<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Fusing realities in human-robot social interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Authors(s)</strong></td>
<td>Dragone, Mauro; Duffy, Brian R.; Holz, Thomas; O'Hare, G. M. P. (Greg M. P.)</td>
</tr>
<tr>
<td><strong>Publication date</strong></td>
<td>2006-05</td>
</tr>
<tr>
<td><strong>Conference details</strong></td>
<td>37th International Symposium on Robotics ISR/Robotik, Munich, Germany, May, 2006</td>
</tr>
<tr>
<td><strong>Item record/more information</strong></td>
<td><a href="http://hdl.handle.net/10197/4543">http://hdl.handle.net/10197/4543</a></td>
</tr>
</tbody>
</table>
FUSING REALITIES IN HUMAN-ROBOT SOCIAL INTERACTION

Mauro Dragone  
University College Dublin  
Ireland

Brian R. Duffy  
University College Dublin  
Ireland

Thomas Holz  
University College Dublin  
Ireland

Gregory M.P. O'Hare  
University College Dublin  
Ireland

Speaker: Mauro Dragone, University College Dublin, Ireland (mauro.dragone@ucd.ie)

Topic: Simulation techniques for system design
Keywords: Robot-Human Interaction, Multi-Agent System, Virtual Characters, Mixed Reality

Abstract

As robots become more and more embedded in our physical and social environment, their integration into our social interaction space necessitates mechanisms which manage these new social contexts. While considerable work has been invested in developing strong human-like robots in order to arguably augment the human-robot interaction experience, the core complexity and significant costs with such an approach render it difficult to justify for practical real-world applications. This paper discusses the use of augmented reality as a tool to bypass this issue and allow the designer, and subsequent user, to easily choose an aesthetic with associated social behavioural mechanisms. Not only does this allow a user to perceive an associated form with a robot, but also allows many people to perceive alternate forms for the same robot, a degree of social customisation which has been impossible until now.

Introduction

The development of the field of social robot research in recent years has lead to a strong need for coherent control frameworks for heterogeneous robots, which support explicit social interaction between robots and between these robots and humans. Human-robot interaction research involves empowering a robot with the social functionality required to engage human participants in some form of directed social engagement. The humanoid form has traditionally been seen as the obvious strategy for integrating robots successfully into these environments. Such systems often involve building robotic devices with a degree of anthropomorphic representation (head, body, facial expressions, hand gestures, etc) (see [16] for a discussion). The control systems generally employ key human-centric interaction modalities such as speech and even models of emotion in order to realise as natural a social interaction as possible. In hardware-only domains, behavioural resolution for facial expressivity, for example, is very limited. Strong anthropomorphic paradigms in HRI overly increase people’s expectations of the system’s performance, and subsequently severely load the behavioural complexity required to succeed. Mori highlighted the problematic issues found in developing anthropomorphic systems with “The Uncanny Valley” [30]. His thesis is that the more closely a robot resembles the human, the more affection it can engender through familiar human-like communication references. However, there is a region in the design space where the robot appears uncanny and weird. Issues of speed, resolution and expression clarification based on often very subtle actions provide for a highly complex design arena.

This paper discusses the use of augmented reality as a tool to bypass these issues and allow the designer, and subsequent user, to easily choose an aesthetic with associated social behavioural mechanisms. Within this paper we describe and employ our multi-agent middleware the Social Situated Agent Architecture (SoSAA) [14] and associated tools [12] in order to develop and interact with agents composed of physical robotic bodies and virtual avatars represented on top of them through augmented reality overlay. The following sections outline this new approach towards developing flexible social machines and summarise our research to date.

Bridging the Gap between HCI and HRI

The recent pervasiveness of robotic platforms has acted as a catalyst for a large body of research investigating human-robot interaction. This research has shown that humans treat computers as equal social partners if they behave in a socially competent manner [37], thus facilitating natural interaction. To fulfil this goal, robots need to exhibit social and emotional intelligence, not only for the good of the human but also for that of the robot [9].
Part of this intelligence is the ability of human and robot to understand and anticipate the other’s intention [9][48]. Humans convey this information through explicit communication as well as an array of non-verbal cues [25], which have been studied, among others, by traditional digital animators who have developed a set of principles based on their observations, e.g. anticipation (providing cues to an action before undertaking it) [46]. Van Breemen [48] argues that applying such tried and tested concepts can help humans understand the robot’s intention. Bartneck has shown that affective expressions of machines can appear to humans as convincing as those of a human – regardless of the level of abstraction – as long as the expressions remain distinct from each other [8] and consistent with user expectations [40].

However, due to physical constraints expresional capabilities of robots remain restricted when compared to humans (see [7] for a recent review of robotic user interfaces) and thus fail to convey subtler meanings of intention or emotion. First steps towards building a realistic human-like companion with rich visual expressiveness have been taken [19], but still suffer from limitations and high cost. On the other hand, in a wide range of robotic platforms for which a robot-human interface was not included in the initial design (i.e. see Figure 2a) implementing such capabilities is difficult, especially with the current lack of tools available for such task. Furthermore, once a robot is built, it is more or less fixed in its appearance, which cannot be tailored to the individual user or modified to adapt to other circumstances.

In contrast, virtual characters offer a number of advantages compared to physical robots. The can possess highly expressive interfaces that are as convincing as a human [8], are comparably cheap, can be easily adapted (personalised) and are also capable of actions that are impossible in the real world, such as mutating the form of their avatar [28]. They constitute a convenient and readily available medium for the investigation of Human-Computer Interaction and can be utilised in Human-Robot Interaction studies as well [41].

However HCI and HRI, though sharing common issues, are by no means identical and should be regarded as distinct fields of research [9]. Yamato et al. have demonstrated that, even though a virtual avatar can exercise more influence on humans, robots are regarded as more familiar [50]. Robots also seem to appear more convincing, more entertaining and offer a greater feeling of shared space than animated characters [24]. It is therefore important to remain mindful of such differences.

A logical conclusion is then to combine the two worlds, virtual and physical. A first incarnation of this principle is the addition of a screen to the robot. This has been widely used as a form of telepresence in remote conferencing scenarios [52], and as robot-human interface for social robots [7], but we believe this approach suffers some crucial drawbacks. The screen can only be seen from one angle and cannot be adapted for each individual user. Furthermore, the characters and their actions are still limited to the screen. While 2D embodied virtual agents can point to things in a 2D environment, they cannot effectively point (or look) in a 3D space. Another consequence of this is that user studies and experiments investigating these skills and their influence on human-robot social engagement (e.g. [43][42]) are often restricted to physical robotic platforms, sometimes with limited expressive and sensory-motor competences (e.g. stationary robots) compared to much more expensive humanoid mobile robots.

Our thesis is that mixed reality [29] offers a potential solution to these discrepancies. Only if real and virtual world are seamlessly integrated can the gap between the two be really bridged. We therefore introduce the notion of a Mixed Reality Agent (MiRA), i.e. an agent consisting of real and virtual components that can be perceived as a single entity availing of an augmented reality display, e.g. a Head Mounted Display (HMD) (see Figure 1).

For sake of clarity, we observe that there are different interpretations of mixed and augmented reality but since our agents live in both the real and the virtual world and can be realised across the whole of Milgram’s Virtuality Continuum [29], we find that they are indeed mixed reality agents, as opposed to being ‘merely’ augmented agents. We believe that this approach offers some compelling advantages. The merging of the real and the virtual facilitates the agent in overcoming the limitations of each. A mixed reality agent exhibits tangible physical presence while offering rich expressional capabilities and personalisation features that are complex and expensive to realise with physical robots.

Contrary to screen-based solutions, the virtual avatar is now only constrained by the user’s field of view. As the mixed reality agent is embodied in the real and the virtual world it can manipulate both and can interact with physical as well as with virtual objects. This interaction transcends the boundaries between the real and the virtual world; the virtual avatar can be perceived, for example, pointing or looking at real things, while the physical robot can steer around virtual obstacles. Also, as the virtual part is rendered individually on each user’s equipment, it offers the compelling possibility of personalising the content for each user.
A distinct disadvantage, at the moment, is the cumbersome and expensive hardware imposed on each user. However, this situation is on the verge of a change as both head-mounted displays and wearable computers become cheaper and less invasive [26]. Currently, our research investigates the applicability of mixed reality agents in two distinct contexts, namely:

*Figure 1. View of the same scene with a physical robot and with a mixed reality agent*

**Human-Robot Interaction in Ambient Space**
Our group investigate tools and architectures for the control of a collective of (heterogeneous) mobile robots; their coordination and their integration in a larger social context populated by other diverse robots, software agents and humans. The advantages of combining the unique capabilities of robots, software agents and humans have been reinforced through such projects as RAP (Robot-Agent-People) [39] in terms of a proxy-based architecture for flexible teamwork among agents interconnected by an explicit communication medium. Our work adopts a wider perspective by also embracing the notion of ambient social intelligence [2], the result of endowing ubiquitous (e.g. wearable) devices with autonomous behaviours for the creation of intelligent environments.

It is our thesis that mobile robotics are a compelling instance of those artefacts which comprise and deliver the ambient space. Clearly, mixed reality agents leverage the diffusion of wearable computing devices interconnected by wireless networks. Also, in this context, augmented reality interfaces enable sophisticated interactions between robots and humans by enhancing their mutual awareness thanks to the exploitation of the necessary network link. Identification and spatial referencing (i.e. tracking of humans and their gaze) during social engagement, for example, is greatly facilitated by availing of the same tracking information employed for the realisation of the augmented reality overlay.

**Prototyping and Experimentation Tools for the Development of Social Robots**
We also investigate mixed reality agents, and augmented reality in general, in aid of the prototyping and development of physical, fully autonomous, social robots. The change in perspective brought by collective and social robotics exacerbates some of the issues associated with the development of robotic systems, in particular those related with the conduction of the necessary experiments and the subsequent evaluation or comparison of the components under development.

A common practice in robotics is to build architectures having significant components that transfer across physical and computer-simulated environments and subsequently use these computer simulations to test the robot controllers. Similarly, virtual characters are widely employed as a validation or preliminary step toward the development of skills for physical robots.

In other fields, e.g. in manufacturing industries in the aviation and automotive sectors, Collaborative Virtual Environments (CVE), distributed virtual spaces in which people can meet and interact with others, and augmented reality have been systematically employed for engineering and prototyping purposes. Within the same fields, hardware-in-the-loop simulations (HILS) [27] have also become a consolidated practice for the testing of complex distributed embedded control systems, often in conjunction with virtual and augmented reality visualisation. These solutions have the advantage of preserving the richness and the complexity of system-environment interaction while allowing researchers to test and scrutinise the behaviour of specific isolated components.
Robotics research has recently started to explore the utilisation of similar powerful techniques (e.g. [45][3]). These frameworks allow the instantiation and the monitoring of experiments that combine real and simulated entities, e.g. simulated robots working along physical ones in virtual or mixed reality environments, but they do not support the kind of mixed reality agents (half real/half simulated) that we describe in this paper. Consequently, social skills for robots (e.g. gaze behaviour for a robot head), which require engagement with a human, must still be developed and tested using distinct environments, first relying on work on virtual characters then on fully implemented prototypes.

Immersion of human users in virtual environments inhabited by simulated virtual robots can obviously respond to some of the needs of this research. However, this solution has the same disadvantages as fully simulated experiments for the development of traditional robotic behaviours as simulation can rarely match the complexity of real environments. Engaging a virtual robot in a virtual environment is substantially different from engaging a robot that is physically present in the real environment, as the physical robot is capable of sensing, reacting to and operating upon the real environment while, at the same time, exhibiting the intended expressivity and manifesting the behaviour under development thanks to its virtual component. We therefore advocate mixed reality agents as a prototyping tool for the development of social robots in situations in which it is useful to test social behaviours in conjunction with physical presence in real environments.

In the following sections we describe the SoSAA framework, which is employed for the realization of mixed reality agents and illustrate some of the work we produced to date.

The SoSAA Middleware

In delivering the agent components in the architecture we commission the Socially Situated Agent Architecture (SoSAA) [14], an agent-based middleware for the development of distributed autonomous systems based upon Agent Factory (AF) [10].

AF implements a Belief- Desire-Intention (BDI) [36] model of agency that uses the mental attitudes of Beliefs, Desires and Intentions in order to represent, respectively, the information, motivational, and deliberative states of the agents. The role of these attributes is to provide the agent with a usable description of both present and future states of the agent’s environment. AF agents employ practical reasoning techniques to deliberate upon their perceived situation, update their mental state and select a future line of action in pursuance of system goals.

AF defines an interface layer based upon abstract actuator and perceptor modules that are used to connect the reasoning engine with domain specific systems. This mechanism is instrumental for the ability to create hybrid robotic systems [11] combining deliberative components with behaviour-based controllers and also for controlling both real and virtual agents (e.g. simulated robots or virtual characters) [15].

To enable collaboration among any type of agents within a specific application, AF agents make use of a FIPA (http://www.fipa.org) compliant Agent Communication Language (ACL). Through ACL directives such as request, inform, propose, or reject, the agents can influence each other’s conduct.

The Virtual Robotic Workbench

The connectivity for the realisation of mixed reality agents is based upon the multicast, multi-channel messaging service of the Virtual Robotic Workbench (VRW) [14], an agent-based collaborative virtual environment for social robotic agents.

Current CVE systems are the result of a convergence of research interest within the VR and Computer-Supported Cooperative Work (CSCW) communities. Within the CSCW community, the ability of CVEs to simulate the physical presence of users in a shared environment has been instrumental in enabling complex social interactions which has previously eluded technologies such as audio and video-conferencing and shared desktop applications. For efficient collaborative work, CVE systems necessitate HCI techniques as well as specific instruments facilitating the coordination and the social engagement between participants. Increasingly more systems also incorporate agent-based techniques (e.g. [49]) to improve the usability of the system. Agents act as virtual representatives of human users and as fully autonomous intelligent entities co-habiting the same-shared space.

The core features of the VRW are the immersion of both robots and humans in a networked collaborative environment supporting robot-robot and human-robot interaction. For this purpose, the VRW contains an XML protocol for the transport of both ACL directives and telemetry information (i.e. sensor, positional and tracking updates). The XML tagging mechanism is easily extendable and can naturally support data streaming at different rates and of different content
type. For mixed reality agents, the workbench messaging system is also exploited for the dynamic discovery of peer observers and for advertising static data concerning the avatars and the marker cubes thus fully distributing the configuration of the application.

**Robot Hardware**

Our work employs wheeled robots, specifically, ER1 (www.evolutionrobotic.com) with SICK (www.sick.de) LMS laser scanners and Nomad Scout robots (see Figure 1), with sonar sensing. Every robot in the team is also equipped with image sensing and simple ball pushing devices enabling basic ball-handling behaviours in the context of robotic soccer experiments.

**Augmented Reality**

Figure 2 illustrates the typical setting for the realization of mixed reality agents. It depicts a user wearing a computer and a see-through head-mounted display with associated digital camera and microphone. Through the HMD, the user can see the live scene, depicting the real robots, augmented by superimposing synthetic imagery showing the associated avatars and the other components of the augmented reality interface.

The overlay between real and virtual images is realized by tracking the position and orientation of the HMD of each observer in the coordinate frame of each observed robot. This information is then used to align the image of the virtual character with the associated real robot. As a simple and cost-effective way to implement this tracking, we place 3D physical markers on top of the robots and track them in the video captured from the HMD camera.

Tracking and rendering functionalities in the current implementation are based upon ARToolkit [23], a software library for the recognition and pose estimation of square markers within a camera image. We arranged five different markers upon the visible faces of a cube (visible in Figure 2, on top of the Nomad Scout robot and below the snowman avatar) to make the robot traceable from all angles. In order to improve the precision and the reliability of the tracking, a Kalman Filter [22] is used to integrate the observation of multiple faces of the cube over time. As in [34] we implemented a shared tracking system in which multiple observers (i.e. different users and/or fixed cameras) may communicate with each other in order to collaborate in the tracking effort and to increase both precision and coverage. The VRW messaging service is exploited for the dynamic discovery of peer observers and the dissemination of data through the network. Such data includes tracking observations, robot telemetry (i.e. positional updates), and other information regarding the behaviour and the cognitive state of the robotic agents and the associated virtual characters. Furthermore, the VRW manages the run-time distribution of configuration data (e.g. broadcasting the picture of markers upon the observation of a previously unknown cube), without the need for explicit mirroring of complex configuration files. More details of the tracking and communication systems can be found in [14].

It is important to note here that, at present, the marker based optic tracking is a cheap and easy solution to augmented reality but it suffers from a couple of shortcomings, most notably, the sensitivity to different lighting conditions and the failure to track when the cubic marker is partially occluded. However, a number of researchers have recently developed marker free vision based tracking (e.g. [47]) that could be used to track the pose of the robot itself instead of a marker, even when it is partially occluded by other objects in the environment.
Multi-Agent System Integration

Mixed reality agents look to seamlessly integrate physical robots with augmented reality overlays in real-world environments. To this end, the SoSAA framework offers distributed behavioural rendering through managing the appearance of both simulated robots and virtual characters (e.g. virtual heads or full body avatars, see Figure 3b) associated with real robots.

The functionalities of mixed reality agents emerge from the collaboration of a network of distributed agents: agents in control of the robotic platforms (robotic agents), agents managing the user interfaces (user interface agents), and agents in control of the virtual avatars (avatar agents). In the current implementation (see Figure 2a), for each user, both the user interface agent and the avatar agent reside on the viewers wearable computer and are connected to the other participating robots and users via the VRW messaging system.

This multi-agent system is thus composed by virtual and real elements, which ought to exhibit cohesion and behavioural consistency to the observer. This is partly facilitated by the adopted robot-centric solution for augmented reality tracking, but in itself is insufficient. In order to be a believable component of the mixed reality agent, the behaviour of the virtual component needs to exhibit a degree of awareness of its surroundings, comparable to a robot being physically embodied through an array of physical sensors and actuators. The instrument for such situatedness is the update sensor stream, which notifies the avatar agent about the nature and relative position of the obstacles, and other objects of interest perceived by the robot. This allows the direct-control of deictic and anticipatory animations [46], to project the intention of the mixed reality agent toward perceived objects of interest.

Animations can also be associated with specific functions performed by the physical robot or to increase the human understanding of the robot’s state. For example, the avatar could nod or salute to acknowledge a user’s vocal command, or shrug its shoulders if the input has not been understood. Being virtual artefacts, these animations are not just limited to ‘natural’ human-like forms, but can also include more complex effects involving other virtual objects. A flash bulb could for example be displayed when the robot agent is initiating a new task or activating a new plan, or a virtual soccer ball visualised near the head of the avatar whenever the real ball is recognised in the robot’s camera image (see Figure 5), thus informing an observer that the robot is recognizing the ball and tracking its position.

Robot Agent. The robot agent manages the robot conduct. It is also connected with the VRW messaging system that is used to receive observation updates, user commands and to update other participants regarding its internal state (i.e. sending telemetry and sensorial information when requested or reporting its BDI mental state to the observer). Robotic control is achieved through the integration of this BDI intentional layer with a reactive, behaviour-based robotic system [4]. In order to enable the soccer game scenario, in addition to obstacle avoidance and navigation capabilities, the robotic controller includes object recognition functionalities (based on blob colour tracking) used to recognize the ball and the other robots and low-level reactive behaviours for dribbling, kicking and passing the ball. The control system also includes a spatial referencing sub-system used for the definition and maintenance of way-points which can be associated to objects tracked with the on-board camera or initialised to arbitrary positions registered with the robot odometry system. Successively, these way-points can be used as inputs for behavioural modules (i.e. for the behaviours FaceObject or MoveToward). The mechanism enables behaviour’s persistence. Thus the robot can keep approaching the ball even when it gets momentarily occluded. It also supports multiple foci of interest. Thus the robot can turn toward the user and then return to pursuing the ball.

User Interface Agent. The user interface agent controls the display of text and other 2D graphic overlays in the user HMD. Through them, the user can be informed of details of the task and the state of the other participants (both robots and humans). This agent also processes user utterances availing of the IBM ViaVoice™ speech recognizer and the Java Speech API (http://java.sun.com/products/java-media/speech/). The vocal input may be used to trigger a predefined set of commands in order to configure the local interface (i.e. requesting to tune on a particular VRW channel) or issue requests to the robots. To do this, the recognized text is matched and unified with a set of templates and the result transformed in the correspondent ACL directive, e.g. request(?robot, follow(right)).

Avatar Agent. Since the rendering of the virtual component is performed on the wearable computer connected to the human observer’s Head Mounted Display, an Avatar Agent is in charge of the behavioural rendering for each robot in the scene. This agent is responsible for monitoring state information and events transmitted by both robot and observer through the VRW and to keep the behaviour of the virtual character in line with the perceived activity. This is achieved by activating a set of animations (e.g. hand gestures and facial expressions) available to the avatar, a degree of flexibility difficult to achieve on non-augmented platforms. Using an agent in this context avoids the need for a statically pre-defined mapping, enabling instead a more reasoned, and more easily definable control over the avatar’s appearance. For example, the avatar agent may use a stereotypical identity description for the robot and the avatar (see [46] for more details) as well
as a profile of the observer (i.e. his identity and preferences) in order to personalize the form of the avatar and dynamically bind classes of events to the specific visual clues and animations.

**Gaze Tracking and Spatial Referencing**

Since proper overlay of virtual images onto the user’s field of vision requires exact knowledge about the position and gaze of the user, we decided to avail of this knowledge by forwarding it to the observed robot via the workbench data dissemination service and to use gaze direction to influence the robot’s behaviour, to spatially reference virtual and real objects, and to identify which individual out of a team is the focus of the interest of the user in a multi-robots scenarios (in case more than one robot is tracked in the HMD camera image).

Since the user gaze is hidden behind the HMD, it is approximated through the orientation of the camera on top of the HMD, thus effectively reflecting the users head orientation. The robot infers the position of the observer and the direction of his gaze in its own local frame of reference and then projects the user’s gaze vector onto the floor plane.

**Behavioural Rendering in Mixed Reality Agents**

The mixed reality agents have access to a library containing a number of avatar forms they can use as their virtual component. For human-like avatar forms we employ the H-Anim 1.1 standard [18]. The H-Anim specifications define a standard description of the virtual human body in form of a set of segments organised hierarchically and connected by joints. Figure 3a shows one of the humanoid avatars used in conjunction with a Nomad Scout robot. The avatar is part of the distribution of the Virtual Character Animator software by ParallelGraphic and is compliant with the standard.

Based on the VRML/X3D standard for 3D representation, the definition of a H-Anim avatar can be complemented with the other features available in these standards so as to represent human-like characters with different degrees of visual sophistication, from caricature to photo-realistic avatars. Furthermore, a humanoid animation may then be defined in terms of relative movements (i.e. rotation, translation, scale) of these joints over time.

The H-Anim 1.1 specifications, in fact, do not extend the basic support for animation available in VRML. They also do not currently include the ability to represent high-level behaviours (e.g. walk, wave) or define facial expressions. Nonetheless, the standard has acted as catalyst for a large body of research in the 3D and virtual reality communities which share an interest in the incorporation of human-like avatars in virtual or hybrid environments (see [5][38][21][44]). Several working groups are currently developing standards and tools for humanoid animations including facial expressions [6] and more general behaviour specifications (e.g. [51][35]). We feel that our mixed reality agents would greatly benefit from avatars that are able to exhibit complex facial expressions, like gaze behaviour natural language processing and lip synthesis.

In this work, an abstract interface is defined between the BDI agent and the behavioural rendering in the form of perceptors and actuators, to be used by the avatar agents. This design favours a smooth transition to a more sophisticated behaviour-specification system, while at the same time, the avatar agent treats the virtual head or virtual limbs as mechanical devices, using the same interface intended to real hardware. Our current implementation is based upon OpenVRML [32], an open source C++ library enabling a programmatic access to the VRML model. With this mechanism in place, there are two alternative means to execute animate the model, which correspond to different degrees of control over the animation:

- Indirect control, by triggering the activation of pre-stored animations.
- Direct control, by manipulating the avatar’s joints.

Both options enable the use of pre-stored definitions of the animations, which is another advantage of using the H-Anim standard. For example, realistic animations created with key-frame editors (such as Virtual Character Animator) or with motion-capturing technology (as in [44]) can be stored independently from the avatar definition and re-used for different avatars. The main advantage of indirect control is that the interpolation between frames is delegated to the VRML engine, thus freeing our interface from this duty. On the other hand, direct control could also be used to produce the animation at run-time, i.e. availing of (more computationally demanding) forward or inverse kinematics techniques. Alternatively, pre-stored animation could be treated as templates and parameterised before execution, as in [21].

In order to meet our objectives and to achieve a compromise between speed, simplicity and behavioural realism, our design allows the use of both pre-stored and run-time animations by handling them with, respectively, synchronous and asynchronous actuator modules. In addition, in order to approximate a desired behaviour, we segment each H-Animation by using both sequenced and parallel activation of such actuators. The use of such mechanisms can be better illustrated with the example of a mixed reality agent Boy (detailed in section 5.2 and shown in figure 7), which involves playing robotic soccer with a human participant. The Boy avatar points to the ball detected by the robot camera (whose lens can be seen in
the front of the robot just below the red bumper ring). In this example, a continuous animation (triggered by a synchronous actuator activated at start-up) controls the avatar, simulating the breathing pattern (including the rhythmic expansion and contraction of the chest) while an asynchronous actuator (activated when the robot perceives the ball) directly controls the joints involved in the pointing behaviour. A similar combination can be used for the manipulation of facial expression in conjunction with the eye gaze.

A Mixed Reality Agent as a Museum Guide

This work incorporates strong notions of perceptual identity in artificial systems through the use of stereotypes, character (perceived identity) and roles [17]. The SoSAA framework provides a flexible mechanism where users can customise both the agent’s virtual persona and how this is managed through explicit mechanisms for artificial identity. While each agent’s representation is fundamentally grounded on a unique identity, these personalisation mechanisms allow users to select their own preferred avatars in both virtual and augmented reality applications.

As an example of the type of situation where these features could be used, consider a mixed reality agent guide to a museum that is physically embodied as a robot, travelling around the museum, guiding visitors and providing information about the different exhibits. Furthermore, when the user is equipped with a head-mounted display (HMD), the agent’s virtual embodiment, an avatar overlaid on the robot (as shown in Figure 3a), can be seen. Together, this combines the advantages of an expressive virtual agent with the physical presence of a robot.

![Figure 3.(a) The avatar can use gestures and other non-verbal expressions to interact with the user (b) Adult visitors are greeted by an adult avatar form while kids are greeted by a child avatar form.](image)

As a result of this combination of the physical and the virtual, the guide is capable of complex virtual expression, while being physically present within the environment. Furthermore, the virtual avatar can be tailored to the user’s personal model. For example, an adult may see the avatar as an adult whereas a child may see a child avatar (see Figure 3b). As it moves around within the physical space, the virtual avatar animates to help point out key items. For instance, in Figure 1 (right), the avatar points, and directs its gaze, towards a ball. It should be noted that, as the avatar is 3D, the object that it is pointing out can be easily discerned by the user.

As the user is guided around the museum, the virtual avatar can mutate its form to help the user visualise the points that it is illustrating. As an example of this, the guide’s operation within a natural history museum is shown in Figure 4. Specifically, these show the guide’s appearance as it approaches two exhibits, one containing a comparison of carnivore skulls, the other a skeleton of a Giant Irish Elk. At each exhibit, the guide augments its discussion of the exhibit by adopting an appropriate avatar form. In this case, it mutates into a lion while discussing the carnivore skulls (Figure 4a) and a stag while discussing the Elk (Figure 4b).

Throughout the course of the interaction with the user, the guide’s actions are controlled by the agent’s deliberative mechanism. Decisions on where to move the robot, which avatar form to adopt, and which animation to carry out are made based upon the agent’s BDI reasoning. This allows the agent to respond to changes in the environment. When a user
becomes disinterested in a particular exhibit and moves away, the guide may adopt its actions and change its discussion in order to reflect this. Deliberative agent control also allows for the creating of a more dynamic, and hence more believable, agent interaction.

---

**Figure 4.** (a) The guide adopts a lion avatar form at an exhibit of carnivore skulls
(b) The guide changes its form to a stag when discussing the Giant Irish Elk

---

**Human-Robot Collaboration Experiment**

In order to realise practical demonstrations of mixed reality agents, we devised an experiment that avails of the capabilities of our robots to recognize and move a soccer ball. The intent of the experiment was to demonstrate some of the functionalities of mixed reality agents and serve as preliminary study for a larger user trial. In particular, the experiment tests the spatial referencing system, the gaze tracking functionalities and some social behaviour in situations that require collaboration between humans and, potentially, multiple robots. The higher-level skill we adopted was the ability to fetch a soccer ball and deliver it to the user who requested it.

We conducted a series of tests [13] with one Nomad Scout robot operating in a coffee area within the School of Computer Science and Informatics within University College Dublin. In the experiment we adopted a snowman avatar that clothes the tracking cube and part of the robot (see Figure 5).

We asked three different users to try the system. They were instructed prior to the experiment as to the vocabulary they could use with the robot’s speech recognition interface (like ‘move there’, ‘bring me the ball’, etc). In order to help in replicating the test with these different users, they were instructed to remain standing within a half-meter of a given position. This was located approximately in the centre of the experiment area and marked with a circle, one-meter in diameter, on the floor. We ran a number of different trials with each user, changing both the starting position of the robot and the location of the ball at every trial. In order to increase the complexity of the task we also placed armchairs in the environment to create obstacles between the ball, the user and the robot.

By way of example, Figure 5 shows a sequence of snapshots taken during one trial. An armchair was located in front of the user, just between the ball and the robot, with the latter pointing away from the user. In the picture, the mixed reality agent is shown turning toward the user before saluting and listening to the first instruction. In Frame 3, the user imparts a ‘move’ command. In the picture a red arrow, connecting the base of the robot with the user’s gaze, indicates the suggested direction. Subsequently, the robot starts to move in that direction while avoiding the armchair along the way. In Frame 4 the user imparts a ‘turn left’ command causing the robot to turn and sense the ball with the on-board camera. A virtual soccer ball appears near the avatar indicating to the user that the robot has detected the ball. When the user asks for it (‘bring me the ball’) the robot initiates the grab behaviour (Frame 6) and then moves toward the user with the ball between its arms.
A Prototyping and Visualisation Tool

Mixed reality agents can help reduce the cost of hardware development and allow for a greater range of visual functionality to be used to monitor on-line robotic experiments in unstructured and previously unknown environments. We currently employ mixed reality agents for assisting the development of our team of mobile robots. For example, through augmented reality we can have several researchers, each wearing a HMD, monitoring experiments involving either a single robot or a team of robots. This equips the researchers with a live representation of the robots’ activities while they are working on a designed task (e.g. testing a new sensory-motor behaviour or perception skill) instead of having to rely only on offline analysis and visualization tools. The augmented reality equipped experimenter can instead query the robots mental state, debug the live system, and experience a rich visual feedback while observing the real robotic system at work.

Figure 6a shows a mixed reality agent extended with the visualisation of the beams emitted by the Polaroid sonar ring with which the Nomad Scout robot is equipped. In addition, the same mechanisms implemented to increase the human understanding of the robotic system during human-robot collaborative work are also employed during the development process. In other words, we treat this process, consisting of task-oriented trial and observation of partially developed
behaviours, as just another phase in the life cycle of the robot. In this development phase, the goal of the mixed reality agent is to collaborate to the effort by facilitating the debugging of the activities internal to the robotic controller.

To this end, animations and other visual cues in the virtual avatar are used to attract the attention of the experimenter to particular situations (e.g. when the robot is unable to overcome an obstacle) or evidence meaningful events in the BDI reasoning process of the robotic agent (e.g. by flashing a virtual light bulb to inform about a newly activated plan). Finally, having implemented the avatar’s animation mechanism within the same framework we adopted to interface with other robotic hardware (e.g. the robots’ differential drive systems), we now have a powerful system that can be used to test robotic social skills before their actual implementation in hardware while still availing of physical presence of both human and robot in the real environment. In order to demonstrate this use of the mixed reality agents, we are currently in the process of testing an extension to the behaviour repertoire of the humanoid robot “JoeRobot” (Figure 6b) through experiments with the “Kid” mixed reality agent (Figure 3a). Among the behaviours of interest, we are developing a gaze behaviour for the control of the direction of the robot gaze during the fetch-ball collaborative task. The use of mixed reality agents will allow us to test this behaviour in conjunction with a real robotic platform that will sense and act upon the real world, the same world inhabited by the observer.

Therefore we neither need a physical simulation of the ball dynamic nor a simulation of the robot’s image processing. We can carry our experiments simulating different dynamic capabilities of the robotic head, before we actually proceed to change the hardware of the physical robot, either by updating the servo-mechanisms in control of the head movement, or completely changing the design/appearance of the robot head, and finally porting the developed behaviours.

Conclusions and Future Work

The role of the robot in our physical and social space will change drastically over the next decade. While by no means a solved problem, the development of relatively robust control and sensor technologies has allowed researchers to experiment with different hardware and software configurations. To date, the majority of these which have employed virtual reality have only done so in the context of simulations or remote representations for tele-presence and tele-operation. This work embraces virtual and augmented reality from a whole new perspective. The MiRA framework helps address a number of fundamental issues that occur when robots become a part of our social space, including the complex issues of behaviour expressivity and similar human-centric social interaction paradigms. Considerable research has started to investigate those key mechanisms which promote and help manage social interaction between man and machine. This work proposes the adoption of mixed reality agents in the context of ambient space and provides a flexible framework for rapid prototyping of both aesthetic and behavioural features to help develop coherent social systems.

In addition to their appeal to human-robot interaction studies, the implementation of mixed reality agents poses a number of new and complex technological challenges (i.e. the synchronization between the behaviour of the physical and the virtual components and their integrated identity to the user) that we believe are worthy of investigation. Future work will include a user study, including a more formal introduction to the system and the behavioural capabilities of the robot, an assessment of the spatial and perceptual skills of the users and a questionnaire to investigate the usability and other aspects of the system. Such a study will be used to investigate how specific components of the system are helpful in improving the understanding of the robot’s state and its intentions and which are the most effective for human-robot interaction.

Acknowledgments

This research is based upon works supported by the Science Foundation Ireland under Grant No. 03/IN.3/1361.

References


[53] Zhao J. and Badler N., “Inverse kinematics positioning using nonlinear programming for highly articulated figures”, ACM Transactions Graph., vol.13, 1994, pp313-336
Authors

Mauro Dragone received a degree in Computer Science (110/110 Cum Laude) from Bologna University, Italy in 1998. After graduation he worked in a variety of software companies as a software engineer. Thereafter he returned to academia and is now currently a Ph.D. student within the School of Computer Science and Informatics, University College Dublin (UCD), Ireland. His research interests include Multi-agent systems, robotics and virtual and augmented reality.

Brian R. Duffy received his M.Eng.Sc. and Ph.D. degrees from the University College Dublin, in 1994 and 2000, respectively. From 1994 to 1996, he was researcher at the Institut National des Sciences Appliquées de Lyon, France, before spending two years at the Fraunhofer Institute, Germany. From 2001 until 2004, he conducted research in the field of social robotics and human-machine interaction at the MIT Media Lab Europe, Ireland establishing the Anthropos research group, and co-directed nationally funded research projects in the field of multi-agent systems and human-computer interaction. Brian joined the Affective Social Computing Lab at Institut Eurécom, Sophia Antipolis, France in 2005 and conducts research into affective social robot control and human-machine interaction.

Brian is the author of over 50 technical publications including conference proceedings, journal papers, and book chapters. His research interests include artificial intelligence, social robotics, human-machine interaction, affective computing, virtual and augmented reality. He is a Chartered Engineer, holds the title of Eur.Eng. and is an active member of the IEEE and IIE Societies.

Thomas Holz is a Ph.D. student at the School of Computer Science and Informatics at University College Dublin. His research interests include augmented reality and its implications on agents and mobile robotics. He is a graduate of the University of Applied Sciences Harz, Germany and worked on augmented reality at the Fraunhofer Institute for Computer Graphics in Darmstadt, Germany in the GEIST project.

Gregory O'Hare is a Senior Lecturer at the School of Computer Science and Informatics at University College Dublin (UCD) and is Director of the Department's Practice and Research in Intelligent Systems & Media (PRISM) Laboratory. His research focuses on Multi-Agent Systems and on mobile and ubiquitous computing. O'Hare has a B.Sc. in Computer Science and an M.Sc in Information Technology from the University of Ulster. He is a Fellow of the British Computer Society and a Science Foundation Ireland (SFI) Investigator. Gregory is the author of over 200 peer-reviewed publications in journals and international conferences on the broad area of Multi-Agent Systems and Pervasive Computing.