Say Hi to Eliza

An Embodied Conversational Agent on the Web

Gerard Llorach[1,2,3]✉ and Josep Blat[1]

[1] Interactive Technologies Group, Universitat Pompeu Fabra, Barcelona, Spain
{gerard.llorach, josep.blat}@upf.edu
[2] Medizinische Physik and Cluster of Excellence Hearing4all', Universität Oldenburg
[3] Hörzentrum Oldenburg GmbH, Oldenburg, Germany

Abstract. The creation and support of Embodied Conversational Agents (ECAs) has been quite challenging, as features required might not be straight-forward to implement and to integrate in a single application. Furthermore, ECAs as desktop applications present drawbacks for both developers and users; the former have to develop for each device and operating system and the latter must install additional software, limiting their widespread use. In this paper we demonstrate how recent advances in web technologies show promising steps towards capable web-based ECAs, through some off-the-shelf technologies, in particular, the Web Speech API, Web Audio API, WebGL and Web Workers. We describe their integration into a simple fully functional web-based 3D ECA accessible from any modern device, with special attention to our novel work in the creation and support of the embodiment aspects.

Keywords: embodied conversational agents, web technologies, virtual characters

Fig. 1. Interface of the web application

J. Beskow et al. (Eds.): IVA 2017, LNAI 10498, pp. 255-258, 2017.
DOI 10.1007/978-3-319-67401-8_34

1 Introduction

We present an implementation where our main contribution is the support of the 3D embodiment and the integration of web technologies. We demonstrate that a 3D ECA in the web browser is feasible using the right tools and libraries, with the work of one or two experts during two weeks. In addition, our system is open and standars based, so that better/alternative artificial intelligence or other components, can be easily connected. Indeed, our web-based ECA uses the following web/open source components:

1. **Listening and Speaking**: speech recognition and synthesis with the Web Speech API.
2. **Understanding, thinking and replying**: ELIZA [1] as artificial conversational entity; rule-based nonverbal behaviors described in [2] for gaze and head motions.
3. **Embodiment**: creation of virtual characters with Makehuman and Blender; support, real-time rendering and animations with WebGLStudio [3]; facial tracking with javascript libraries (jsfeat) [4] within a Web Worker allowing the agent to follow the user with the gaze.

Currently there are no WebGL-based ECAs with 3D virtual characters with the advanced features we present to the best of our knowledge.

2 Related Work

RAG LiteBody [8] was one of the first web implementations but the ECAs only allowed the user to choose from a set of sentences as input and the embodiment was 2D and based on Adobe Flash. A first WebGL implementation of a talking head can be seen in [12], but the facial features need to be processed by the server and it is not interactive. The company Existor [9] with products such as Cleverbot, Cleverscript and Evie, is one of the few that supports web-based ECAs, although they are more specialized on the creation of the chatbot system and use off-the-shelf web components for speech processing. They embody the agent through 2D video-realistic facial expressions synthesis by means of morphing with Adobe Flash plugins. Our approach does not need any plugins, it is based on open source components and allows the users and researchers to modify any components and to create their own virtual characters.

3 Components

Speech Recognition, Speech Synthesis, Dialogue System The Web Speech API permits to use local OS services and external services by URI. In the Chrome browser, Google services are the default configuration, and provide real-time, incremental speech recognition in several languages and dialects with high accuracy as well as speech synthesis through a few lines of code.

For demonstration purposes we used one of the first chatbots: ELIZA from the mid 60s. We integrated the chatbot as a script on the web client [13], adding some nonverbal behavior such as shaking the head when negation words (no, not, n't) are spoken by the agent and head nods while listening and speaking [2]. It is important to note that more sophisticated dialogue systems can be as easily connected, and provide a more conversationally capable ECAs.

Embodiment Among the tools to create 3D humanoid characters, Makehuman, Autodesk Character Generator, Poser, Mixamo (Fuse) and Daz Studio, we used Makehuman to create virtual characters and Blender to add the blend shapes and optimize the models.

Applications such as Unity3D and Unreal Engine permit to rapidly create, visualize and export 3D scenes and games with little coding and compatibility problems for different OSs. However, web plugins are becoming unsupported by some browsers, in particular the Unity Web Player by Chrome. We chose WebGLStudio, a 3D scene editor and game engine, as it has better tools and components to easily to integrate virtual characters than other web engines such as PlayCanvas [10] and Clara.io [11] and it is supported in Chrome.

Our implementation supports some basic BML commands such as gaze and head nods in WebGLStudio; uses a web-based system which automatically generates facial expressions based on the two values of valence and arousal as proposed in [12] and a web-based lip-syncing implementation proposed in [13]. Only eight blend shapes are needed for facial expressions, where three of them are used for the lip-sync, which is quite cost-effective when creating new characters.

We used a facial tracking library [4] and implemented it in Web Workers to extract the position of the user's face relative to the camera so that the ECA directs its gaze at the user. Facial expression analysis libraries and algorithms could also be implemented using such Web Workers, so that the nonverbal behavior of the user could be extracted and used in the dialogue.

4 Results, Discussion and Future Work

A novel web-based ECA was implemented successfully, fulfilling all basic requirements. The timing performance of each component was tested over 100 interactions on a PC (Windows 8 x64 2.50GHz, NVIDIA GeForce GT 750M) with Google Chrome v56.0.2924.87 and a internet connection of 45Mbps with the following results: SST $mean = 166.67ms$ ($sd = 86.47$); TTS $mean = 233.01ms$ ($sd = 251.39$); ElizaAI $mean = 3.72ms$ ($sd = 1.88$); Total processing time $mean = 392.66ms$ ($sd = 254.09$). Thus, the processing time of an interaction with the system (STT, TTS and AI) is less than a second, an acceptable pause in natural human conversations [14]. The system was developed in two weeks of work, using WebGL, BML and facial tracking libraries.

Open web-based ECAs such as ours will allow researchers to carry out large user studies more easily and possibly in a more standardized way, a contribution to advance in the ECAs research field.

Acknowledgements

This research has been partially funded by the Spanish Ministry of Economy and Competitiveness (RESET TIN2014-53199-C3-3-R), by the DFG research grant FOR1732 and by the European Commission under the contract number H2020-645012-RIA (KRISTINA) and under the the Marie Sklodowska-Curie grant agreement No 675324 (ENRICH). Special thanks to Volker Hohmann and Sergio Sagayo for revisions and counseling and to Javi Agenjo for developing WebGLStudio and helping out with all the technical challenges.

References

1. Weinzembaum, J.: ELIZA A Computer Program for the Study of Natural Language Communication Between Man And Machine. In: Communications of the ACM, 9 (1), pp. 36–45 (1966)
2. Ruhland K., Peters C. E., Andrist S., Badler J. B., Badler N. I., Gleicher M., Mutlu B., McDonnell R.: A Review of Eye Gaze in Virtual Agents, Social Robotics and HCI: Behaviour Generation, User Interaction and Perception. In: Computer Graphics Forum 34, 6, 299–326. (2015)
3. Agenjo, J., Evans, A., Blat, J.: WebGLStudio: a pipeline for WebGL scene creation. In: Proceedings of the 18th International Conference on 3D Web Technology, 79–82. (2013)
4. Zatepyakin E.: JavaScript Computer Vision library (jsfeat), `https://github.com/inspirit/jsfeat`
5. Romeo, M.: Automated Processes and Intelligent Tools in CG Media Production. PhD thesis, 119-148 (2016)
6. Llorach, G., A. Evans, J. Blat, G. Grimm, V. Hohmann. Web-based live speech-driven lip-sync. In: 8th International Conference on Games and Virtual Worlds for Serious Applications (VS-Games) (2016)
7. Kopp S., Krenn B., Marsella S., Marshall A.N., Pelachaud C., Pirker H., Thrisson K. R., Vilhjlmsson H.: Towards a Common Framework for Multimodal Generation: the Behavior Markup Language. In: Proceedings of the 6th international Conference on Intelligent Virtual Agents (IVA'06), Gratch J., Young M., Aylett R., Ballin D., Olivier P. (Eds.). Springer-Verlag, Berlin, Heidelberg, 205–217, (2006)
8. Bickmore T., Schulman D., Shaw G.: DTask and LiteBody: Open Source, Standards-Based Tools for Building Web-Deployed Embodied Conversational Agents. In: Ruttkay Z., Kipp M., Nijholt A., Vilhjlmsson H.H. (eds) Intelligent Virtual Agents. IVA 2009. Lecture Notes in Computer Science, vol 5773. Springer, Berlin, Heidelberg (2009)
9. Existor, `http://www.existor.com/`
10. PlayCanvas, `http://playcanvas.com/`
11. Clara.io, `https://clara.io/`
12. Leone G. R., Cosi P.: LUCIA-webGL: a web based Italian MPEG-4 talking head, In AVSP-2011, 123-126 (2011).
13. Landsteiner, N., `http://www.masswerk.at/elizabot/`, (2005)
14. Sacks, H., Schegloff, E., Jefferson, G.. A Simplest Systematics for the Organization of Turn-Taking for Conversation. In Language, 50(4), 696–735 (1974)