CasandRA: A Screenplay Approach to Dictate the Behavior of Virtual Humans in AmI Environments

Evropi Stefanidi, Asterios Leonidis(✉), Nikolaos Partarakis, and Margherita Antona

Institute of Computer Science (ICS), Foundation for Research and Technology – Hellas (FORTH), Heraklion, Cretet, Greece
{evropi, leonidis, partarak, antona}@ics.forth.gr

Abstract. Intelligent Conversational Agents are already employed in different scenarios, both in commerce and in research. In particular, they can play an important role in defining a new natural interaction paradigm between them and humans. When these Intelligent Agents take a human-like form (embodied Virtual Agents) in the virtual world, we refer to them as Virtual Humans. In this context, they can communicate with humans through storytelling, where the Virtual Human plays the role of a narrator and/or demonstrator, and the user can listen, as well as interact with the story. We propose that the behavior and actions of multiple, concurrently active Virtual Humans, can be the result of communication between them, based on a dynamic script, which resembles in structure a screenplay. This paper presents CasandRA, a framework enabling real-time user interaction with Virtual Humans, whose actions are based on this kind of scripts. CasandRA can be integrated in any Ambient Intelligence setting, and the Virtual Humans provide contextual information, assistance, and narration, accessible through various mobile devices, in Augmented Reality. Finally, they allow users to manipulate smart objects in AmI Environments.

Keywords: Intelligent conversational agent · Virtual humans · Storytelling · Augmented reality · Ambient intelligent environment

1 Introduction

Virtual Humans are embodied agents, existing in virtual environments, that look, act and interact with humans in a natural way. The incorporation of Virtual Humans (VHs) in Ambient Intelligence (AmI) environments can enhance the social aspects of interaction, offering natural anthropocentric communication [1]. In this context, Intelligent Conversational Agents (ICAs) enable the interaction of the VHs with humans, and can be combined with end-user development (EUD), to author and define the behavior of the agents, as well as the AmI environments [2]. In the context of EUD, providing users with intelligent tools that support authoring and creative processes is important [3], as user-generated content sharing has become a cultural phenomenon and interactive storytelling crafts are the focus of increasing interest [4].

C. Stephanidis and M. Antona (Eds.): HCII 2019, CCIS 1088, pp. 57–66, 2019.
https://doi.org/10.1007/978-3-030-30712-7_8

Storytelling and VHs, as well as game-like interfaces, have been introduced to replace or supplement GUIs [4]. Our focus, however, does not lie on how these stories are created, but on how to enact them, through the VHs; that is, on story telling rather than story creation. Storytelling traditionally relies on a predefined plot, concerning facts and occurrences, and involves presentation means such as speech, poses, and gestures of the narrator, in our case the VH, as well as representations of narration; that means textual as well as visual aids (e.g. pictures, videos, presentations, etc.) [4].

Storytelling is usually based on a script, the "screenplay", a term also used in filmmaking. In "The Screenwriter's Workbook" [5], screenwriter Syd Field defines a screenplay as a story told in words and pictures, so that in addition to reading the dialogue, the reader of a screenplay can read what the camera sees [6]. In our case, the readers are the VHs, who "read" the script, and act appropriately. We thus adapt the concept of screenplay to a conversation between VHs, who coordinate in order to enact a story. According to [7], a conversation is an interactive dialogue between two agents, which in our case are the VHs. In CasandRA, the conversation flow between the VHs is encoded in scripts, dictating their behavior (i.e. actions, movements, etc.).

In this paper we present CasandRA, a framework that enables real-time user interaction with ICAs, in the form of VHs, in Augmented Reality (AR), within AmI environments. These VHs can provide help and information, as well as act as story-tellers. Moreover, they allow users to use natural voice-based interaction to get information, as well as configure and manipulate various smart artifacts of the AmI environment. Our novelty lies in the communication protocol between the VHs, which dictates their behavior and intelligence. This protocol does not limit itself to speech, but also posture, movement, facial animation, etc., as well as the sharing of content with the users (images, video, presentation etc.). One VH (the narrator) is responsible for coordinating both itself and all the other VHs (demonstrators), who may be acting in the same or different devices. This protocol allows real-time interaction and is based on a dynamic script, which resembles a screenplay. The usage of this scripting technique facilitates the scalability and reusability of the script. Each script consists of sections which get selected for execution dynamically during run-time, depending on the interaction with the user.

2 Related Work

Several studies highlight the advantages of VHs, as they can elicit better results with regard to social presence, engagement and performance. In [8, 9] users favored interacting with an agent capable of natural conversational behaviors (e.g., gesture, gaze, turn-taking) rather than an interaction without these features. Moreover, research in [10] demonstrated that an embodied agent with locomotion and gestures can positively affect users' sense of engagement, social richness, and social presence. Finally, with respect to engagement, participants in [11] could better recall stories of robotic agents when the agent looked at them more during a storytelling scenario.

Regarding contextual awareness, and the capability of VHs to interact with their surroundings, [12] discusses the perception of changes to the environment as well as the ability to influence it with a VH, and concludes that this approach can increase

social presence and lead to realistic behavior. With respect to environmental awareness, the research in [13–15] indicated that a VH in AR exhibiting awareness of objects in a physical room elicited higher social presence ratings.

VHs are investigated in various research projects, with different systems offering conversational abilities, user training, adaptive behavior and VH creation. For example, the ICT Virtual Human Toolkit [16, 17] offers a flexible framework for generating high fidelity embodied agents and integrating them in virtual environments. Embodied Conversational Agents as an alternative form of intelligent user interface are discussed in depth in [18]. Finally, in [19] Maxine is described, an animation engine that permits its users to author scenes and VHs, focusing on multimodal and emotional interaction.

In the same context, VHs have been proven effective as museum storytellers, due to their inherent ability to simulate verbal as well as nonverbal communicative behavior. This type of interface is made possible with the help of multimodal dialogue systems, which extend common speech dialogue systems with additional modalities just like in human-human interaction [20]. However, employing VHs as personal and believable dialogue partners in multimodal dialogs entails several challenges, because this requires not only a reliable and consistent motion and dialogue behavior, but also appropriate nonverbal communication and affective behavior. Over the last decade, there has been a considerable amount of success in creating interactive, conversational, virtual agents, including Ada and Grace, a pair of virtual Museum guides at the Boston Museum of Science [20], the INOTS and ELITE training systems at the Naval Station in Newport and Fort Benning [21], and the SimSensei system designed for healthcare support [22]. In the FearNot! application VHs have also been applied to facilitating bullying prevention education [23].

Existing approaches have been proven successful but target specific application, communication and information provision contexts. However, in order to unleash the power of Virtual Humans as conversational agents in smart environments, there are still several open challenges imposed by the radically changing ICT domain. Such challenges are mainly stemming from the need to address AmI ecosystems that have dynamic behavior and may offer unstructured and even chaotic interactions with unpredicted user groups in fluid contexts, changing through the dynamic addition and modification of smart devices and services. Current approaches do not provide a holistic method suitable for the broad needs imposed by AmI environments. A step towards this direction is the work in [1], where Bryan is presented, a virtual character for information provision who supports alternative roles and can be integrated in AmI environments.

Regarding the conversation between VHs, [7] presents a system for automatically animating conversations between multiple human-like agents. They focus on how to synchronize speech, facial expression, and gesture, so as to create an interactive animation dialogue. In [24] a new language and protocol for agent communication is described, called Knowledge Query and Manipulation Language, focusing on the dialogue between the agents.

Our approach goes a step further, by delivering CasandRA, a platform that allows multiple VHs to interact with humans in AR, in an AmI environment, by following a straightforward, screenplay-like dynamic script. The behavior of the VHs is dictated by these scripts, allowing them to appear across different mobile devices as well.

Furthermore, our approach enhances the storytelling aspect, as it is performed by multiple VHs collaborating within the AmI environment, to offer a richer and more natural storytelling experience, inspired by the structure of screenwriting scripts in the film and theater industries. CasandRA's infrastructure is implemented in a way that allows scalability, reusability and easy integration of new scripts defining the VHs' behavior and storytelling. Finally, CasandRA allows users to get information about their surroundings in real-time and manipulate smart objects both in the virtual and real world.

3 Requirements

The high-level requirements that CasandRA satisfies have been solidified through an extensive literature review and an iterative requirements elicitation process, based on multiple collection methods, as outlined below:

1. Brainstorming, where mixed groups of developers were involved (AmI usability experts and end users)
2. Focus groups with end users
3. End-users who were requested to perform typical everyday activities, in the context of the "Intelligent Home" simulation space located at the AmI Facility (http://ami.ics.forth.gr/) of FORTH-ICS, in order to evaluate how VHs could be of assistance
4. Scenario building during co-design processes, where experts and end-users were formulating scenarios of use together

The following requirements were derived for CasandRA:

R1. **Human-likeness of the Virtual Humans**: The system should allow for natural, human-like interaction between the VHs and people. This means the VHs should be as realistic as possible, as well as user-friendly.

R2. **Context-awareness**: The VHs should be aware of the context, meaning they should have behaviors corresponding to their context of use; for example, when they are deployed in a smart museum exhibition, they should be aware of the existing artifacts and any relevant stories about them.

R3. **Smart object discovery and manipulation**: The VHs should be able to discover the smart objects in an AmI environment, what can be done with them, and be able to manipulate and configure them.

R4a. **Dynamic dialogue between the VHs**: The conversation between the VHs should be provided through a dynamic script which will dictate their behavior. Dynamic means that different sections of the script are selected to be executed at runtime, depending on the interaction with the user and the dialogue flow that occurs.

R4b. **Hierarchy in the Conversation between the VHs**: The conversation between the VHs should be structured, i.e. follow a hierarchy. There should be:

- One master, the moderator of the conversation (or narrator), who gives "commands" through the dialogue.
- One or more slaves (or demonstrators). They are the ones receiving the commands from the narrator.

R5. **Scalability, reusability, extensibility**: The system's architecture should support: (a) scalability for addition of more complex interaction scenarios, (b) reusability in different applications and AmI contexts, (c) extensibility to support future addition of other AmI services and functions.

4 System Description

CasandRA is a framework enabling VHs to interact with users in AmI environments of various contexts (e.g. Intelligent Homes, Museums), and provide information, smart object manipulation, and storytelling services. This interaction takes place in AR, i.e. users can utilize their mobile devices (smart phones, tablets), to view their surroundings in AR, enhanced with the VHs and the functionalities they provide.

CasandRA is currently deployed in the "Intelligent Home" simulation space located at the AmI Facility (http://ami.ics.forth.gr/) of FORTH-ICS. Inside this environment, everyday user activities are enhanced with the use of innovative interaction techniques, artificial intelligence, ambient applications, sophisticated middleware, monitoring and decision-making mechanisms, and distributed micro-services. A complete list of the programmable hardware facilities of the "Intelligent Home", that currently include both commercial equipment and technologically augmented custom-made artifacts, can be found in [2].

4.1 Architecture

CasandRA's architecture, depicted in Fig. 1, consists of different components, coordinated by the *Agent Behavior Script Manager* (ABSM). The *Scripts* component refers to the dynamic scripts driving the behavior of the VHs and their interactions, which are structured as screenplays, i.e. a dialogue between them. Each Script is comprised of different "sections", which correspond to different interaction scenarios. The Scripts are dynamic, meaning that, depending on the context and the user input, which is processed through the ParlAmI framework [25], different sections of them are executed. ParlAmI namely receives and modifies the raw user input, using Chatscript[1], and is targeted to facilitating the definition of behaviors in AmI spaces.

The ABSM combines ParlAmI input with information from the AmI-Solertis [26] platform, which is used for service and object discovery. With the data that these two frameworks stream to the ABSM, it then instructs the Narrator Script regarding which sections of the Script should be executed at a given moment. The Narrator Script is responsible for communicating with the available Demonstrator(s), and instruct them which section of their Script they should execute; in more detail, each section of the Demonstrator Script corresponds to a specific "line" in a section for the Narrator Script. This means that the Narrator script includes a command (e.g. *Demonstrate Looming technique*), that constitutes a section for the Demonstrator, i.e. it is interpreted to one or more commands for him. This can be better understood by viewing the sample script

[1] Wilcox, B.: Chatscript. (2011) http://chatscript.sourceforge.net/.

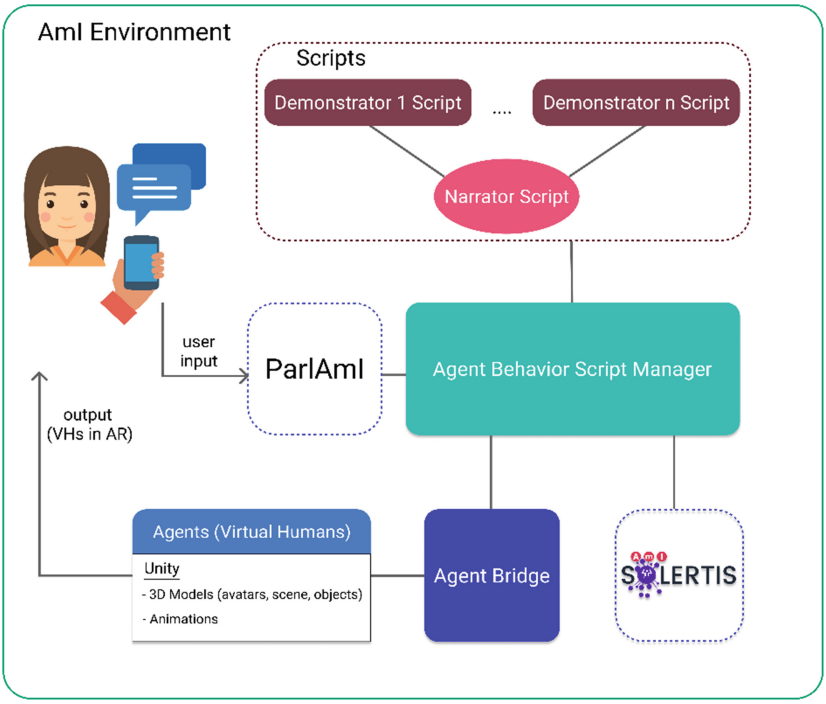

Fig. 1. CasandRA's architecture.

for the Narrator and Demonstrator Scripts in Fig. 2(a) and (b) respectively. We assume the context of an "Arts and Crafts" museum, where storytelling regarding the art of weaving with a loom is taking place. In this example, *Loom Storytelling* and *Behavior Definition* constitute different sections of the Narrator Script, while in the Demonstrator Script, *Demonstrate Looming Technique* and *Demonstrate Used Objects* are also different sections. For instance, when the *Demonstrate Looming Technique* section of the Demonstrator script gets executed, the Demonstrator should walk towards the Loom, sit in front of it, and begin to show how weaving with the Loom is conducted. After that, the Demonstrator should stand up. This is described by the commands visible in the Demonstrator Script below.

The ABSM allows a constant flow of information between all its components; thus, when the Narrator Script dictates that a VH should perform an action (e.g. say something, perform a certain physical movement), this information is passed on to the *Agent Bridge*. While the Narrator and Demonstrator Scripts constitute "high level" abstractions of the behavior of the VHs in natural language, the Agent Bridge (Fig. 2 (c)) converts them into separate "low-level" commands. These commands are then propagated to the Unity engine so as to execute the designated animations of the VHs in the virtual space. In reality, these functions constitute a remote API to an internal Unity module that implements them; for example, *PlayAnimation("StandUp")* is translated through that API to the corresponding Unity code, as depicted in Fig. 3.

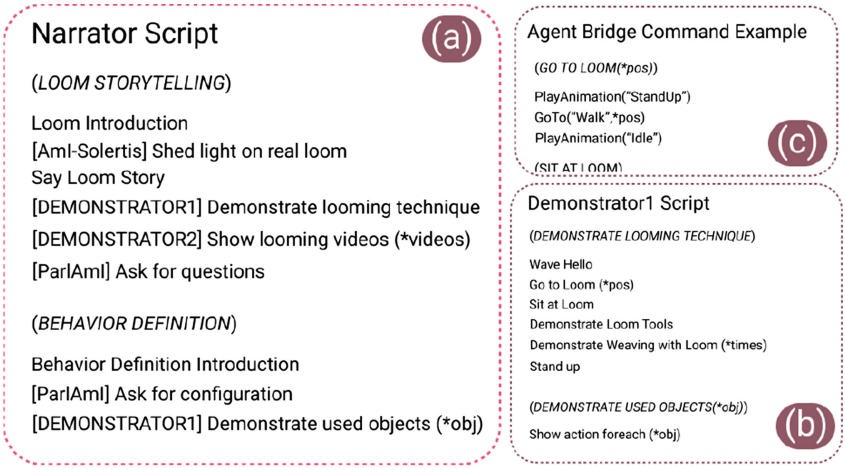

Fig. 2. (a) Example of narrator script. (b) Example of demonstrator script. (c) Example of agent bridge command

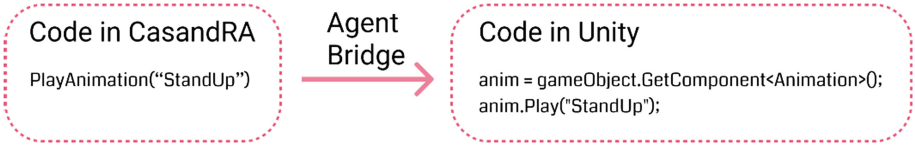

Fig. 3. The agent bridge translates CasandRA commands into Unity commands.

4.2 Use Cases of VHs in AmI Environments

The CasandRA framework can be integrated and used in any AmI environment consisting of smart devices and can provide Virtual Humans who can assist in various context scenarios. This section presents three such examples of VHs powered by CasandRA.

Assistance. The VHs can provide information and assistance to users, or even demonstrate how something works (e.g. tutorial of an Intelligent space or artifact). This kind of interaction is either explicitly initiated, after a user's request (e.g. "how can I use this switch", "what will the weather be like tomorrow", "is John home"), or implicitly, when the agent relies on contextual monitoring to detect when the user needs assistance. In the latter case, users can freely dismiss the agent if they do not want help. As an example of this type of interaction, when the user goes to bed, the agent can suggest turning off the heating or inform the user about any lights that remain switched on. Depending on the user's response, the agent learns to either turn them off automatically, or not bother the user in the future regarding this matter.

Behavior Definition. Besides providing information, the VHs can also be used as "virtual butlers", as a means of defining behavior scenarios in the context of the AmI

environment. For example, users can ask for the direct manipulation of a physical smart artifact (e.g. "turn the TV on channel 5"), or they can create rules, dictating the behavior of smart devices and their surrounding intelligent environment under certain conditions (e.g. "do not turn on the lights if I walk in the children's bedroom after 10 p.m.").

Storytelling. CasandRA can also provide immersive storytelling experiences, in various contexts. One VH plays the role of the storyteller (Narrator), and there can be one or more Demonstrators present, enhancing the story by making it come alive, with appropriate content sharing and demonstration. For example, in the context of a museum exhibiting arts and crafts, the Narrator explains the history and origins of a craft, e.g. on an interactive smart board, while a Demonstrator, visible on the personal mobile device of the museum visitor, can showcase how this craft was performed. Another Demonstrator can then appear on the mobile device, moving around the AR space and pointing out other relevant museum artifacts and exhibits. Storytelling is currently being enhanced based on the formulation of a protocol to transform Heritage Crafts to engaging cultural experiences. This protocol defines the process of capturing information from multiple sources including motion capturing, and representing such information appropriately, so as to generate narratives that can be then transformed to CasandRA scripts. This work is conducted under the European Union's Horizon 2020 research and innovation program under grant agreement No. 822336 (Mingei).

5 Plans of Evaluation

Regarding the evaluation of the storytelling functionality, this will be done in the near future in the context of the European Union's Horizon 2020 research and innovation program under grant agreement No. 822336 (Mingei). The main goals will be to validate: (a) the protocol that is used to generate narratives, (b) the ability of CasandRA to transform narratives to engaging stories in intelligent environments, and (c) the exploitation of CasandRA in the context of Heritage Crafts training, where the VH acts as a tutor for craftsmanship education, based on the computer-aided authoring of interactive experiences that involve manual procedures, use of simple machines, and tools in Augmented Reality.

6 Discussion

This paper has presented CasandRA, a framework enabling the interaction between humans and VHs in AmI environments in AR, for information and assistance provision, smart object manipulation, as well as storytelling. It utilizes a novel technique for the definition of the behavior of the VHs, by employing screenplay-like dynamic scripts for the definition of the behavior of the VHs, whose execution can be modified in real-time, according to the interaction with the user. In particular, there are two different types of scripts: (i) the Narrator's script, responsible for orchestrating the behavior of that VH (i.e. directly instructs him what to say and how to act), and dictating the behavior of any Demonstrators; and (ii) the Demonstrators' script, which

internally defines their behavior, and is externally exposed to the Narrator, so as to accommodate the overall scenario.

Our immediate plans include a user-based, full-scale evaluation of CasandRA's interface and functionalities, in order to assess its usability. Future improvements involve: (a) the ability to transform the narratives that will be provided by the Mingei project to CasandRA's scripts, for enriching the storytelling aspect of the system; (b) the deployment of the CasandRA framework on actual heritage sites; and (c) the introduction of an Avatar editing module, to enable users to select among available avatars for the VHs (Narrator/Demonstrators), allowing for further customization.

Acknowledgments. Part of the work reported in this paper is being conducted in the context of the European Union's Horizon 2020 research and innovation program under grant agreement No. 822336 (Mingei).

References

1. Birliraki, C., Grammenos, D., Stephanidis, C.: Employing virtual humans for interaction, assistance and information provision in ambient intelligence environments. In: Streitz, N., Markopoulos, P. (eds.) DAPI 2015. LNCS, vol. 9189, pp. 249–261. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-20804-6_23
2. Stefanidi, E., Foukarakis, M., Arampatzis, D., Korozi, M., Leonidis, A., Antona, M.: ParlAmI: a multimodal approach for programming intelligent environments. Technologies **7**, 11 (2019)
3. Riedl, M.O., Rowe, J.P., Elson, D.K.: Toward intelligent support of authoring machinima media content: story and visualization. In: Proceedings of the 2nd International Conference on INtelligent TEchnologies for Interactive enterTAINment, p. 4. ICST (Institute for Computer Sciences, Social-Informatics and … (2008)
4. Spierling, U., Grasbon, D., Braun, N., Iurgel, I.: Setting the scene: playing digital director in interactive storytelling and creation. Comput. Graph. **26**, 31–44 (2002)
5. Field, S., Field, S.: The Screenwriter's Workbook. Dell, New York (1984)
6. Honkanen, S.: Stepping Inside the Story: Writing Interactive Narratives for Virtual Reality (2018)
7. Cassell, J., et al.: Animated conversation: rule-based generation of facial expression, gesture & spoken intonation for multiple conversational agents. In: Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques, pp. 413–420. ACM (1994)
8. Cassell, J., Thorisson, K.R.: The power of a nod and a glance: envelope vs. emotional feedback in animated conversational agents. Appl. Artif. Intell. **13**, 519–538 (1999)
9. Cassell, J., Bickmore, T., Vilhjálmsson, H., Yan, H.: More than just a pretty face: affordances of embodiment. In: Proceedings of the 5th International Conference on Intelligent User Interfaces, pp. 52–59. ACM (2000)
10. Kim, K., Boelling, L., Haesler, S., Bailenson, J.N., Bruder, G., Welch, G.: Does a digital assistant need a body? The influence of visual embodiment and social behavior on the perception of intelligent virtual agents in AR. In: IEEE International Symposium on Mixed and Augmented Reality (2018)
11. Mutlu, B., Forlizzi, J., Hodgins, J.: A storytelling robot: modeling and evaluation of human-like gaze behavior. In: 2006 6th IEEE-RAS International Conference on Humanoid Robots, pp. 518–523. Citeseer (2006)

12. Chuah, J.H., et al.: Exploring agent physicality and social presence for medical team training. Presence Teleoperators Virtual Environ. **22**, 141–170 (2013)
13. Kim, K., Bruder, G., Welch, G.: Exploring the effects of observed physicality conflicts on real-virtual human interaction in augmented reality. In: Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology, p. 31. ACM (2017)
14. Kim, K., Maloney, D., Bruder, G., Bailenson, J.N., Welch, G.F.: The effects of virtual human's spatial and behavioral coherence with physical objects on social presence in AR. Comput. Animation Virtual Worlds **28**, e1771 (2017)
15. Kim, K., Schubert, R., Welch, G.: Exploring the impact of environmental effects on social presence with a virtual human. In: Traum, D., Swartout, W., Khooshabeh, P., Kopp, S., Scherer, S., Leuski, A. (eds.) IVA 2016. LNCS (LNAI), vol. 10011, pp. 470–474. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-47665-0_57
16. Hartholt, A., et al.: All together now. In: Aylett, R., Krenn, B., Pelachaud, C., Shimodaira, H. (eds.) IVA 2013. LNCS (LNAI), vol. 8108, pp. 368–381. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-40415-3_33
17. Virtual Human Toolkit. https://vhtoolkit.ict.usc.edu/
18. Cassell, J., Sullivan, J., Churchill, E., Prevost, S.: Embodied Conversational Agents. MIT Press, Cambridge (2000)
19. Baldassarri, S., Cerezo, E., Seron, F.J.: Maxine: a platform for embodied animated agents. Comput. Graph. **32**, 430–437 (2008). https://doi.org/10.1016/j.cag.2008.04.006
20. Swartout, W., et al.: Virtual museum guides demonstration. In: 2010 IEEE Spoken Language Technology Workshop, pp. 163–164. IEEE (2010)
21. Campbell, J.C., Hays, M.J., Core, M., Birch, M., Bosack, M., Clark, R.E.: Interpersonal and leadership skills: using virtual humans to teach new officers. In: Proceedings of Interservice/Industry Training, Simulation, and Education Conference, Paper (2011)
22. DeVault, D., et al.: SimSensei kiosk: a virtual human interviewer for healthcare decision support. In: Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems, pp. 1061–1068. International Foundation for Autonomous Agents and Multiagent Systems (2014)
23. Aylett, R., Vala, M., Sequeira, P., Paiva, A.: FearNot! – an emergent narrative approach to virtual dramas for anti-bullying education. In: Cavazza, M., Donikian, S. (eds.) ICVS 2007. LNCS, vol. 4871, pp. 202–205. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-77039-8_19
24. Finin, T., Fritzson, R., McKay, D., McEntire, R.: KQML as an agent communication language. In: Proceedings of the Third International Conference on Information and Knowledge Management, Gaithersburg, Maryland, USA, pp. 456–463. ACM (1994)
25. Stefanidi, E., Korozi, M., Leonidis, A., Antona, M.: Programming intelligent environments in natural language: an extensible interactive approach. In: Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference, pp. 50–57. ACM (2018)
26. Leonidis, A., Arampatzis, D., Louloudakis, N., Stephanidis, C.: The AmI-Solertis system: creating user experiences in smart environments. In: Proceedings of the 13th IEEE International Conference on Wireless and Mobile Computing, Networking and Communications (2017)