eEVA as a Real-Time Multimodal Agent Human-Robot Interface

P. Peña[1(✉)], M. Polceanu[3], C. Lisetti[2], and U. Visser[1]

[1] University of Miami, Coral Gables, FL 33146, USA
pedro@cs.miami.edu
[2] ENIB Brest, UMR 6285, Lab-STICC, 29200 Brest, France
[3] Florida International University, Miami, FL 33199, USA

Abstract. We posit that human-robot interfaces that integrate multimodal communication features of a 3-dimensional graphical social virtual agent with a high degree of freedom robot are highly promising. We discuss the modular agent architecture of an interactive system that integrates two frameworks (our in-house virtual social agent and robot agent framework) that enables social multimodal human-robot interaction with the Toyota's Human Support Robot (HSR). We demonstrate HSR greeting gestures using culturally diverse inspired motions, combined with our virtual social agent interface, and we provide the results of a pilot study designed to assess the effects of our multimodal virtual agent/robot system on users' experience. We discuss future directions for social interaction with a virtual agent/robot system.

Keywords: Human-robot interaction · Service and social robots · Intelligent virtual agents · Culturally-aware robotics · Culturally-aware virtual agents

1 Introduction

Twenty years ago, research has shown that humans respond positively to social cues when provided by computer artefacts [21]. With the emerging introduction of robots in social spaces where humans and robots co-exist, the design of socially competent robots could be pivotal for human acceptance of such robots. Humans are very skilled at innately reading non-verbal cues (*e.g.*, emotional signals) and extrapolating pertinent information from body language of other humans and animals [24]. Although some robots are currently capable to portray a small collection of emotional signals [12], robots social abilities are currently very limited. Recently, the use of virtual interactive social agents as main user interface (UI) has been shown to enhance users' experience during human-computer interactions in contexts involving social interactions (*e.g.*, health assistants, tutors, games) [7,14]. Yet robots intended to engage in social dialogs and physically collaborate with humans do not have virtual social agents as user interface.

We posit that human-robot interfaces that integrate multimodal communication features of a social virtual agent with a high degree of freedom robot might enhance users' experience with, and acceptance of, robots in their personal spaces, are highly promising, and need to be investigated. However, according to Matarić et al. [15], in order to avoid a mismatch between the expectations of the human and the behavior of the robot during human-robot interaction (HRI, henceforth), the natural integration of all the modules of the robot responsible for social, physical, and cognitive abilities is of utmost importance.

We have started to address this social HRI challenge by developing a multimodal human-robot interface for the Toyota's Human Support Robot (HSR, designed to help people in homes or offices) which integrates the RoboCanes agent and the Embodied Empathetic Virtual Agent (eEVA) developed by FIU's VISAGE lab. The RoboCanes agent is responsible for managing and controlling navigation, object manipulation, grasping, among other physical actions, while the VISAGE agent is responsible for recognizing and displaying social cues involving recognizing the user's facial expression and speech, synthesizing speech with lip-synchronization, and portraying appropriate facial expressions and gestures.

We created a greeting context for the pilot study of our first social human-HSR interactions with our RoboCanes-VISAGE interface (described in Sect. 4) by designing a small set of greeting gestures to personalize Toyota HSR with its users greeting preferences (and to establish some initial rapport in future more advanced studies): the Toyota HSR generates greeting gestures from four different cultures such as waving-hand (Western), fist-bump (informal Western), Shaka (Hawaii), and bowing (Japan) greeting gestures (for details see Sect. 4). The HSR's gesture greetings are performed based on the user's spoken selection of one of the four greetings and our pilot questionnaire aims to assess the impact of combining the virtual agent interface on the user's experience (*e.g.*, feelings of enjoyment, boredom, annoyance, user's perception of the robot's friendliness or of competence). Future directions for social interaction with a virtual agent/robot system are discussed in Sect. 5.

2   Related Work and Motivation

Human-Robot Interfaces: Human-robot interfaces that utilize multimodal features (*e.g.*, nonverbal and verbal channels) to communicate with humans has been a current trend in HRI [1,2,9,22], but has demonstrated to be very challenging due to the high-dimensional space of these channels. Therefore, theories and ideas from plethora of fields (*e.g.*, Neuroscience, psychology, and linguistics) have come together to develop new algorithms to create a more natural interface to communicate with humans. However due to hardware constraints and current A.I. technologies, developing an agent and robot that can communicate with humans at the level of human-human interaction has not been possible. Consequently, human-robot interfaces that are simple yet intuitive have been developed to help with tasks that require assistance for humans. An example of

these interfaces is the graphical user interface. Depending on the task, it is easier for the user to interact with a robot using a graphical user interface with 3D graphic rendering of the world to select objects or tasks for the robot to perform [4], than with speech recognition and synthesis as proposed with our approach.

Nagahama *et al.* [16] developed an interactive graphical interface for users that are not able to grab an object by themselves. The interface allows the user to specify the object the user wants the robot to fetch by clicking on the object on the screen. Hashimoto *et al.* [8] created a simple interface that has four different modes or windows to give Toyota HSR tasks or monitor the robot.

Nonverbal gestures (*e.g.*, arm gestures) to communicate with the robot and assist with tasks have also been used. Kofman *et al.* developed a human-robot interface that allows a user to teleoperate a robotic arm with vision [13]. There are also human-robot assistive interfaces developed with haptic and visual feedback [6,23]. Human-robot interfaces that are connected to the human brain have also been developed [20]. Qiu *et al.* developed a brain-machine interface that is able to control an exoskeleton robot through neural activity. There is also a recent trend of Augmented Reality (AR) human-robot interfaces to help users visualize an environment from another location in their physical environment [25].

Although there has been recurring research in human-robot interfaces, the communication between humans and robots through graphical interfaces is limited because the interaction between the human and the robot is constrained by the screen where the interface resides in, and it does not offer nonverbal and verbal communication as a medium of communication. Augmented and virtual reality is a promising interface but it is also limited by the hardware, equipment, and the lack of physical realism, *i.e.*, virtual characters cannot interact with the physical world. A promising yet an immature technology is the integration of virtual agents which offers the social realism that robots require and integration of robotics which offers the physical realism that virtual agents require.

Social Virtual Agents with Robots: Because virtual characters can use their sophisticated multimodal communication abilities (*e.g.* facial expressions, gaze, gesture) [17], to coach users in interactive stories [10], establish rapport (with back channeling cues such as head nods, smiles, shift of gaze or posture, or mimicry of head gestures) [18], communicate empathically [19], and engage in social talk [11], they have the potential of becoming as engaging as humans [7]. The integration of a virtual agent with social robots has been very limited and only given small attention. On example of a robot with a social virtual agent as a human-robot interface is GRACE (Graduate Robot Attending ConferencE) which was built by Simmons *et al.* [22] to compete in the AAAI Robot Challenge that required GRACE to socially interact with humans in a conference.

The Thinking Head research [9] was performed in conjunction with artist Stelarc where the facial characteristics of Stelarc were used for the animated head. Cavedon *et al.* developed an attention model for the Thinking Head that used backchanneling cues and eye gaze [5]. The Thinking Head resides in various robots such as a robot arm's end-effector and in a mobile robot.

Other human-robot interfaces include head-projection systems where a projector projects an animated face onto a mask [1, 2]. These systems allow an animated avatar to display complex facial expressions not yet possible with robotic hardware.

However, none of these previous approaches studied robots with manipulative capabilities that are able to produce gestures, appropriately combined with the social verbal and non-verbal cues of a virtual agent. Yet, many of the emerging and future human-robot interactions are or will require socially and culturally appropriate robots. Therefore rather than utilize a robot as a platform for a virtual character to enable movement in the physical world such as the literature discussed in this section, we developed an agent that takes advantage of the social-emotional capabilities of social virtual agents (*e.g.*, anthropomorphic agent, natural language, and nonverbal gestures) with the physical capabilities of the robot (high degree of freedom arm and mobile base of the HSR robot) that can work as a synchronized system which exhibits features from human-human interactions such as simple greetings (*e.g.*, robot greets user saying "hello" and waving arm based on the users' spoken utterance, discussed in Sect. 4) to enhance the social interaction with the user. In the following section, we will explain the architecture of the virtual agent and robot to understand how these two systems interact with each other while it is providing a synchronized interface for the user.

3 Modular Architecture for Real-Time Multimodal User-Interface Agents

3.1 RoboCanes-VISAGE: Integration of Two Agent-Based Frameworks

The system architecture of the RoboCanes-VISAGE affective robot agent consists of two separate frameworks: one developed by FIU's VISAGE lab (eEVA framework) and the other developed by UM's RoboCanes lab (RoboCanes framework). As described earlier, the RoboCanes agent is responsible for physical actions, such as managing and controlling navigation, object manipulation, grasping. The VISAGE agent is responsible for recognizing and displaying social cues involving recognition of the user's facial expression and speech, speech synthesis with lip-synchrony, and portray of appropriate facial expressions and gestures.

Since our goal is to integrate two existing agent-based systems (namely the eEVA and RoboCanes agents), in order for the integration of eEVA and RoboCanes modules to cooperate seamlessly, a higher-level framework has been designed and implemented to manage both systems accordingly. This was accomplished by integrating the inputs of eEVA and of the RoboCanes agent under one decision making process rather than treating both systems separately. By doing this, eEVA and RoboCanes agent act as one agent and their behavior is synchronized.

More specifically, in order to integrate both systems together, the frameworks communicate through the Standard ROS Javascript Library, roslibjs[1].

---

[1] http://wiki.ros.org/roslibjs.

This library facilitates both frameworks to communicate through web-sockets. Therefore the user input in eEVA is transported from these web-sockets to the RoboCanes framework, and the robot generates motions based on the requests from the user.

3.2 eEVA: A Framework for Building Empathic Embodied Virtual Agents

The default HSR user interface (UI) is shown in Fig. 1(a), and it is our aim to use our empathic embodied virtual agent (eEVA) shown in Fig. 1(b) to enhance user experience while interacting with HSR. While eEVA's UI is a 3D animated agent, it is driven by a fully integrated web-based multimodal modal system that perceives the user's facial expressions and verbal utterances in real time which controls the displays of socially appropriate facial expressions on its 3D-graphics characters, along with verbal utterances related to the context of the dialog-based interaction. eEVA's facial expressions are currently generated from the HapFACS[2] open source software developed by the VISAGE lab for the creation of physiologically realistic facial expressions on socially believable speaking virtual agents [3].
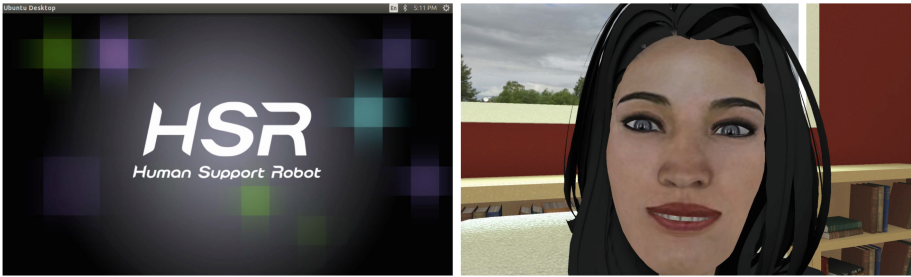
(a) HSR default visual screen (with eEVA's added speech recognition and synthesis) as user interface (b) eEVA visual screen (with eEVA's speech recognition and synthesis) as HSR user interface

Fig. 1. Human-robot interfaces

eEVA Components: The two basic components of the eEVA architecture consist of modules and resource generic types. The principle of a module is to robustly implement a single concrete functionality of the overall system. A module is defined by the task that it solves, the resources it requires for solving the given task, and the resources it provides (which may be further used for other purposes within the system). In other words, a module receives an input which is the resource it requires and it has an output which is the resource it provides. Modules are further categorized by their resource handling: sensors (*i.e.*, modules which only provide resources), processors (*i.e.*, with both required and provided

---

[2] http://ascl.cis.fiu.edu/hapfacs-open-source-softwareapi-download.html.

Table 1. List of eEVA current modules.

Ref. No.	Type	Short name	Function description
1	Sensor	ChromeSpeech	Speech recognition using Google Chrome API
2	Processor	HapCharacter	Virtual character controller (body and face)
3	Processor	UserChoice	User interface for interacting with eEVA
4	Processor	WinSAPISynth	Speech synthesis using Windows SAPI
5	Effector	WebGLScene	Default 3D scene rendering
6	Effector	ROSHandler	ROS Communication through roslibjs

resources), and effectors (*i.e.*, modules which require resources but produce no further data for system use). The list of eEVA modules and third-party libraries is shown in Table 1.

Sensors: Sensors are modules that provide an output but do not have a processed input. An example of a sensor in the eEVA framework is the ChromeSpeech module which uses Google Speech API to recognize speech from the user by using the head microphone of HSR as shown in Table 2. The final speech text from the user is processed by this module and provides a UserText and UserCommand resource that can then be required by another module such as an effector or processor. Hence, sensors are modules that receive input from the environment.

Processors: Processors are modules that require and provide resources. The modules process inputs from the sensors and then request the effectors to do an action. Hence, these modules extract information and make a decision. Since the interaction in the pilot study is turn-taking, the UserChoice module displays the choices the user can say (*i.e.*, the greetings discussed in Sect. 4). The virtual agent uses Windows SAPI to generate speech. It is important to note that majority of modules fall into the processor category and the collection of these modules define the behavior of the agent.

Effectors: The effectors are modules that require resources but do not further process other resources. Effectors are the modules that perform an action on the environment and are responsible for displaying system data such as the 3D virtual scene, the agent's behavior, text, and other information to the user. The effectors are the modules that are visible to the user and affect the perception of the sensors. The communication between eEVA and RoboCanes is done through an effector, ROSHandler. ROSHandler requires UserText resource from a sensor, ChromeSpeech module, and sends this resource through roslibjs (roslibjs deals with wrapping this resource in a format that ROS understands).

3.3 RoboCanes Components

On the robotic side, we use Toyota HSR which is an exemplary platform to embody the integration of the University of Miami (UM) RoboCanes agent with

the FIU VIrtual Social AGEnt (VISAGE). Our RoboCanes framework is an extension of the ROS[3] architecture that runs on the HSR.

The RoboCanes framework is developed in the ROS environment and it is also modular. In pursuance of gesture synthesis, the RoboCanes framework consists of a motion library node that uses MoveIt![4] and Toyota Motor Corporation (TMC) action servers. The relevant node for this research is the manipulation node.

Motion Planner: The motion planner node uses the MoveIt! library and the OMPL[5] library through MoveIt! to generate motions. The motions are requested by the eEVAHandler which handles the communication between both frameworks. The eEVAHandler processes the request from eEVA and decides which gesture to generate based on the input of eEVA. This results in the robot generating motions of the physical robot through ROS. In Fig. 2, eEVA is running on HSR, and Fig. 1(b) shows how eEVA is presented on Toyota HSR. All the relevant HSR components are listed in Table 2. The actuators shown in Table 2 are used in parallel to generate the motions discussed in Sect. 4.

Fig. 2. eEVA running on Toyota HSR

Table 2. Listing of most significant Toyota HSR hardware components. The highlighted components are used for the pilot study.

Type	Short name	Function description
Sensor	Head Microphone	Speech Recognition
Sensor	Head 3D sensor	RGBD Camera
Sensor	Head Stereo Camera	Perception
Sensor	Head Wide Camera	Perception
Sensor	Laser Range Sensor	Perception
Actuator	Arm	Manipulation (Gesture synthesis)
Actuator	Head Display	Displays eEVA
Actuator	Head	Tilts and pans (Gesture Synthesis)
Actuator	Body	Goes up and down (Gesture synthesis)
Actuator	Gripper	Grasping (Gesture synthesis)
Actuator	Omni-Directional Base	Motion
Actuator	Speakers	eEVA voice

---

[3] http://www.ros.org/.

[4] http://moveit.ros.org/.

[5] https://ompl.kavrakilab.org/.

4 Pilot Study: Culturally-Sensitive Greetings on HSR with RoboCanes-VISAGE

We investigated what the effects of a multimodal virtual agent as a UI are, and whether we can develop a multimodal virtual agent UI that is more enjoyable than a robot without such a UI.

We aimed at testing the following hypotheses:

– **H1:** Users find eEVA's 3D character with speech recognition as the HSR UI more enjoyable and competent over an HSR robot UI with speech recognition without eEVA's 3D character.
– **H2:** eEVA's 3D character as UI with speech recognition does not make the HSR UI with speech recognition more eerie, annoying, or boring compared to the HSR robot default UI with speech recognition.

In our pilot study, the user stood about one meter away from the robot in the lab, and the interaction exhibited turn-taking behavior. Each interaction was initiated by eEVA greeting the user: **"Hi, I am Amy. How is it going? How do you greet?"**. eEVA uses Google Chrome API for speech recognition and Windows SAPI for speech synthesis (see Table 1 and Sect. 3.2). After eEVA received the user's greeting preference, the user greeted the robot from four greetings (see below), and the robot portrayed the corresponding pre-greeting gesture. The interaction is concluded when the robot performs the greeting gesture chosen by the user. When the robot finishes greeting the user, the user is allowed to get greeted by the robot again (study setup is shown in Fig. 3).

We established four short social interactions with the RoboCanes-VISAGE framework. The four greetings identified below represent diverse forms of greeting, which vary to reflect cultural influences via the HSR's robot specific motions, coupled with the eEVA human-robot interface: 1. Japanese greeting (**Bow**) as shown in Fig. 3(a). When the user says, "hello" in Japanese, "Konnichiwa", the robot lifts its torso and bows by tilting its head forward. 2. **Fist bump** as shown in Fig. 3(b). When the user says, "Hey, bro!", the robot lifts its torso and moves its arm forward while closing its fist. The user is able to pound the fist of the robot. (this is the only interaction that involves physical contact with the user). 3. **Shaka**, the Hawaiian greeting as shown in Fig. 3(c). When the user says, "Shaka", the robot performs a Shaka gesture. The Shaka gesture involves the robot lifting its hand and moving it side to side. 4. **Hand Waving** greeting. When the user says, "hello", the robot moves its hand up and down, *i.e.*, simulating a wave arm motion.

4.1 Participants

There were a total of 32 participants from the University of Miami Computer Science department that took part in the pilot study (age $M = 41$, $SD = 13$). There was a total of 17 females and 15 males that completed the experiment. Data from one participant was excluded because the participant did not complete the whole questionnaire.
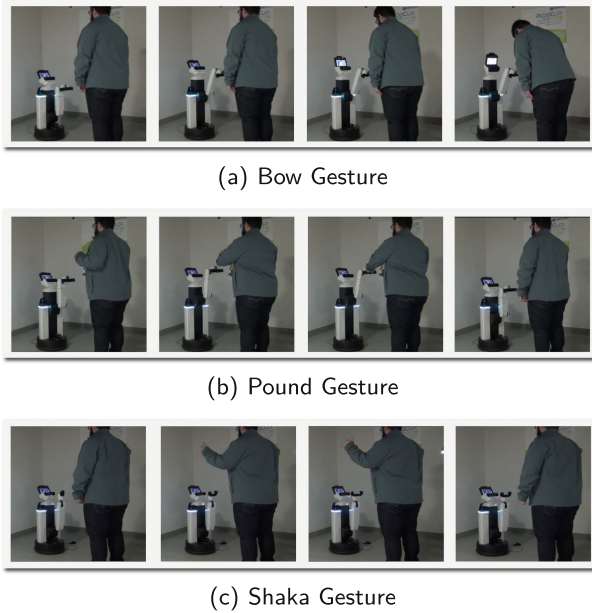
(a) Bow Gesture

(b) Pound Gesture

(c) Shaka Gesture

Fig. 3. Gestures used for pilot study

4.2   Experiment Design and Procedure

A small number of participants interacted with Toyota HSR with eEVA's voice, and the screen of the robot had the visual default HSR splash screen as shown in Fig. 1(a). We compared their interaction experience with users who interacted with Toyota HSR with eEVA's 3D character as the visual interface element and eEVA's voice as shown in Fig. 2.

We split the participants into two groups: one group of 19 participants (age $M = 40$, $SD = 12$) who interacted with Toyota HSR with eEVA (face and voice Fig. 1(b)) and another group of 13 participants (age $M = 41, SD = 13$) who interacted with Toyota HSR with eEVA's voice and HSR default screen (see Fig. 1(a)). At the end of the interaction, we asked the participants to fill out a questionnaire with 7-point Likert scales about how they felt about the interaction of the robot and their feelings toward the robot itself, and conducted an unstructured interview for qualitative data.

4.3   Results

The data was analyzed using the Mann-Whitney test. For this experiment, $n_1 = 19$ and $n_2 = 13$ with a critical $U = 72$. An alpha level of 0.05 was used to analyze the data. There was no significant difference reported for each Likert scale. The competent category was very close to the critical U-value but was not significant enough. No significant differences were found between both groups with regards

Table 3. Overall impression of eEVA as a human-robot interface

Category	eEVA			eEVA's voice only			Mann-Whitney test	
	Q1	Median	Q3	Q1	Median	Q3	U ($Critical\ U = 72$)	p
Enjoyable	6	6	7	5	7	7	108.5	0.56
Boring	1	2	3	1	1	2	101	0.39
Natural	2	4	5	2	3	4	108	0.55
Friendly	5.5	6	7	4	6	6	107.5	0.54
Competent	5	6	7	4	4	6	92	0.22
Scary	1	1	2.5	1	1	1	111	0.63
Annoying	1	1	1	1	1	1	105.5	0.49

[*]Significant $p < 0.05$ (Likert scales are 7-point scales)

to age ($p = 0.79$) and experience interacting with robots ($p = 0.42$). Details can be seen in Table 3.

4.4 Discussion

Although no significant differences were found in all categories, interesting conclusions can be made from this pilot study. First, it is important to note that no significant difference was found in the *scary, annoying*, nor *boring* category. Therefore our second hypothesis **H2**, *eEVA does not make the human-robot interaction more eerie, annoying, or boring* is supported by our results. We concluded that eEVA as an virtual agent human-robot interface might be acceptable to users.

The first hypothesis, **H1**, is not supported by our quantitative results. However the qualitative data we acquired in the study revealed interesting observations that we will investigate in future research. For example, participants in the study requested to interact with HSR for a longer period. One user asked "Will the robot say something else?", and another user asked "Can it do something else?" These observations indicate that a longer interaction might be needed to allow the user to interact with eEVA for a longer period of time to generate an accurate evaluation. This also indicates that users enjoyed the HSR interaction enough to want longer interactions with it, which is a measure of engagement; many users asked, "Can I try all four greetings?" (in fact, 100% of all users used all four greetings). We also noticed that users who interacted with eEVA were trying to get closer to the screen suggesting that the size of HSR's screen might also have an effect on the interaction (*i.e.*, in this case, the HSR screen might be too small to generate an effect in the experience of the interaction).

Another factor in the interaction that might deter our results to be statistically significant is the current hardware of Toyota HSR which evokes aspects of a human face: the two stereo cameras and the wide angle camera on the Toyota HSR resemble two eyes and a nose. During the interaction, users were seen gazing at HSR's stereo cameras rather than the screen. One user mentioned the stereo cameras were distracting when interacting with eEVA.

Henceforth, in future formal studies we plan to investigate the following questions, among others: Does eEVA on different screen sizes on the HSR affect the user's experience such as user's feelings or user's perception of the robot's characteristics? Does HSR's anthropomorphic features (two stereo cameras as eyes and wide angle camera as nose) affect the user's experience such as user's feelings or user's perception of the robot's characteristics? If the answer to the previous question is yes, do users prefer eEVA as a human-robot interface for Toyota HSR without an anthropomorphic face, Toyota HSR with an anthropomorphic face but without eEVA, or both, eEVA and anthropomorphic face?

5   Conclusions and Future Work

In this article, we described a system that integrates both frameworks (eEVA and RoboCanes) under one synchronized system that takes human input such as eye gaze and user speech, and outputs a personalized human-robot interface with greeting gestures.

Our pilot study to assess the effects of eEVA as a human-robot interface for Toyota HSR revealed no significant differences in enjoyment, friendliness, competence, uncanniness, and other categories when comparing Toyota HSR with and without eEVA. We concluded that eEVA's character does not make Toyota HSR more uncanny, boring, or annoying.

In our future research, we will make a formal experiment to study further effects of eEVA on Toyota HSR. This will include making the interaction with Toyota HSR for a longer period of time to answer users' wish to interact longer with the robot (with or without eEVA).

References

1. Abdollahi, H., Mollahosseini, A., Lane, J.T., Mahoor, M.H.: A pilot study on using an intelligent life-like robot as a companion for elderly individuals with dementia and depression. arXiv preprint arXiv:1712.02881 (2017)
2. Al Moubayed, S., Beskow, J., Skantze, G., Granström, B.: Furhat: a back-projected human-like robot head for multiparty human-machine interaction. In: Esposito, A., Esposito, A.M., Vinciarelli, A., Hoffmann, R., Müller, V.C. (eds.) Cognitive Behavioural Systems. LNCS, vol. 7403, pp. 114–130. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-34584-5_9
3. Amini, R., Lisetti, C., Ruiz, G.: HapFACS 3.0: FACS-based facial expression generator for 3D speaking virtual characters. IEEE Trans. Affect. Comput. **6**(4), 348–360 (2015)
4. van der Burgh, M., et al.: Tech united eindhoven@ home 2017 team description paper. University of Technology Eindhoven (2017)
5. Cavedon, L., et al.: "C' Mon dude!": users adapt their behaviour to a robotic agent with an attention model. Int. J. Hum Comput Stud. **80**, 14–23 (2015)

6. Cowan, R.E., Fregly, B.J., Boninger, M.L., Chan, L., Rodgers, M.M., Reinkensmeyer, D.J.: Recent trends in assistive technology for mobility. J. Neuroeng. Rehabil. **9**(1), 20 (2012)

7. Gratch, J., et al.: Can virtual humans be more engaging than real ones? In: Jacko, J.A. (ed.) HCI 2007. LNCS, vol. 4552, pp. 286–297. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-73110-8_30. http://dl.acm.org/citation.cfm?id=1769622

8. Hashimoto, K., Saito, F., Yamamoto, T., Ikeda, K.: A field study of the human support robot in the home environment. In: 2013 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO), pp. 143–150. IEEE (2013)

9. Herath, D.C., Kroos, C., Stevens, C.J., Cavedon, L., Premaratne, P.: Thinking head: towards human centred robotics. In: 2010 11th International Conference on Control Automation Robotics & Vision (ICARCV), pp. 2042–2047. IEEE (2010)

10. Hill, R.W., Gratch, J., Marsella, S., Rickel, J., Swartout, W., Traum, D.: Virtual humans in the mission rehearsal exercise system. Kunstliche Intelligenz (KI J.) **17**(4), 5–10 (2003). Special issue on Embodied Conversational Agents

11. Klüwer, T.: "I Like Your Shirt" - dialogue acts for enabling social talk in conversational agents. In: Vilhjálmsson, H.H., Kopp, S., Marsella, S., Thórisson, K.R. (eds.) IVA 2011. LNCS (LNAI), vol. 6895, pp. 14–27. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-23974-8_2

12. Knight, H., Simmons, R.: Expressive motion with x, y and theta: Laban effort features for mobile robots. In: 2014 RO-MAN: The 23rd IEEE International Symposium on Robot and Human Interactive Communication, pp. 267–273. IEEE (2014)

13. Kofman, J., Wu, X., Luu, T.J., Verma, S.: Teleoperation of a robot manipulator using a vision-based human-robot interface. IEEE Trans. Industr. Electron. **52**(5), 1206–1219 (2005)

14. Lisetti, C., Amini, R., Yasavur, U.: Now all together: overview of virtual health assistants emulating face-to-face health interview experience. KI - Künstliche Intelligenz **29**(2), 161–172 (2015). https://doi.org/10.1007/s13218-015-0357-0

15. Matarić, M.J.: Socially assistive robotics: Human augmentation versus automation. Sci. Robot. **2**(4) (2017). eaam5410

16. Nagahama, K., Yaguchi, H., Hattori, H., Sogen, K., Yamamoto, T., Inaba, M.: Learning-based object abstraction method from simple instructions for human support robot HSR. In: 2016 IEEE International Conference on Advanced Intelligent Mechatronics (AIM), pp. 468–475. IEEE (2016)

17. Pelachaud, C., Bilvi, M.: Computational model of believable conversational agents. In: Huget, M.-P. (ed.) Communication in Multiagent Systems. LNCS (LNAI), vol. 2650, pp. 300–317. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-44972-0_17

18. Pelachaud, C.: Modelling multimodal expression of emotion in a virtual agent. Philos. Trans. Roy. Soc. London. Ser. B Biol. Sci. **364**(1535), 3539–3548 (2009). http://www.ncbi.nlm.nih.gov/pubmed/19884148

19. Predinger, H., Ishizuka, M.: The empathic companion: a character-based interface that addresses user's affective states. Appl. Artif. Intell. **19**, 267–285 (2005)

20. Qiu, S., Li, Z., He, W., Zhang, L., Yang, C., Su, C.Y.: Brain-machine interface and visual compressive sensing-based teleoperation control of an exoskeleton robot. IEEE Trans. Fuzzy Syst. **25**(1), 58–69 (2017)

21. Reeves, B., Nass, C.I.: The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. Cambridge University Press, Cambridge (1996)

22. Simmons, R., et al.: Grace: an autonomous robot for the AAAI robot challenge. Technical report, Carnegie Mellon University (2003)
23. Suero, E.M., et al.: Improving the human-robot interface for telemanipulated robotic long bone fracture reduction: Joystick device vs. haptic manipulator. Int. J. Med. Robot. Comput. Assist. Surg. 14(1) (2018)
24. Wolpert, D.M., Doya, K., Kawato, M.: A unifying computational framework for motor control and social interaction. Philos. Trans. Roy. Soc. B: Biol. Sci. **358**(1431), 593–602 (2003)
25. Yew, A., Ong, S., Nee, A.: Immersive augmented reality environment for the tele-operation of maintenance robots. Procedia CIRP **61**, 305–310 (2017)