes for multiple platforms, including Oculus and HTC VIVE. Projects can be compiled through Unity for a variety of platforms, including iOS or Android devices. AR headsets usually provide a Unity developers kit as well that can be installed as a Unity package. Although the development process remains virtually the same up until compilation, the speech recognizer provided with VAIF works only in Windows 10.

4 APPLICATIONS

VAIF has been used in commercial and research applications. One application is a commercial fighting game. When the players face the main characters, VAIF handles their conversations, which enables the game to use speech recognition to upset the enemies, to ally with them, to convince them to join the player's cause, or to stall the fight for a few moments, effectively changing the game's outcome. A second application involves a research project located in a futuristic marketplace. The goal of this project is to examine and experiment with user initiative in multiparty non-linear interactions (see Figure 1).

5 LIMITATIONS

One of the major limitations of VAIF involves the difficulty of integrating external technologies into the characters. Integration of independent stand-alone models coded in different languages (e.g., back-channels, turn-taking, prosody, discourse managers) are particularly difficult, as they must either be compiled within Unity and C#, or connected through a network and message the input and output across applications. Although it is possible for us to integrate some of these systems, the complexity and system knowledge required to do so makes one of the most-requested features unusable for average users.

Another limitation is the speech recognizer, which is restricted to Windows 10 devices. Although this is usually enough for self-contained applications, additional options are needed for cross-platform compatibility. The research team has connected Google Speech and IBM Watson to precursors to VAIF, but this requires additional external programs and network connections that place similar or worse restrictions on cross-platform use.

Finally, even when extensive written and video documentation is provided, to fully exploit VAIF's capabilities users need to create several timelines and moderate the flow of the interaction through events (i.e., prevent events or conversations from happening if other events or conversations with different characters have not occurred yet), which requires familiarity with the system and knowledge about social agents.

A tool, regardless of how good it is, does not necessarily create great interactions and ECAs. Just as a great game engine does not guarantee great games, our research efforts focus not only on the tools, but also character, dialog, and interaction design. This involves gameplay development, usability testing, and trial-and-error. Virtual agents are a multidisciplinary collaboration among behavioral scientists, writers, artists, voice actors, and programmers. The tool is meant to provide a shortcut for most of the technical implementation details, but socially aware agents go well beyond that.

6 FUTURE WORK

VAIF is the result of five years of research and over a dozen virtual-agent applications. It is an important step towards an easy-to-use social-agent development tool. We are currently using VAIF to create multi-agent applications as test-beds for studying human-ECA interaction. VAIF could be enhanced and expanded in several key respects.We plan to provide additional speech recognizers and integrate them to the extent possible to improve cross-platform compatibility, to simplify the visualization of the events through tree representations of timelines to make branching easier to observe and design from the developer's perspective, to develop a co-op mode, where multiple people can experience and interact within the same virtual environment and intelligent virtual agents.

ACKNOWLEDGMENTS

We thank Adriana Camacho, Diego Rivera, Alex Rayon, Michelle Afravi, Julie Hinojos, and Aaron Rodriguez for their contributions to the development of VAIF.

REFERENCES

[1] Justine Cassell. 2000. *Embodied conversational agents*. MIT press.

[2] Jonathan Gratch, Jeff Rickel, Elisabeth André, Justine Cassell, Eric Petajan, and Norman Badler. 2002. Creating interactive virtual humans: Some assembly required. *IEEE Intelligent systems* 17, 4 (2002), 54–63.

[3] Arno Hartholt, David Traum, Stacy C Marsella, Ari Shapiro, Giota Stratou, Anton Leuski, Louis-Philippe Morency, and Jonathan Gratch. 2013. All together now. In *International Workshop on Intelligent Virtual Agents*. Springer, 368–381.

[4] Anton Leuski and David Traum. 2011. NPCEditor: Creating virtual human dialogue using information retrieval techniques. *Ai Magazine* 32, 2 (2011), 42–56.

[5] Isabel Briggs Myers. 1962. The Myers-Briggs Type Indicator: Manual (1962). (1962).

[6] David Novick, Iván Gris Sepulveda, Diego A Rivera, Adriana Camacho, Alex Rayon, and Mario Gutierrez. 2015. The UTEP AGENT System. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*. ACM, 383–384.

[7] Robert Plutchik. 1980. A general psychoevolutionary theory of emotion. *Theories of emotion* 1, 3-31 (1980), 4.

[8] Marcus Thiebaux, Stacy Marsella, Andrew N Marshall, and Marcelo Kallmann. 2008. Smartbody: Behavior realization for embodied conversational agents. In *Proceedings of the 7th international joint conference on Autonomous agents and multiagent systems-Volume 1*. International Foundation for Autonomous Agents and Multiagent Systems, 151–158.