Time to go ONLINE! A Modular Framework for Building Internet-based Socially Interactive Agents

Mihai Polceanu*
polceanu@enib.fr
ENIB
Brest, France

Christine Lisetti
lisetti@cis.fiu.edu
Florida International University
Miami, USA

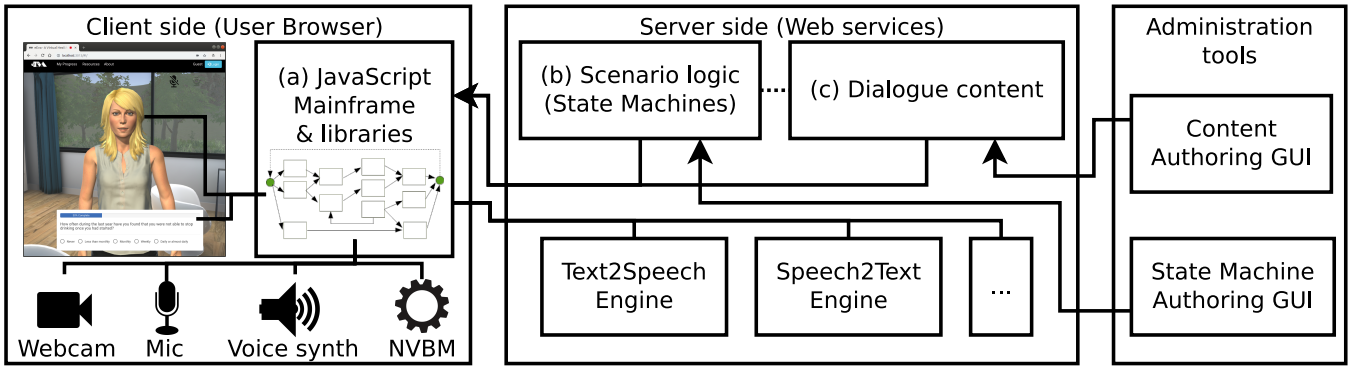

Figure 1: System diagram illustrating main components: the modular JavaScript mainframe (a) which controls the client side application by interconnecting the character input, output and non-verbal behavior model (NVBM) with the scenario logic (b) and dialogue content (c).

ABSTRACT

Although socially interactive agents have emerged as a new metaphor for human-computer interaction, they are, to date, absent from the Internet. We describe the design choices, implementation, and challenges in building EEVA, the first fully integrated platform-independent framework for deploying realistic 3D web-based social agents: with real-time multimodal perception of, and response to, the user's verbal and non-verbal social cues, EEVA agents are capable of communicating rich customizable content to users in real time, while building and maintaining users' profiles for long-term interactions. The modularity of the EEVA framework enables it to be used as a testbed for agents' social communication model development of increasing performance and sophistication (e.g. building rapport, expressing empathy). We discuss a case study in which we show how we used the EEVA framework to create dialog content for a health agent to deliver an online tailored behavior change health intervention, and we show its feasibility by analyzing the response time of the system over the Internet.

CCS CONCEPTS

• **Computing methodologies** → **Intelligent agents**; • **Computer systems organization** → **Real-time system architecture**; • **Applied computing** → *Health care information systems.*

KEYWORDS

web-based 3D character; multimodal interaction; real-time virtual counseling

ACM Reference Format:
Mihai Polceanu and Christine Lisetti. 2019. Time to go ONLINE! A Modular Framework for Building Internet-based Socially Interactive Agents. In *ACM International Conference on Intelligent Virtual Agents (IVA '19), July 2–5, 2019, PARIS, France.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3308532.3329452

1 MOTIVATION

As human-computer interaction (HCI) has become increasingly present in daily life contexts involving socio-emotional content (e.g. medicine, education, entertainment), socially interactive virtual agents – also known as Embodied Conversational Agents (ECA) or as Intelligent Virtual Agents (IVA) – have emerged over the past decade as a new metaphor for HCI to address users' need for natural interfaces simulating human-human conversations.

Building an IVA, however, is no easy feat and presents many interdisciplinary challenges. Whereas having socially appropriate interactions can be challenging even for humans at times, generating artificial social behaviors requires a mix of technology, psychology and art. Indeed, social appropriateness during dialogues, requires (apart from choosing an appropriate topic) knowing how to use

*Current affiliation: University of Greenwich, UK.

IVA '19, July 2–5, 2019, PARIS, France

ACM ISBN 978-1-4503-6672-4/19/07.
https://doi.org/10.1145/3308532.3329452

different channels of communication to establish and maintain rapport via verbal and non-verbal cues [6, 23], such as respectful eye contact [7], motor mimicry and synchoronous postures [3], expression of facial and other nonverbal social cues that are congruent with verbal utterances and emotional states, among others.

In spite of such complexity, IVA researchers have leveraged latest progress in affective computing [2, 4, 24] to build agents with subtle social cues and responses [1, 7, 14, 16, 16, 20, 25]. IVAs are becoming able to establish some rapport [8], express (some) empathy [9, 11, 15].

In spite of their success, however, IVA development did not scale with the now ubiquitous connected devices and latest progress on 3D graphics that can be rendered on internet browsers.

Whereas a few attempts have been made to build web-based 3D ECAs [12, 18, 22], their implementation is still very rudimentary, and none provide an integrated framework for web-based IVA development, including social cues modeling and dialog generation.

In health care, where human personnel are vastly outnumbered by people who need aid, virtual health agents (also referred to as virtual health coaches) capable of screening or providing empathic support to individuals, anytime anywhere, about their lifestyles (e.g. alcohol, drug or nicotine consumption, exercise or lack of, eating habits) have not only been found promising by healthcare research, but also better accepted by users than text-only computer-based interventions [11]. Other health-related agents have been emerging [5, 10, 19, 21], but their lack of availability on the web diminishes their potential impact.

In order to be effective, a health agent needs to be easily accessible (via common communication devices, at any time), usable (have an easy to use interface), enjoyable (provide a positive user experience), responsive to user's emotional behaviors (establish and maintain rapport) and scalable (accommodate an increasing number of users without computational overhead).

2 CASE STUDY: SOCIALLY INTERACTIVE HEALTH AGENT

The EEVA framework (Figure 1) is a cross-platform system (Fig. 2) that can be used to develop a web-based ECA capable of delivering behavior change health interventions. The current case study consists in delivering a brief motivational interviewing (BMI) intervention for at-risk behaviors such as alcohol consumption, over eating, smoking, lack of exercise.

As detailed in [13], the content of any BMI is clearly structured into a sequence of four steps, in addition to an initial greeting and a closing statement (with potential referral of resources for healthy lifestyles) [13]: 1) screening the person's lifestyle with a series of questionnaires; 2) providing normative feedback about the person's lifestyle; 3) if the person is found to have lifestyle patterns placing them at risk (as determined in steps 1-2), assessing what level or readiness to change the at-risk behavior(s) the person is experiencing (from not at all, to unsure, to ready); and 4) collaborating with the person to create a behavior change plan that is aligned with the level of readiness determined in step 3.

Each step has a number of questions that prompt the user to input one answer, multiple choices, or typed or spoken natural language. The system output consists in the feedback given by the

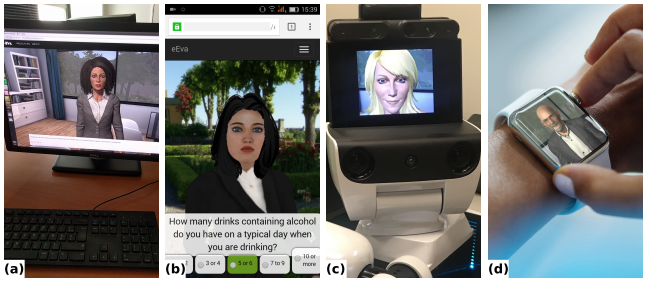

Figure 2: EEVA running on different platforms: (a) desktop, (b) mobile phone, (c) autonomous robot [17], (d) smartwatch concept.

virtual character, along with visual content such as text, images, videos and HTML. The framework provides a set of authoring tools created to facilitate the creation of content for diverse use cases. The interface allows the content creator to input multimedia (visual and audio content) relevant to a particular scenario that is to be executed by the virtual character.

The EEVA framework also enables the creation and retrieval of user models that can be used to tailor and personalize the interaction with the ECA. The user's answers to the various questionnaires are saved to provide personalized feedback – in our case, normative feedback about their [un/]healthy life-style. The user model consists in storing the user's answers to the agents' utterances, along with a set of scores that are calculated based on these results. These scores allow scenario branching, using conditional guards.

2.1 Runtime Evaluation over the Web

To evaluate our framework, we deployed it on a custom built server running Ubuntu 14.04 LTS with a Quad-core 2GHz CPU and 4GB RAM and measured the response time of critical system components on connections between two continents (see Table 1). The server was configured with SSL/TSL encryption, in order to satisfy the WebRTC security standards for accessing the user's camera and microphone. This also permitted a smoother user experience, as during non-encrypted connections the user consent must be asked for each input device reactivation (standard security policy).

We tested two main types of network connections available to the public: broadband and 4G mobile data. All experiments were performed under "first run" conditions - *i.e.* no cache mechanism was used, to simulate the first time a new user would connect to the system. The majority of the launch time consists in loading the 3D visual data (the 3D character and surrounding scene), taking on average 30 and 25 seconds on 4G and Broadband respectively, while the lightweight version amounts to 16 and 10 seconds respectively. Each component is asynchronously loaded, thereby limiting the additional overhead.

While using caching techniques removes this significant load time for returning users, real time interaction is dependent on the response time of speech synthesis and recognition (among other factors). The experiment results show that our framework can maintain realtime interaction on both connection types, amounting to an average of 1 second to produce each spoken sentence and fast

continuous speech recognition using the functionality built into the Chrome browser.

Table 1: Average response time and standard deviation analysis for EEVA (in milliseconds); 4G/Broadband connections over the Internet between North America and Europe; caching disabled (first run).

Functionality	4G mobile data	Broadband internet
Unity 3D character	30018 +/- 663	24626 +/- 1910
Light 3D character	16612 +/- 1471	10934 +/- 2008
TTS (sentence)	939 +/- 381	551 +/- 141
TTS (word)	72 +/- 40	44 +/- 23
Speech recognition	~30 (Offline processing)	
Entire HTTP request	1124 +/- 166	784 +/- 66
DOM loading	2313 +/- 80	1635 +/- 224

The experiments (Table 1) show that the main distributed functionalities of the EEVA framework do allow real-time interaction and acceptable loading times even for the first run. Scaling up to a large number of users does not imply a linear increase of server capacity, as after the download, most critical components run on the client-side, while only some lightweight communication with the server is performed.

3 FUTURE WORK ADDING EMPATHIC CUES

Future work will involve carrying out experiments and evaluations of nonverbal models of behavior by end-users of the health agent system, in terms of the realism of the IVA behaviors, as well as the end-users' perceived sense of rapport with the IVA delivering the health intervention.

We will also investigate data-driven approaches to modelling the character's behavior and deploying them in directly into the client-side environment, to be integrated with the EEVA mainframe. This brings a significant advantage: the end-users' facial images would not need to exit the user's personal device for the system to function, thereby removing any potential privacy concerns about sharing identifiable facial images over the network.

ACKNOWLEDGMENTS

This work has been funded by the National Science Foundation (NSF) award #1423260. Thanks to G. Ruiz, D. Rivero, S. Lunn, E. Henley, S. Bolivar for their contribution, and to anonymous reviewers for valuable feedback.

REFERENCES

[1] R. Amini, C. Lisetti, and G. Ruiz. 2015. HapFACS 3.0: FACS-based facial expression generator for 3D speaking virtual characters. *IEEE Transactions on Affective Computing* 6, 4 (2015). https://doi.org/10.1109/TAFFC.2015.2432794

[2] Maryam Ashoori, Chunyan Miao, Majid Nili, and Mehdi Amoui. 2008. Economically inspired self-healing model for Multi-Agent Systems. In *Proceedings of the IEEE/WIC/ACM International Conference on Intelligent Agent Technology, IAT 2007*, Vol. 2. 261–264. https://doi.org/10.1109/IAT.2007.80

[3] Janet B. Bavelas, Alex Black, Charles R. Lemery, and Jennifer Mullett. 1986. "I show how you feel": Motor mimicry as a communicative act. *Journal of Personality and Social Psychology* 50, 2 (1986), 322–329. https://doi.org/10.1037/0022-3514.50.2.322

[4] Rafael A Calvo and Sidney D'Mello. 2010. Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing* 1, 1 (2010), 18–37. https://doi.org/10.1109/T-AFFC.2010.1

[5] Fiorella de Rosis, Nicole Novielli, Valeria Carofiglio, Addolorata Cavalluzzi, and Berardina De Carolis. 2006. User modeling and adaptation in health promotion dialogs with an animated character. *Journal of Biomedical Informatics* 39, 5 (2006), 514–531. https://doi.org/10.1016/j.jbi.2006.01.001

[6] JE Grahe. 1999. The importance of nonverbal cues in judging rapport. *Journal of Nonverbal behavior* 23, 4 (1999), 253–269. http://www.springerlink.com/index/V8U30855W38M4673.pdf

[7] Ouriel Grynszpan, Jean Claude Martin, and Philippe Fossati. 2017. Gaze leading is associated with liking. *Acta Psychologica* 173 (2017), 66–72. https://doi.org/10.1016/j.actpsy.2016.12.006

[8] Lixing Huang, Louis-philippe Morency, and Jonathan Gratch. 2011. Virtual Rapport 2.0. In *International Conference on Intelligent Virtual Agents, Intelligence, Lecture Notes in Artificial Intelligence*. Springer-Verlag Berlin Heidelberg, 68–79.

[9] Joris H. Janssen. 2012. A three-component framework for empathic technologies to augment human interaction. *Journal on Multimodal User Interfaces* 6, 3-4 (2012), 143–161. https://doi.org/10.1007/s12193-012-0097-5

[10] C. LeRouge, K. Dickhut, C. Lisetti, S. Sangameswaran, and T. Malasanos. 2016. Engaging adolescents in a computer-based weight management program: Avatars and virtual coaches could help. *Journal of the American Medical Informatics Association* 23, 1 (2016). https://doi.org/10.1093/jamia/ocv078

[11] Christine Lisetti, Reza Amini, Ugan Yasavur, and Naphtali Rishe. 2013. I can help you change! an empathic virtual agent delivers behavior change health interventions. *ACM Transactions on Management Information Systems (TMIS)* 4, 4 (2013), 19.

[12] Gerard Llorach and Josep Blat. 2017. Say Hi to Eliza. In *International Conference on Intelligent Virtual Agents*. 255–258. https://doi.org/10.1007/978-3-319-67401-8_34

[13] William R Miller, R Gayle Sovereign, and Barbara Krege. 1988. Motivational interviewing with problem drinkers: II. The Drinker's Check-up as a preventive intervention. *Behavioural and Cognitive Psychotherapy* 16, 4 (1988), 251–268.

[14] Magalie Ochs, Catherine Pelachaud, and Gary McKeown. 2017. A User-Perception Based Approach to Create Smiling Embodied Conversational Agents. *ACM Transactions on Interactive Intelligent Systems* 7, 1 (2017), 1–33. https://doi.org/10.1145/2925993

[15] Ana Paiva, Iolanda Leite, Hana Boukricha, and Ipke Wachsmuth. 2017. Empathy in Virtual Agents and Robots. *ACM Transactions on Interactive Intelligent Systems* 7, 3 (2017). https://doi.org/10.1145/2912150

[16] Catherine Pelachaud. 2009. Modelling multimodal expression of emotion in a virtual agent. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 364, 1535 (dec 2009), 3539–48. https://doi.org/10.1098/rstb.2009.0186

[17] Pedro Pena, Christine Lisetti, Mihai Polceanu, and Ubbo Visser. 2018. eEVA: Real-time Web-based Affective Agents for Human-Robot Interface. In *RoboCup 2018: Robot World Cup XXII*. Montreal, Canada, Springer International Publishing.

[18] Vikram Ramanarayanan, David Pautler, Patrick Lange, and David Suendermann-Oeft. 2018. *Interview with an avatar: A real-time cloud-based virtual dialog agent for educational and job training applications.* Technical Report Research Memorandum No. RM-18-02. Princeton, NJ: Educational Testing Service. 1–8 pages.

[19] Albert Rizzo, Russell Shilling, Eric Forbell, Stefan Scherer, Jonathan Gratch, and Louis-Philippe Morency. 2016. Chapter 3 âĂŞ Autonomous Virtual Human Agents for Healthcare Information Support and Clinical Interviewing. *Artificial Intelligence in Behavioral and Mental Health Care* (2016), 53–79. https://doi.org/10.1016/B978-0-12-420248-1.00003-9

[20] K. Ruhland, C. E. Peters, S. Andrist, J. B. Badler, N. I. Badler, M. Gleicher, B. Mutlu, and R. McDonnell. 2015. A Review of Eye Gaze in Virtual Agents, Social Robotics and HCI: Behaviour Generation, User Interaction and Perception. *Computer Graphics Forum* 34, 6 (2015), 299–326. https://doi.org/10.1111/cgf.12603

[21] Mark R Scholten, Saskia M Kelders, and Julia EWC Van Gemert-Pijnen. 2017. Self-guided web-based interventions: Scoping review on user needs and the potential of embodied conversational agents to address them. *Journal of medical Internet research* 19, 11 (2017).

[22] Jessica Schroeder, Chelsey Wilks, Kael Rowan, Arturo Toledo, Ann Paradiso, Mary Czerwinski, Gloria Mark, and Marsha M. Linehan. 2018. Pocket Skills: A Conversational Mobile Web App To Support Dialectical Behavioral Therapy. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2018)* (2018), 1–15. https://doi.org/10.1145/3173574.3173972

[23] L. Tickle-Degnen and Robert Rosenthal. 1990. The nature of rapport and its nonverbal correlates. *Psychological Inquiry* 1, 4 (1990), 285–293. http://www.tandfonline.com/doi/abs/10.1207/s15327965pli0104{_}1

[24] Johannes Wagner, Florian Lingenfelser, Tobias Baur, Ionut Damian, Felix Kistler, and Elisabeth André. 2013. The Social Signal Interpretation (SSI) Framework Multimodal Signal Processing and Recognition in Real-Time. In *Proceedings of the ACM Multimedia Conference*. 1–4.

[25] U. Yasavur, C. Lisetti, and N. Rishe. 2014. Let's talk! speaking virtual counselor offers you a brief intervention. *Journal on Multimodal User Interfaces* 8, 4 (2014).