Leveraging Multimodal Dialog Technology for the Design of Automated and Interactive Student Agents for Teacher Training

David Pautler, Vikram Ramanarayanan, Kirby Cofino, Patrick Lange & David Suendermann-Oeft
Educational Testing Service R&D
90 New Montgomery St, San Francisco, CA
<dpautler,vramanarayanan,plange,suendermann-oeft>@ets.org

Abstract

We present a paradigm for interactive teacher training that leverages multimodal dialog technology to puppeteer custom-designed embodied conversational agents (ECAs) in student roles. We used the open-source multimodal dialog system HALEF to implement a small-group classroom math discussion involving Venn diagrams where a human teacher candidate has to interact with two student ECAs whose actions are controlled by the dialog system. Such an automated paradigm has the potential to be extended and scaled to a wide range of interactive simulation scenarios in education, medicine, and business where group interaction training is essential.

1 Introduction

There has been significant work in the research and development community on the use of embodied conversational agents (ECAs) and social robots to enable more immersive conversational experiences. This effort has led to the development of multiple software platforms and solutions for implementing embodied agents (Rist et al., 2004; Kawamoto et al., 2004; Thiebaux et al., 2008; Baldassarri et al., 2008; Wik and Hjalmarsson, 2009). More recently, there has also been a push towards developing ECAs that are empathetic (Fung et al., 2016) and are directed toward specific educational applications such as computer-assisted language learning (CALL) (Lee et al., 2010), including the possibility of providing targeted feedback to participants (Hoque et al., 2013). The degree of realism and immersiveness of the interaction experience can elicit varying behaviors and responses from users depending on the nature and design of the virtual interlocutor (Astrid et al., 2010).

Figure 1: Screenshot of the two virtual student avatars that teacher candidates interact with

2 Task Design

The task we used for our prototype implementation asks participants to imagine themselves in the role of a 2nd grade teacher leading a classroom discussion on the purpose and function of Venn diagrams with two ECAs designed to behave as students (see Figure 1). We provided participants with a stimulus Venn diagram (shown in Figure 2) in which one item, *fish*, is purposefully placed in the wrong place to serve as a catalyst for a small-group discussion. The learning goals for the discussion are to effectively evaluate the Venn diagram for its accuracy, while considering the similarities and differences between lakes and oceans. Further, one of the ECAs is designed to manifest a certain misunderstanding of this particular Venn diagram—that fish belongs outside all the circles—but the ECA does not reveal this misunderstanding unless it is asked to comment. The teacher candidate must engage both students in conversation, diagnose potential misunderstandings, and then correct those misunderstandings through dialog interactions.

Proceedings of the SIGDIAL 2018 Conference, pages 249–252,
Melbourne, Australia, 12-14 July 2018.

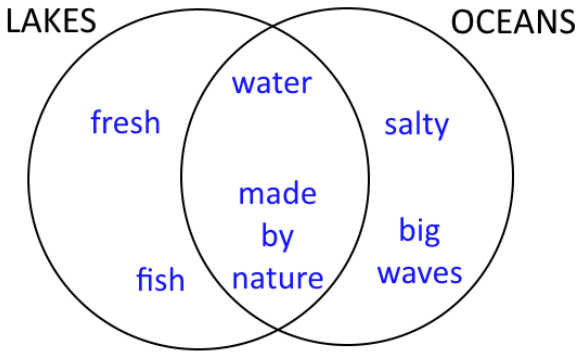

Figure 2: The Venn diagram that the trainee discusses with the ECAs

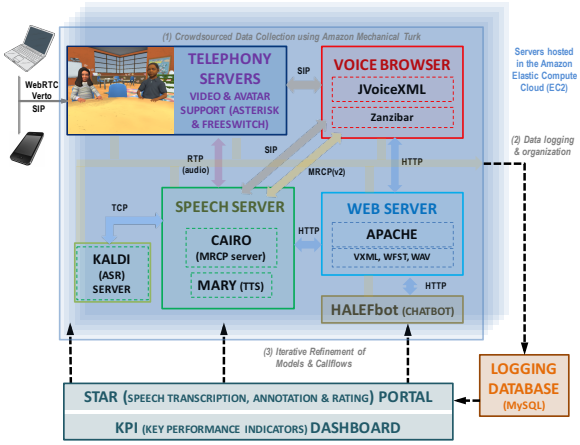

Figure 3: The HALEF multimodal dialog framework with ECAs to support educational learning and assessment applications.

3 System Design and Implementation

This section first describes our existing dialog framework. It then discusses the authoring process, in which the final step is integration of the 3D classroom user interface (UI) with the HALEF dialog system[1]. Note that work described in this paper builds on our previous efforts in building virtual avatars for job interviewing (see for example (Ramanarayanan et al., 2016; Cofino et al., 2017)). While designing such experiences for users and authors, we aim for several high-level goals:

- *The simulation must be available to potential users across the globe with as little setup as possible.* This goal implies that we avoid requiring software to be installed, if possible, and that we make the experience as accessible as possible.
- *The activity must be realistic and immersive.* Research has shown that engagement is higher with on-screen ECAs than without (and higher yet with physical embodiments such as robots) (Sidner et al., 2005; Rich and Sidner, 2009), and higher engagement might provide more effective training.
- The authoring tools/resources must be as open, low-cost, easy-to-use, and well-supported as possible.
- It must be possible to control the ECAs remotely from the HALEF system and to sync the mouth motions and gestures of the ECAs with the audio of the ECAs' speech.

To fulfill these goals, we decided to use the Unity 3D[2] authoring tool, because it allows a game to be built as a WebGL[3] resource that can be hosted in a web page, thereby saving users from having to install anything. The following subsections describe how we integrated a Unity WebGL resource with HALEF.

3.1 Resources for Authoring

We used the Blender 3D modeling tool[4] to create several of our scenes and ECAs[5]. We also explored creating animations through the motion-capture capabilities of Microsoft Kinect. While both these methods are effective and complement each other, we found both of these to have a steeper learning curve than application designers (content matter experts who are not necessarily expert software engineers) might find acceptable, and they both require substantial time and expertise to develop ECAs of optimal quality. Therefore, going forward, we will work toward creating and maintaining an open repository of scenes, characters, and animations created by game-authoring experts[6].

When scenes, characters, and animations are assembled in Unity and built, they are still non-responsive because there is no way of sending commands (yet). One must add code to the web

[1] http://halef.org

[2] https://unity3d.com/

[3] https://developer.mozilla.org/en-US/docs/Web/API/WebGL_API

[4] https://www.blender.org

[5] We worked off assets originally created for us by Mursion, Inc.

[6] The public repository might not include the 3D models shown here as they are proprietary.

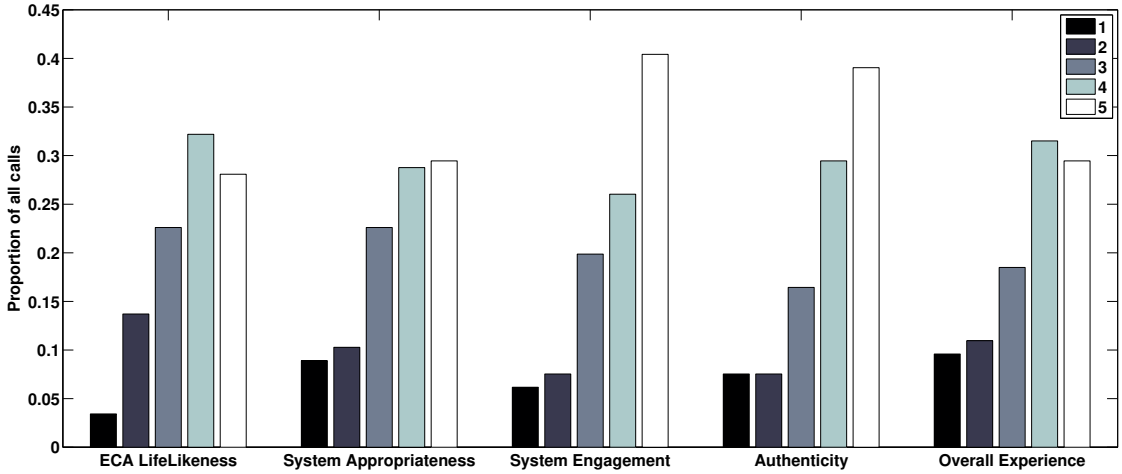

Figure 4: Crowdsourced ratings aggregated from 146 calls during user acceptance tests on Amazon Mechanical Turk.

page to receive commands over the network, as well as to the Unity files in order to route commands to a particular character. We bundled code to support these functions into a new Unity "WebGL template" that is easy to import into new Unity projects. The code includes a JSON configuration file that specifies all information required to connect to the HALEF dialog system. After an author imports this template, she updates the HTML, CSS, and JSON to fit the task (e.g. showing a static image of a Venn diagram), she builds the template as a "WebGL build", and the result is a set of files comprising a website.

For the backend, the author creates a dialog callflow using the Eclipse-based OpenVXML toolkit[7]; the author exports the callflow as a Java-based WAR file and HALEF hosts it on an Apache Tomcat server, similar to the way many HTML-only applications that have dynamic server-based logic are hosted.

To control ECAs from a callflow, the callflow must have nodes containing scripts that send commands over the network to the website. These commands include references to animations that should be triggered, as well as the ECA that should perform them. When an ECA speaks, the command that triggers the audio and mouth motions just identifies the ECA and the audio file. Part of front-end configuration is a sequence of animation-like "blendshape" settings to move the mouth into different phoneme-related shapes (this sequence of blendshape settings is generated from a forced alignment speech recognition tool that is currently proprietary).

4 User Acceptance Tests

We used the Amazon Mechanical Turk crowdsourcing platform to do user acceptance testing (UAT). We collected data from 146 crowd workers interacting with the ECAs. Following their interaction, the workers were also requested to rate, on a scale from 1–5 (with 1 being least satisfactory and 5 being most satisfactory), the following:

1. *ECA Lifelikeness*: How realistic and life-like were the ECAs over the course of the interaction?
2. *Appropriateness*: How appropriate were the system (or ECAs') responses to the questions posed by the user?
3. *Engagement*: How engaged were users while interacting with the ECAs?
4. *Authenticity*: How authentic were the responses of the ECAs, considering that they were supposed to represent students?
5. *Overall Experience*: How was the overall user experience interacting with the application?

Figure 4 plots the results of this user survey. We observe that users gave predominantly positive ratings to all aspects of the survey, with a majority proportion assigning ratings of 4 or 5. This also suggests that the lifelikeness of the ECAs and the appropriateness of system responses warranted the most improvement.

[7] https://sourceforge.net/p/halef/openvxml

5 Conclusions

We have presented a multimodal dialog-based teacher training application involving more than one virtual agent to create an immersive and interactive classroom simulation experience. Future work will look at leveraging the results of our user acceptance tests to improving the naturalness of the ECAs and the interaction, as well as in designing the simulation to be more adaptable to the engagement level of users. We will also explore the addition of more student avatars and different situational contexts.

Acknowledgments

We would like to thank David Dickerman and Eugene Tsuprun, who helped with the design and creation of the Venn Diagram task.

References

M Astrid, Nicole C Krämer, Jonathan Gratch, and Sin-Hwa Kang. 2010. it doesnt matter what you are! explaining social effects of agents and avatars. *Computers in Human Behavior*, 26(6):1641–1650.

Sandra Baldassarri, Eva Cerezo, and Francisco J Seron. 2008. Maxine: A platform for embodied animated agents. *Computers & Graphics*, 32(4):430–437.

Kirby Cofino, Vikram Ramanarayanan, Patrick Lange, David Pautler, David Suendermann-Oeft, and Keelan Evanini. 2017. A modular, multimodal open-source virtual interviewer dialog agent. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pages 520–521. ACM.

Pascale Fung, Dario Bertero, Yan Wan, Anik Dey, Ricky Ho Yin Chan, Farhad Bin Siddique, Yang Yang, Chien-Sheng Wu, and Ruixi Lin. 2016. Towards empathetic human-robot interactions. *arXiv preprint arXiv:1605.04072*.

Mohammed Ehsan Hoque, Matthieu Courgeon, Jean-Claude Martin, Bilge Mutlu, and Rosalind W Picard. 2013. Mach: My automated conversation coach. In *Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing*, pages 697–706. ACM.

Shin-ichi Kawamoto, Hiroshi Shimodaira, Tsuneo Nitta, Takuya Nishimoto, Satoshi Nakamura, Katsunobu Itou, Shigeo Morishima, Tatsuo Yotsukura, Atsuhiko Kai, Akinobu Lee, et al. 2004. Galatea: Open-source software for developing anthropomorphic spoken dialog agents. In *Life-Like Characters*, pages 187–211. Springer.

Sungjin Lee, Hyungjong Noh, Jonghoon Lee, Kyusong Lee, and Gary Geunbae Lee. 2010. Cognitive effects of robot-assisted language learning on oral skills. In *INTERSPEECH 2010 Satellite Workshop on Second Language Studies: Acquisition, Learning, Education and Technology*.

Vikram Ramanarayanan, Patrick Lange, David Pautler, Zhou Yu, and David Suendermann-Oeft. 2016. Interview with an avatar: A real-time engagement tracking-enabled cloud-based multimodal dialog system for learning and assessment. In *Proceedings of the Spoken Language Technology (SLT) Workshop, San Diego, CA*.

Charles Rich and Candace L Sidner. 2009. Robots and avatars as hosts, advisors, companions, and jesters. *AI Magazine*, 30(1):29.

Thomas Rist, Elisabeth André, Stephan Baldes, Patrick Gebhard, Martin Klesen, Michael Kipp, Peter Rist, and Markus Schmitt. 2004. A review of the development of embodied presentation agents and their application fields. In *Life-Like Characters*, pages 377–404. Springer.

Candace L Sidner, Christopher Lee, Cory D Kidd, Neal Lesh, and Charles Rich. 2005. Explorations in engagement for humans and robots. *Artificial Intelligence*, 166(1-2):140–164.

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*, pages 151–158. International Foundation for Autonomous Agents and Multiagent Systems.

Preben Wik and Anna Hjalmarsson. 2009. Embodied conversational agents in computer assisted language learning. *Speech Communication*, 51(10):1024–1037.