Towards Rapid Development of Conversational Virtual Humans using Web3D Technologies

Mohan Zalake
Alexandre Gomes de Siqueira
Krishna Vaddiparti
University of Florida
Gainesville, Florida

Pavlo Antonenko
Felix Hamza-Lup
University of Florida
Gainesville, Florida
Georgia Southern University
Savannah, Georgia

Benjamin Lok
University of Florida
Gainesville, Florida

ABSTRACT

In this paper, we propose a framework for the rapid development of conversational virtual humans (VHs) via the web. Web3D formats, such as WebGL, have enabled the use of interactive VHs on the web. Web-based conversational VHs are found to have great potential in improving population health. However, the development of web-based VHs needs significant time and effort which can lead health experts to avoid VH technology. In this paper, we present a case study to highlight the challenges that lead to significant time and effort in the development of web-based conversational VHs. Based on our findings, we propose a framework that uses Web3D formats and has the potential to address identified challenges in the rapid development of web-based conversational VH systems.

CCS CONCEPTS

• **Computing methodologies** → **Virtual reality**.

KEYWORDS

Virtual Human, X3D, Web3D, Conversation, Real-time

ACM Reference Format:
Mohan Zalake, Alexandre Gomes de Siqueira, Krishna Vaddiparti, Pavlo Antonenko, Felix Hamza-Lup, and Benjamin Lok. 2020. Towards Rapid Development of Conversational Virtual Humans using Web3D Technologies. In *The 25th International Conference on 3D Web Technology (Web3D '20), November 9–13, 2020, Virtual Event, Republic of Korea.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3424616.3424727

1 INTRODUCTION

With current advances in 3D graphics on the web and Internet speeds, the use of web-based conversational virtual humans (VHs) has shown great potential to improve population health [Provoost et al. 2017]. Prior work has shown that VHs are effective in promoting positive health behaviors. They enable face-to-face conversations like a real human while maintaining anonymity for users in difficult conversations. As they are deployed via the web, they can be standardized and targeted to large populations at reduced costs.

Web3D '20, November 9–13, 2020, Virtual Event, Republic of Korea

ACM ISBN 978-1-4503-8169-7/20/11.
https://doi.org/10.1145/3424616.3424727

With the current SARS-CoV-2 pandemic and our ongoing work with VHs in health, we predict an increase in the use of web-based conversational VHs in different fields. Several experts from different fields, such as health, aim to address specific problems in their domain using web-based conversational VHs. Web-based conversational VHs require the development of several components, but not limited to, 3D VHs, animations, conversational dialogue, a dialog manager, and rendering VHs on the web. There have been multiple frameworks proposed to streamline the development of VHs using Web3D formats (e.g., X3D, WebGL) [Ieronutti and Chittaro 2005; Jung and Behr 2008]. However, the development process often requires computer science expertise and still suffers from Web-browser compatibility issues, which makes adoption of VH technology challenging for non-computer science experts. The time, cost, and expertise required to develop and deploy VHs on the web can lead experts from non-computer science domains to avoid VH technology. To increase the adoption of VH technology in different domains, the Web3D community could look into the development of standardized tools for web-based conversational VHs.

In this paper, we discuss the challenges in the rapid development of web-based conversational VHs, in the context of health. We then propose a framework to address the identified challenges in the rapid development of web-based conversational VHs.

2 CASE STUDY

We first present a case study to identify and discuss the challenges for the rapid development of conversational VHs. In the case study, our research team aimed to deploy a conversational VH to improve the mental health of college students during the SARS-CoV-2 pandemic. The goal was to establish a framework for the evaluation of VHs as catalytic factors in persuading students to engage in a number of activities that can improve mental health. We identify the following challenges for rapid deployment of web-based conversational VHs at different stages: 1) Developing a conversational script for VH with a mental health expert is a iterative process and introduces difficulties in communication due to differences in backgrounds. 2) Development of additional infrastructure, such as a dialog manager to support complex non-linear conversation with VH, a server to host the interaction, data logging capabilities, are non-trivial and increase the development time of web-based conversational VHs. 3) The implementation of VH features, such as real-time lip synchronization, further adds delay in deployment. 4) The 3D content size must be optimized for reasonable download times when interacting on browsers.

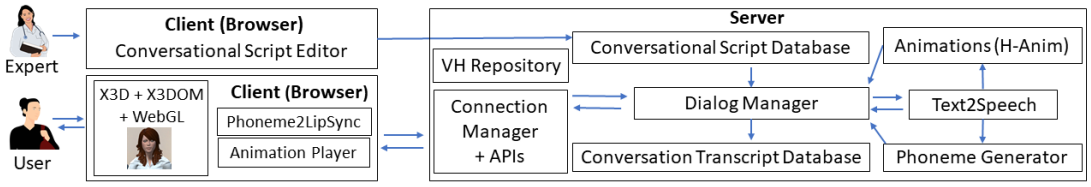

Figure 1: Framework for rapid development of conversational VHs using Web3D technologies.

Before we built such a complex system that addresses the challenges, we empirically evaluated a simple system which showed pre-recorded videos of VH to students in a user study. The findings of user study would help developers decide if the effort in building a complex web-based VH system is worthwhile.

2.1 Establishing the need for the framework

Our study group consisted of 15 senior computer science students (n=15, age =19-27 years, females=2). The students interacted with a Caucasian female VH. The students were asked about the frequency in which they engaged in a number of activities for 20 minutes during two weeks prior to the intervention and after the intervention. The analysis of pre and post-test shows that students spent more time with breathing exercises after the intervention (M = 2.071, SD = 0.917) as opposed to the two weeks prior to the intervention (M = 1.428, SD = 0.646), $t_{13} = 2.590$, $p = 0.022$.

The findings suggest that the intervention with VHs was successful at changing the behavior of students. Developing a framework that can enable the rapid development of VH applications, such as user study described in our paper, can have a significant impact on population health. We now propose a framework that can address the challenges highlighted in the rapid development of VHs.

3 FRAMEWORK

To address the challenges using framework, the system requirements were set up. The system should: 1) enable experts from non-computer science domains to easily create conversational dialogs, 2) be compatible and scalable across platforms, 3) support real-time speech synchronized lip and body animation, 4) generate smaller size content for reasonable download times.

We now describe a framework (illustrated in figure 1) that addresses the requirements. The framework uses a scalable client-server model that enables experts to create conversational VHs and deploy conversational VHs via the web. In the framework, the experts are not required to involve VH developers to update the VH dialogs. The experts can modify VH dialogs using the Conversational Script Editor and choose one of the existing VHs from the repository. The server then creates a web page with a 3D VH which can converse with users using the script authored by experts.

To enable cross-platform compatibility, the framework proposes the use of X3D along with WebGL and X3DOM. X3D enables the use of the H-Anim standard for representing and exchanging VH models and animations through the web (e.g., H-Anim Architecture V2 for representing humanoids in a network-enabled 3D graphics and multimedia environment [Web3D Consortium 2019]). However, including several H-Anim animations with the VH scene can increase the size of the 3D content and thereby increase the initial loading time of a web page. To provide reasonable load times and keep the VH scene generic for reuse, the animations are served dynamically through the server. The dynamic loading of hierarchical data is supported by X3DOM and can further benefit from the ongoing efforts in compression for X3D data transfer [Web3D Consortium 2015].

In the framework, the client initiates a request for the server to start the conversation. The conversation is managed by the Dialog Manager. The dialog manager instantiates a new instance of the conversational script developed by the expert. The dialog manager responds to the client's request with a dialog for VH. Along with dialog, the following data is sent from the server: speech generated from Text2Speech module, different options for user to respond, humanoid body animations generated from Animations module, and phonemes information for lip animation using phoneme generator. Phonemes information, which includes both: types of phonemes and durations, is mapped to respective visemes for lip animation by the client. The viseme mapping is implemented on the client to avoid constraints. Since visemes are not standardized, the client can choose to implement custom visemes and map to phonemes as needed. All conversations are logged on the server for future reference on Conversation Transcript Database. The server also exposes several APIs via Connection Manager so that VHs on any platform can have conversational capabilities.

4 CONCLUSION AND FUTURE WORK

We proposed a framework to address the highlighted challenges in the rapid development of web-based conversational VHs. The framework allowed rapid generation of web-based conversational VHs by providing platform-independent conversational capability for VHs and enabling non-computer science experts to directly modify VH dialogs. The framework has been implemented into a platform called RM^3. Currently, a team of researchers from multiple universities are testing the feasibility of the RM^3 platform for enhancing mental health among university students.

REFERENCES

Lucio Ieronutti and Luca Chittaro. 2005. A virtual human architecture that integrates kinematic, physical and behavioral aspects to control H-Anim characters. In *Web3D Symposium Proceedings*. 75–83. https://doi.org/10.1145/1050491.1050502

Yvonne Jung and Johannes Behr. 2008. *Extending H-Anim and X3D for Advanced Animation Control.*

Simon Provoost, Ho Ming Lau, Jeroen Ruwaard, and Heleen Riper. 2017. Embodied conversational agents in clinical psychology: A scoping review. *Journal of Medical Internet Research* 19, 5 (2017), 1–27. https://doi.org/10.2196/jmir.6553

Web3D Consortium. 2015. X3D Compressed Binary Encoding Activity | Web3D Consortium. https://www.web3d.org/working-groups/x3d/compressed-binary-encoding-activity

Web3D Consortium. 2019. HAnim Architecture V2 | Web3D Consortium. https://www.web3d.org/content/hanim-architecture-v2