Arjun Nagendran*
Remo Pillat
Adam Kavanaugh
Greg Welch
Charles Hughes
Synthetic Reality Lab
University of Central Florida

A Unified Framework for Individualized Avatar-Based Interactions

Abstract

This paper presents a framework to interactively control avatars in remote environments. The system, called AMITIES, serves as the central component that connects people controlling avatars (inhabiters), various manifestations of these avatars (surrogates), and people interacting with these avatars (participants). A multiserver–client architecture, based on a low-demand network protocol, connects the participant environment(s), the inhabiter station(s), and the avatars. A human-in-the-loop metaphor provides an interface for remote operation, with support for multiple inhabiters, multiple avatars, and multiple participants. Custom animation blending routines and a gesture-based interface provide inhabiters with an intuitive avatar control paradigm. This gesture control is enhanced by genres of program-controlled behaviors that can be triggered by events or inhabiter choices for individual or groups of avatars. This mixed (agency and gesture-based) control paradigm reduces the cognitive and physical loads on the inhabiter while supporting natural bidirectional conversation between participants and the virtual characters or avatar counterparts, including ones with physical manifestations, for example, robotic surrogates. The associated system affords the delivery of personalized experiences that adapt to the actions and interactions of individual users, while staying true to each virtual character's personality and backstory. In addition to its avatar control paradigm, AMITIES provides processes for character and scenario development, testing, and refinement. It also has integrated capabilities for session recording and event tagging, along with automated tools for reflection and after-action review. We demonstrate effectiveness by describing an instantiation of AMITIES, called TeachLivE, that is widely used by colleges of education to prepare new teachers and provide continuing professional development to existing teachers. Finally, we show the system's flexibility by describing a number of other diverse applications, and presenting plans to enhance capabilities and application areas.

1 Introduction

The use of virtual characters and associated environments has been widely adopted in training and rehabilitation scenarios over the last several decades. These virtual characters/environments generally offer the flexibility to recreate specific scenarios and events, while doing so in a controlled and consistent manner. Traditionally, virtual characters have autonomous agency—they are

Presence, *Vol. 23, No. 2, Spring 2014,* 109–132
doi:10.1162/PRES_a_00177

---

*Correspondence to arjun@cs.ucf.edu.

driven by a computer program. Advances in artificial intelligence (such as natural language processing and decision trees) have helped create realistic interaction scenarios (e.g., Rizzo et al., 2013). However, there are still several research challenges associated with open-ended interactions. For example, hampered or interrupted flow during bidirectional conversation can result in a reduced sense of scenario plausibility, and processing errors such as speech recognition errors, repeated responses, or inappropriate responses can detract from the experience or cause harm. To address these and other issues, control of virtual characters may involve a human who inhabits (i.e., controls) the character. The character that is being controlled by a human is referred to as an *avatar*. More formally, a *virtual avatar* is described as a perceptible digital representation whose behaviors reflect those executed, typically in real time, by a specific human being (Bailenson & Blascovich, 2004). In a more general sense, avatars can have physical (e.g., robotic), as well as virtual manifestations. The term *human surrogate* is also used when the avatar is intended to represent the human at some remote destination. In this context, persons who drive their remote counterparts (avatars) are referred to as *inhabiters* (Nagendran, Pillat, Hughes, & Welch, 2012) although the term *interactor* is also used when the inhabiter is a highly trained professional capable of embodying many different, disparate avatars. People who interact with the avatars are referred to as *participants*—these can be active participants who directly influence an interaction or passive participants who merely observe the interaction with an intent to either gain knowledge, analyze performance, or provide guidance to active participants during the interactions. Further distinctions of participants and the roles they may assume is provided in Section 3.3 of this paper.

In this paper, we present a framework and its systems architecture that forms the central component to mediating individualized avatar-based interactions. We call our system AMITIES, for Avatar-Mediated Interactive Training and Individualized Experience System. The acronym has dual meaning, as the word “amities” (derived from Old French) indicates peaceful relationships, friendships, and harmony between individuals or groups. This paper is an extended version of our work presented at the Virtual Reality Software and Technology Conference (VRST; Nagendran, Pillat, Kavanaugh, Welch, & Hughes, 2013) in which we described the AMITIES system architecture without focusing on the individual components that form the underlying basis for AMITIES. AMITIES can be thought of as a binding system between three components that are typically involved during interactions: (1) the avatars; (2) their inhabiters; and (3) the participants. This paper addresses the role of AMITIES in bringing together these components for improved avatar-mediated interactions (e.g., see Figure 1) and presents an instantiation of AMITIES as a case study.

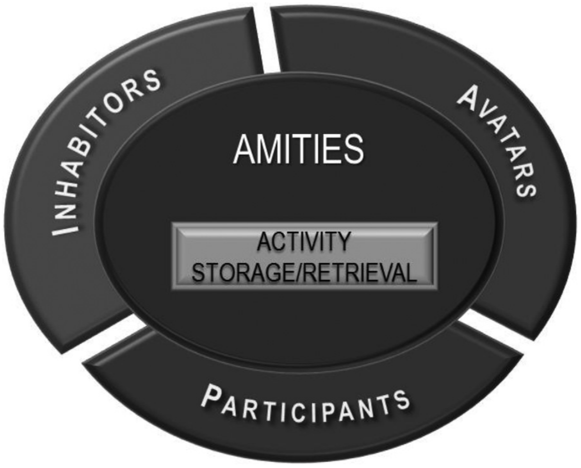


Figure 1. *Components of the proposed system for avatar-mediated individualized interactions.*

The system provides an interface for each of these three components by leveraging technological affordances and avatar mediation to create scenarios that establish, maintain, and preserve user beliefs that are critical to the interaction. In essence, the system attempts to preserve *place illusion* (a sense of “being there/this is happening in my space”) and *situational plausibility* (a sense of “this event is possible”), both of which have been shown to influence human perceptions (Slater, 2009), particularly in virtual-reality-based environments. The AMITIES system features digital puppetry (Hunter & Maes, 2013; Mapes, Tonner, & Hughes, 2011) blended with autonomous behaviors and a network interface to allow inhabiters to control multiple virtual characters seamlessly from remote locations. The system uses a network-efficient protocol during control, thereby minimizing the required bandwidth and hence any associated latencies. Rendering is in the domain of each recipient station and so perceptible lag is avoided. At the user end, the system offers the flexibility for several observers to be involved (passively) during a training session, extending the training impact to additional users. Additionally, the system allows multiple interactors and flexible assignments to support many-to-many, many-to-one, and one-to-many interactor–character scenarios. Within AMITIES is another component that is of value during avatar-mediated interactions. This is called the activity storage/retrieval unit. This subcomponent supports trainers in the processes of tagging and commenting on events, subsequently using these to assist reflection on the part of users (trainees) and supporting detailed after-action reviews.

We start by providing context through discussions of the rationale behind the human-in-the-loop paradigm that forms the basis of the system. We then describe the individual components and the interfaces provided by our system architecture. As a part of these discussions, we also present some of our previous user interfaces for our inhabiters. Our participant and inhabiter interfaces are aimed at intuitiveness and low cost, while retaining the realism of the interaction required during critical personalized training and rehabilitation scenarios.

2 Background

Traditionally, two terms have been used to denote manifestations of virtual humans: avatars and agents. The distinction is based on the controlling entity, which could be either a human (avatar) or a computer algorithm (agent) (Bailenson & Blascovich, 2004). There is a rich set of literature comparing how the agency of a virtual character is perceived by human users (Nowak & Biocca, 2003; Garau, Slater, Pertaub, & Razzaque, 2005). In general, intelligent agents (Wooldridge & Jennings, 1995; Baylor, 2011) are very flexible as they can be replicated easily, can be used during any hour of the day, and are cost-effective human representations. Since avatars are directly controlled by humans, they rely less on the capabilities of the agent's artificial intelligence engine and can convincingly simulate social scenarios and adaptively steer conversations (Blascovich et al., 2002; Ahn, Fox, & Bailenson, 2012).

On the other hand, a recent metastudy comparing the effectiveness of agents and avatars (Fox et al., 2010) found that avatars elicit stronger levels of social influence compared to agents. Similar results were found in game environments (Lim & Reeves, 2010).

While having free-speech conversation with virtual characters is desirable in virtual environments, it is difficult to achieve this through intelligent agents without the use of certain methods that restrict a participant to limited responses (Qu, Brinkman, Wiggers, & Heynderickx, 2013). Due to the open-ended nature of conversations in bidirectional conversations in training and rehabilitation scenarios, our AMITIES system uses human-controlled avatars. This choice of human agency has been made by several systems in the past and has usually been referred to as digital puppetry. As defined in Sturman (1998), digital puppetry refers to the interactive control of virtual characters by humans. This paradigm has been successfully employed for decades in many fields including children's education (Revelle, 2003), games (Mazalek et al., 2009), and interactive networked simulations (Dieker, Lingnugaris-Kraft, Hynes, & Hughes, 2013).

Existing puppeteering systems often map the full range of captured human motion data to an avatar (e.g., Lee, Chai, Reitsma, Hodgins, & Pollard, 2002; Mazalek et al., 2011), but this approach requires specialized motion capture equipment, is prone to noise in the raw data, and requires a high-bandwidth connection to transmit the poses. H. J. Shin, Lee, S. Y. Shin, and Gleicher (2001) use Kalman filters and an analysis of the human's posture to process raw motion capture data in real time and map it to a puppet, but this method still requires a full motion capture system. In the system presented in this paper, the problem of full-body motion capture is circumvented by employing the concept of microposes (Mapes et al., 2011; Nagendran et al., 2012).

Other recent approaches to capturing the human user employ the Kinect system (e.g., Leite & Orvalho, 2011; and Held, Gupta, Curless, & Agrawala, 2012). There are also techniques that concentrate solely on capturing a human's face with high precision (Weise, Bouaziz, Li, & Pauly, 2011). Others have worked on the use of arbitrary control devices to control avatars through genetic programming (Gildfind, Gigante, & Al-Qaimari, 2000), and through collaborative control of virtual puppets (Bottoni et al., 2008).

It should be noted that the human-in-the-loop paradigm used in the presented system draws on parallels from the Wizard-Of-Oz (WOZ) technique (Kelley, 1984) by combining the traditional method with simple artificial intelligence routines that can be triggered by an inhabiter. WOZ is primarily used in the field of human–computer (Dow et al., 2005) and human–robot interaction (Riek, 2012) and refers to an experimental design in which users believe that a system is behaving autonomously, but behind the scenes it is actually operated to some degree by a human. This is noteworthy in this context, since participants' beliefs can be influenced by their expectations or preconceived notions (Nunez & Blake, 2003)—this concept is generally referred to as priming. Although the avatars in the presented AMITIES system are controlled by one or more interactors, we are not actively trying to deceive the user or trainee regarding the human agency; that is, no active priming is involved.

2.1 Challenge Areas

Using virtual characters and associated environments for applications such as training, rehabilitation, and practicing interpersonal skills has several associated challenges. One challenge area is related to the technology affordances of the system—this is one of the several subsets of challenges related to human factors issues in virtual environments (Stanney, Mourant, & Kennedy, 1998; Gross, Stanney, & Cohn, 2005); another challenge is related to the virtual character interaction paradigm, several of which currently exist (Faller, Müller-Putz, Schmalstieg, & Pfurtscheller, 2010; Semwal, Hightower, & Stansfield, 1998). For the experience to be effective, a user's beliefs about the validity of the scenario should be fostered, preserved, and reinforced. Explicit or implicit anomalies during bidirectional communication can result in breaking these beliefs. For instance, it is difficult for a traditional AI system controlling a virtual character to initiate a personalized conversation with a user that takes into account factors such as their attire (e.g., unique clothing or accessories) and relevant context such as items that are present in the interaction setting. Yet a conversation that is customized to include such personalized information can be a very powerful tool in influencing the beliefs (and hence behavior) of the user during the rest of the scenario. This is one of the primary advantages that a human-in-the-loop paradigm affords. In addition, the dynamic flexibility of the interactor-based control affords the opportunity to experiment with factors that influence interactions between virtual characters and users.

For a system that includes a human (interactor) in the loop, there are several specific challenges, including setting up a bidirectional architecture for data flow between the server (human) and client (virtual character); minimizing the utilized network bandwidth and latency while controlling virtual characters; maximizing the robustness to lost or erroneous data; and reducing the cognitive and physical demands on the interactor. The system presented here addresses these challenges, providing a smooth paradigm for virtual character control aimed at providing individualized experiences geared toward training, rehabilitation, and other applications where human interaction is critical.

3 System Description

AMITIES is a system architecture designed for mixed-reality environments that supports individualized experience creation such as education, training, and rehabilitation, and utilizes the marionette puppetry paradigm. The system has evolved over a period of six years with continuous refinements as a result of con-

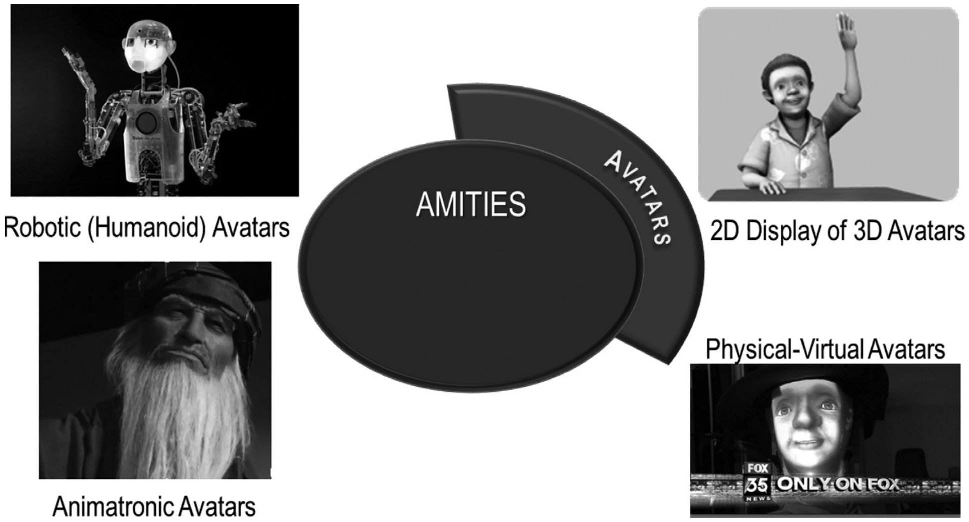


Figure 2. *Some examples of avatar manifestations, controllable by an inhabiter.*

stant use and evaluation. The system has the following features:

1. Custom digital puppetry paradigms (e.g., low-cost, low-demand, both physical and cognitive) interface for inhabiters that allows them to easily participate in the control of the verbal and nonverbal activities of a set of virtual characters;
2. A low-cost, unencumbered interface for users that allows them to employ natural movement and verbal/nonverbal interaction with virtual characters;
3. Seamlessly integrated autonomous behaviors that support one-to-one, one-to-many, many-to-one, and many-to-many avatar-based interactions;
4. A network protocol that supports real-time remote interaction even when dealing with relatively poor network connections; and
5. An integrated activity storage and retrieval system that supports trainers in the processes of tagging and commenting on events, subsequently using these to assist reflection on the part of users and to support detailed analysis via after-action reviews.

3.1 Avatars and Manifestations

We begin with a discussion of AMITIES and the interface it provides for controlling avatars. Avatars, as previously mentioned, are generally human-controlled virtual characters that may either be co-located or have remote presence at a distant location. These have varying degrees of complexity in traits such as appearance, shape, controllable degrees of freedom, and intelligence, among several others. These avatars are commonly seen as 2D representations of 3D avatars—in essence, these are virtual characters that are modeled and rigged by animators to be controlable in real time and are displayed on flat screen surfaces such as TV screens or projected onto viewing surfaces. The same avatar can appear differently, depending on the technology at the perceiving end. For instance, rendering the same avatar with compatibility for a pair of 3D viewing glasses (active/passive) will allow a participant to interact with a virtual 3D representation of this avatar. Similarly, the avatar may have a physical presence in a remote location—one such example is a physical-virtual avatar (Lincoln et al., 2011). These manifestations (physical/virtual) of the avatars can take several forms, a few of which are shown in Figure 2. Other examples of avatar manifestations could include complex robotic (humanoid) or animatronic figures as seen in Figure 2. Some of these avatars may be designed to appear very specific to a person (such as the animatronic avatar in Figure 2), while others offer the flexibility to change appearance. Specifically, the image on the top left

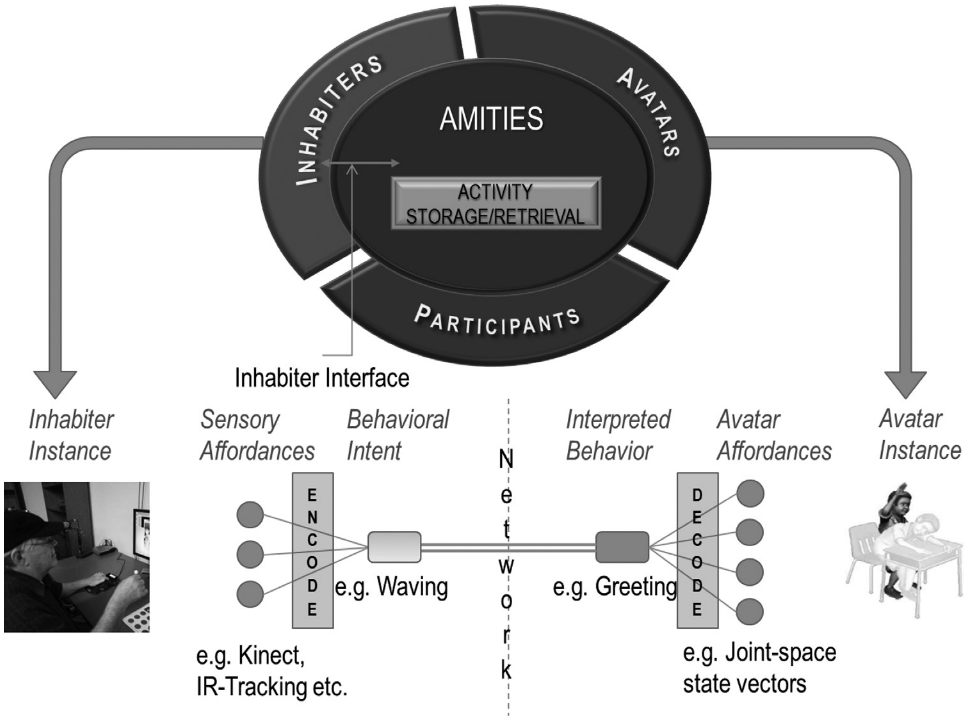


Figure 3. *The interface provided by AMITIES for inhabiters.*

portrays Robothespian, which is a humanoid robot with a rear-projected head for changing appearance, and pneumatic actuation (air muscles) combined with passively loaded elastic elements (springs) and electric motors. What is of importance to note is the requirement for controlling mechanical elements in such avatars. Similarly, the bottom-left image shows an animatronic avatar of a Man of Middle Eastern descent; the avatar's endoskeleton is pneumatically actuated for kinematically compelling gestures and fitted with a silicone-based exoskeleton or skin that deforms to convey realistic facial emotions. The manifestation is generally driven by the needs of the avatar-mediated interaction, where the desire for one trait of an avatar may outweigh the benefits offered by a generic, more flexible version of the same avatar. Similarly, avatar manifestation could vary in the complexity offered in the number of controllable degrees of freedom, the built-in semiautonomous behaviors, their shapes, and so on. For an inhabiter to control these manifestations effectively, the system interface must be opaque to the avatar's specific traits. AMITIES supports this opacity via a control paradigm that captures an inhabiter's intent, encodes it, and transmits it across the network to the avatar instance. The same system then decodes the received packet at the avatar instance and helps realize the desired behaviors on the specific avatar manifestation, including translating this message into the desired mechanical actuation (Nagendran et al., 2012) if required. This concept is further explained in Section 3.2, when the inhabiter's interface is described.

3.2 The Inhabiter Interface

AMITIES provides a multifunctional interface for people controlling their avatar counterparts; these people are referred to as inhabiters. Figure 3 illustrates the stages involved in the control of avatars. An inhabiter station consists of a location in which a person can be tracked via several sensors and perform actions using a wide variety of user-interface devices. The data from the devices and the sensors together form the *sensory affordances* of that particular *inhabiter instance*. Let us assume that the number of sensory affordances provided by an inhabiter instance is N. AMITIES is responsible for interpreting this data and encoding it into a single packet with sufficient information to capture an inhabiter's intent during avatar control; that is, the system processes the individual data streams for all sensors and devices to identify a *behavioral intent* for the inhabiter, such as waving. This constructed packet is then transmitted over the network to the remote location where the avatar resides. At the avatar's end, the information in this packet is interpreted to obtain the desired behavior that must be executed by the avatar instance. AMITIES then takes into account the number of affordances of that particular avatar instance (M) to decode the data into subcomponents required by the avatar, following which the avatar executes the *interpreted behavior* in real time. To illustrate, assume that the avatar instance is a physical-virtual avatar with mechatronic components that control the motion of its rear-projected head. The affordances of this avatar require roll, pitch, and yaw information for the head, and the animation weights (blend-shapes) required to create the facial expressions for the avatar. The received packet contains the behavioral intent information of the inhabiter—for the purpose of clarity, let us assume that this is encapsulated as *disagree*. The interpreted behavior at the avatar's end requires the avatar to execute the behavior *disagree*. The decoded components require the avatar's head to shake from side to side while the facial expression involves a frown and a raised eyebrow. AMITIES extracts this information from the received packet and pushes the velocity profiles (joint-space state vector) for yaw (shake head) to the avatar while also rendering the desired facial expressions via blended animations on the rear-projected head. This is a typical one-to-one mapping of avatar control supported by AMITIES.

In general, AMITIES is capable of aggregating *N sensory affordances* and mapping them onto *M avatar affordances* as required. The system utilizes the same architecture to support one-to-many, many-to-one and many-to-many avatar control. Additionally, AMITIES provides an interface at the inhabiter's end that allows for calibration routines of behavioral intent versus interpreted behavior. For example, an inhabiter can choose to have a specific behavioral intent be mapped onto any other interpreted behavior for each avatar instance as desired. This can be particularly useful when an inhabiter wants to reduce the physical and cognitive demands placed on him or her during multi-avatar control. As an example, a simple behavioral intent such as waving at the inhabiter's end can be mapped onto a more complex interpreted behavior such as standing up, bowing, and greeting at the avatar's end. We should note that in addition to directly controlling an avatar, the inhabiter can also trigger genres of behaviors for individual avatars or groups of avatars in the virtual environment. For instance, an inhabiter can cause an entire virtual classroom consisting of several avatars to exhibit unruly behaviors or limit these behaviors to individual avatars.

3.3 The Participant Interface: Users and Observers

AMITIES classifies participants into two categories, depending on their interaction capabilities with the avatars. The first category is the user/subject (referred to as the *participant-user*) who is an active participant in the avatar-mediated interaction. This participant is directly involved in bidirectional conversations and actively engages in behaviors with the avatar. AMITIES provides an interface that complements the avatar mediation by creating and maintaining the user's beliefs via sensing technology at this end, as shown in Figure 4. For instance, the technology allows a user to be immersed in the environment in which the avatar-mediated interaction is occurring by tracking their motion and correspondingly adjusting the system's response to this motion. Examples include altering the viewpoint of virtual cameras in synchrony with a user's movement to give the user a sense of immersion in virtual environments or autonomously altering an avatar's gaze to look at the user as he or she moves around in the interaction space. Eye gaze has been shown to be an important factor in determining the perceived quality of communication in immersive environments (Garau et al., 2003). Additionally, AMITIES captures and transmits bidirectional audio streams to allow conversations between the participant and the avatar (which is

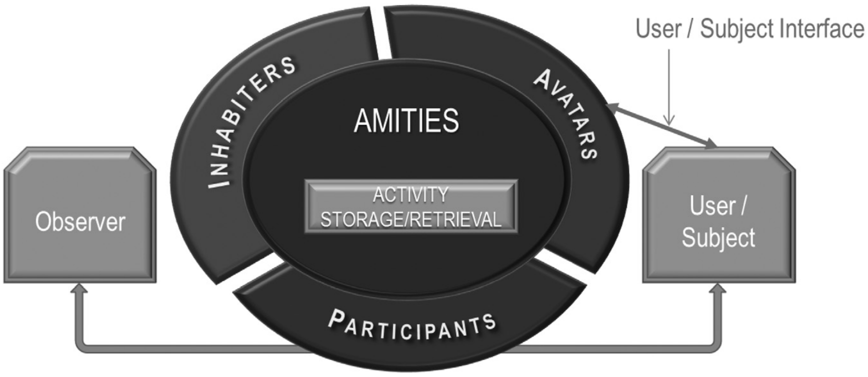


Figure 4. *The AMITIES interface that supports mediated interaction between avatars and users/subjects.*

controlled by its inhabiter). Selective video-streaming capabilities are also offered by the AMITIES interface at this end, allowing an inhabiter to view the user and the remote environment during interactions. While the system supports bidirectional video, this stream from an inhabiter is traditionally not required, since the avatar is the focal point of the interaction for a user. This could be for a variety of reasons, including maintaining anonymity, preventing bias, and masking the actions of an inhabiter during avatar control. A special instance of this case is when an inhabiter chooses to use his or her own video stream to alter the appearance of the avatar so that it resembles him or her. In this case, care must be taken to prevent broadcasting the environment of the inhabiter, since viewing such an environment during the interaction could destroy the belief of situational plausibility as a result of viewing two environments simultaneously, one in which the user is currently located, and the other in which the inhabiter is located. Currently, this is accomplished in AMITIES by using a monochrome background behind the inhabiter that naturally frames his or her face.

The second category of participants is referred to as *participant-observers.* These are participants who are passive and do not directly affect the avatar-mediated interactions. AMITIES provides an interface that does not include sensor technology to track the movements and behaviors of these participants, as shown in Figure 5. This interface allows participant-observers to interact with either the inhabiters or the participant-users. Observers include Subject Matter Experts (SMEs) who can view and influence the interactions indirectly in real time using an audio-uplink to either the inhabiter or the participant-user, depending on the particular application. Other observers may include trainees or simply bystanders who wish to witness the interaction with a view to gathering information. For the purposes of maintaining anonymity and privacy, observer stations are selectively permitted to view the user (trainee), but can hear and see the entire scene that includes the avatars and their environments, allowing observers to gather the gist of the interaction in complete detail. This is accomplished in AMITIES via remote video and audio feeds that are broadcast over the entire system so that all components receive them.

3.4 Activity Storage and Retrieval Module

The Activity Storage and Retrieval (ASR) module is embedded with the AMITIES architecture as shown in Figure 6. The purpose of this module is to record all activity during the avatar-mediated interactions in order to provide both real-time and post-interaction analysis and feedback to participants. To support this, all interface components have read-write access to the activity storage and retrieval module. The module handles the collation of all data streams, including sensor-based data, video data streams, audio data streams, raw control device readings, semiautonomous behaviors, and other

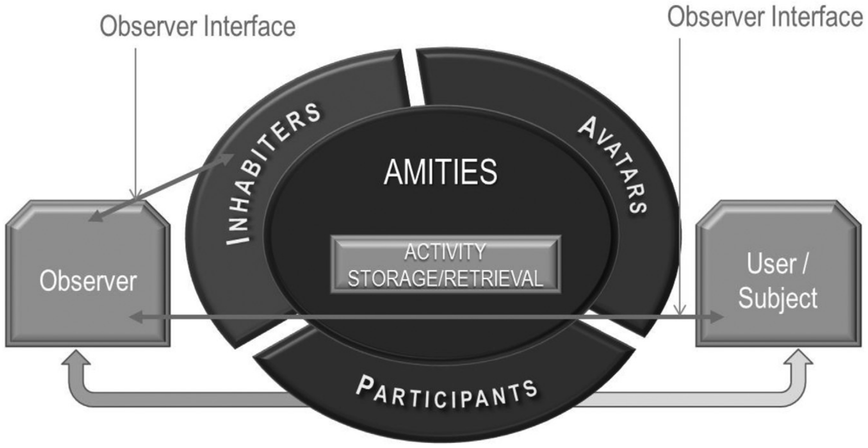


Figure 5. *The interface provided by AMITIES for observers connects them to users as well as inhabiters.*

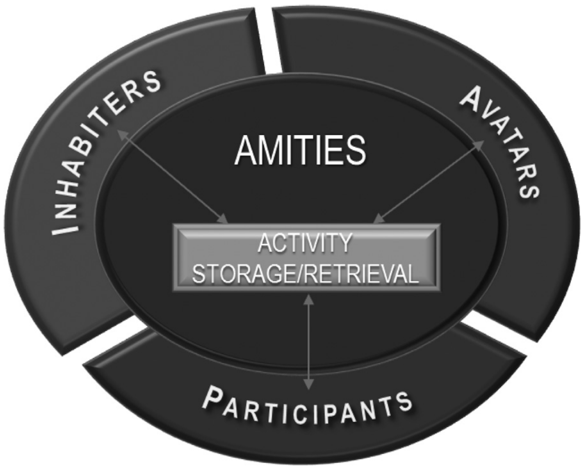


Figure 6. *The activity storage and retrieval module collates data from inhabiters, avatars and participants.*

related information using synchronized time stamps. The ASR module supports live review, after-action review, analysis via visualization tools, and recording and playback of avatar behaviors. In addition, the ASR module logs the avatar's behaviors, allowing a researcher to review a participant's response to specific behaviors. As an example, the visualization tool uses the ASR module's time-stamped audio and video streams to allow a reviewer to step through a section of the interaction while viewing a user's body language during the segment. A quantitative estimate of a user's body motion is obtained via the sensor-based data, allowing a reviewer to analyze the movements of the subject in detail with respect to an avatar's behavior. At the same time, verbal responses during this segment can be analyzed to find statistical measures such as reciprocal response times, initiated response times and so on. An example of using this module for after-action review is shown in Section 5.6.

4 The Scenario Design Process for Using AMITIES

AMITIES provides a flexible framework for controlling expressive, avatar-mediated, human-to-human communication. However, it does not inherently define character personalities or experiences—that exercise is left to the designers, and is usually carried out on a case-by-case basis. Below, we first describe the character and story design process that we have developed, and then we describe some particular cases for which we used this process and the AMITIES framework to create an overall experience for a particular purpose.

4.1 Character and Story Design

The AMITIES framework involves a process for the iterative design and development of the appearance and behaviors of virtual characters, and the context in which these characters operate. This involves artists, SMEs, programmers, and, most importantly, the requirements of users of the resulting system. Components of this include model design and creation, and verbal and nonverbal behavior selection and implementation (puppeteered and automated).

The design process starts with a requirements specification document that identifies the key goals of the interaction—this could be for education or a more intense training scenario such as a mission debrief. Inhabiters then rehearse their avatars' (including the trainee's) behaviors (verbal and nonverbal) using a role-playing approach designed to flesh out the characters' back stories and interaction styles. This involves video and audio recordings of the entire process. Note that this does not result in a traditional script, but rather a mix of story elements, branching logic (roadmaps to important events), and motivations for each character. Individual stages of these role-playing sessions are used for analysis and eventually utilized by the artist(s) and programmer(s).

These are just initial steps to establish the artistic/technical requirements. We then produce concept art. Care is taken to ensure that the demographics and appearances of the avatar are well-suited to and representative of the scenario being created. Once these artistic designs are reviewed and accepted and the role-playing is deemed to have uncovered the collection of required gestures (facial and body), the artists proceed to model development, texturing, and rigging of the characters. This involves identifying key frames (microposes) that support specific behaviors (those uncovered in rehearsal) as well as a broad range of dynamically created behaviors so an inhabiter can react to widely varying interactions with users. Additionally, the artist/inhabiter/programmer team develops animations of specific behaviors such as *annoy others*, *look attentive*, or *act bored*, and create finite state machines that support transitions between scenes, as needed. This results in an operational set of puppets and scenes. With this nearly final product in-hand, the inhabiters perform role-playing rehearsals again, but this time using the AMITIES system, with test participants and observers. The outcome of this process is then a set of characters, microposes, scenes, animations and decision trees that enable the final avatar-mediated interaction experiences.

5 Case Study: TeachLivE—An AMITIES Instance

The plasticity of AMITIES allows a wide range of applications as evidenced by existing projects involving teacher education (Dieker et al., 2013), cross-cultural communication (Lopez, Hughes, Mapes, & Dieker, 2012), interviewing skills for employers and supervisors, protective strategies regarding peer pressure for childen and young adults (Wirth, Norris, Mapes, Ingraham, & Moshell, 2011), debriefing skills training for instructors, client interaction skills for charitable foundation employees, and communication skills development for young adults with autism. Here we describe a specific instance where we applied the above design processes and the AMITIES framework for a particular application.

As shown in Figure 7, AMITIES is the foundation for the TLE TeachLivE Lab, which includes a set of pedagogies, content, and processes, created as an environment for teacher preparation. The environment delivers an avatar-based simulation intended to enhance teacher development in targeted skills. Teachers have the opportunity to experiment with new teaching ideas in the TLE TeachLivE Lab without presenting any danger to the learning of real students in a classroom. Moreover, if a teacher has a bad session, he or she can reenter the virtual classroom to teach the same students the same concepts or skills. Beyond training technical teaching skills, the system helps teachers identify issues such as common misconceptions, for example, in algebra skills, so these can be mitigated, and latent biases, so the teachers can develop practices that mitigate the influence of these biases in their teaching practices. The ability of the system to track movement and time spent with individual students is a great benefit of this program, as it provides objective measures for the teacher and trainer to use during reflection and after-action review.

The TLE TeachLivE Lab has been an ongoing project since 2009, with efforts ramping up in 2012–2014

Figure 7. *Virtual class of five students who can be controlled by an interactor.*

Table 1. *Statistics and Outreach of the TLE TeachLivE Lab*

Number of universities enrolled	42 across the United States
Number of universities in pipeline	About 20 more in the United States
Total teachers that have trained using the system	Nearly 10,000
Sessions and durations	Four Sessions @ 10 min per session
Effective impact and outreach	Nearly 1,000,000 students

with support from the Bill & Melinda Gates Foundation. Table 1 shows the outreach and statistics of the program. Data analysis is currently underway and considered preliminary until approved for release by the funding agencies.

5.1 AMITIES Framework Components in TeachLivE

Figure 8 shows the components of the AMITIES framework instantiated in TeachLivE. The inhabiter is typically referred to as an interactor in the TeachLivE system. These are individuals trained in improvisation, interactive performance, and story development (Erbiceanu, Mapes, & Hughes, 2014), who, with the aid of agent-based (programmatically determined) behaviors, control the avatars in the classroom. A single interactor controls multiple avatars by using the framework's ability to seamlessly switch between avatars while retaining behavioral realism in the avatars that are not directly inhabited. The interactors modulate their voices and behavioral intent in accordance with their avatars and appear completely opaque to a subject interacting with the avatars in the classroom.

The TeachLivE virtual classroom typically consists of five avatars, as seen in Figure 7. Each of these characters has a back story and certain behavioral traits that are unique. The interactor is trained to adhere to these traits during the classroom interaction. For instance, one of the students is very quiet, low-key, intelligent, and not desirous of attention (passive, independent); while another student is very talkative, inquisitive, responsive, and in constant need of attention (aggressive, dependent). The avatars also have built-in autonomous behaviors that can be modulated by the interactor and are capable of exhibiting combinations of group behaviors such as laughing in tandem, or whispering to each other. These group behaviors can be triggered by an

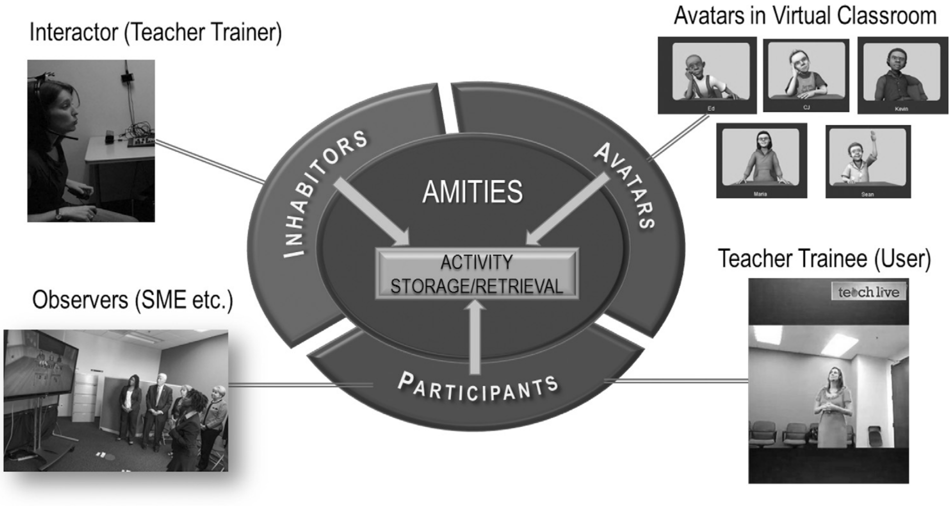


Figure 8. *The AMITIES instance TeachLivE showing an interactor (inhabiter), student avatars, a teacher trainee (participant-user) and SMEs (participant-observers).*

interactor, and will occur on all avatars except the one that the interactor is currently inhabiting to create a realistic classroom environment.

The participant-user/subject is either a teacher trainee (preservice) or an experienced teacher seeking new skills (in-service) whose role during a session is to apply good pedagogy, convey subject-related information, and to manage behaviors in the classroom. The trainees are allowed to experience varying levels of difficulty, all of which are predetermined via discussions between their parent universities or supervisors and SMEs. The difficulty manifests via avatar mediation.

Participant-observers may include bystanders, coders, SMEs, and other trainees who may have already completed the sessions, since we do not want to bias a new trainee by exposing him or her to another trainee's classroom experience.

5.2 System Architecture of TeachLivE

As described previously, the teacher training environment consists of several students (digital avatars) in a virtual classroom, whose space is shared with the real world. Figure 9 shows the architecture of the system, with the individual components and the data flow between them explained in detail in the following section. The illustration is best understood when perceived in the following order. Starting with the inhabiter experience, follow the data streams (control data, audio uplink) to the participant-user, then look at the participant-user experience and follow the data streams (audio/video uplink) back to the inhabiters and the participant-observers, and finish by looking at the data flow between the inhabiters and the observers.

The current AMITIES framework consists of a server-client model that supports bidirectional communication. The server controls the avatars and the camera if necessary (manual camera control). The client displays the scene and allows interaction with it via the virtual camera and an audio interface. The audio interface is responsible for all conversations between the avatars (interactor-controlled) and trainee during the session. The interactor (server) also receives a video feed of the trainee, allowing him or her to assess body language and other nonverbal cues. At the server end, the interactor's intentions (motions) are captured via two independent motion capture systems. Devices that can be used interchangeably for this purpose include infrared cameras, Microsoft Kinect, Razer Hydra, and keypads. This is mapped onto the avatars via a custom animation blending system. The movement of the trainee (at the client)

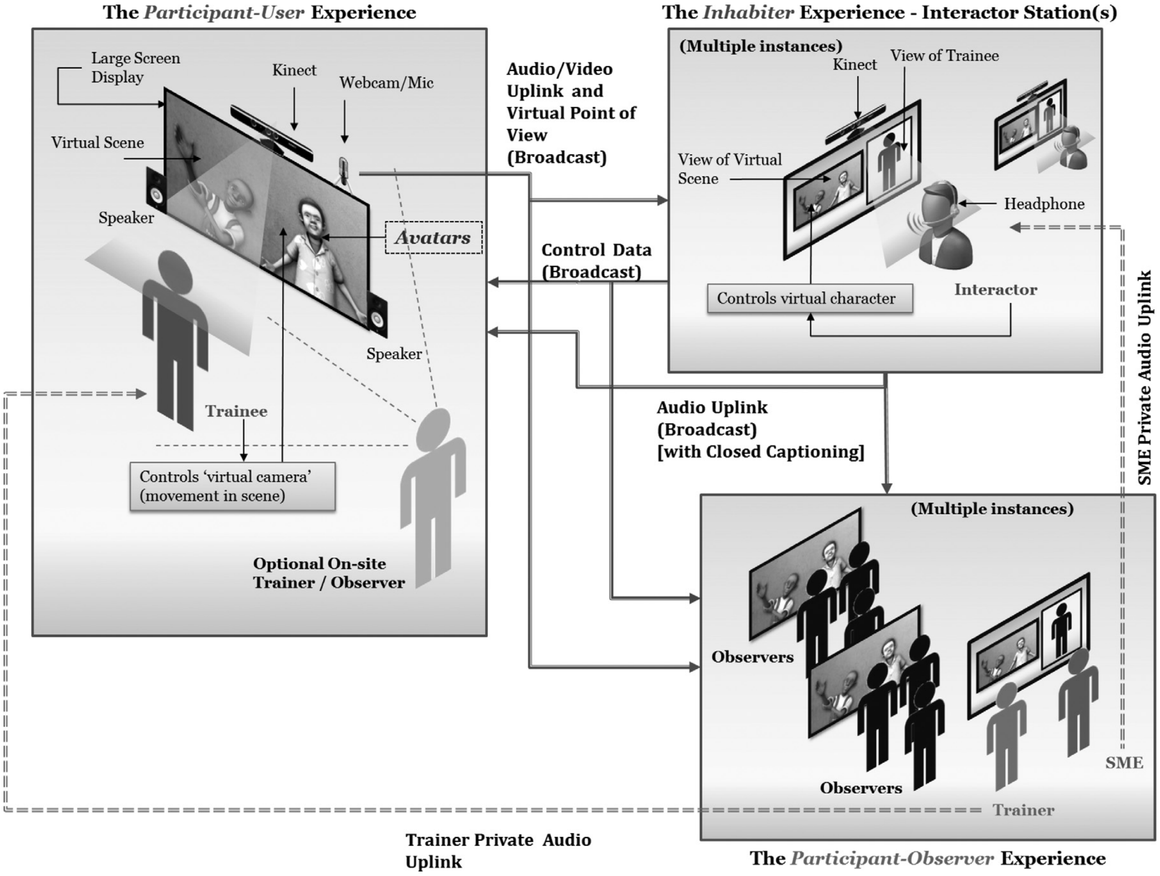


Figure 9. *The complete system showing different AMITIES components (inhabiters, avatars and participants) and the data flow between them. The acronym SME is used to indicate a Subject Matter Expert in the figure.*

in the interaction space controls the camera view during the process. This allows the teacher (trainee) to walk up to specific students in the environment, bend down to achieve eye-to-eye contact, and initiate a focused conversation. This camera view is seen by the interactor, allowing him or her to determine the character that is in focus. In the following sections, we describe each one of these interfaces in detail.

5.3 The Inhabiter Experience—Interactor Station(s)

Central to the AMITIES framework is the concept of the WOZ technique—this is personified by the inhabiter who is responsible for avatar control during the mediated interactions. Inhabiters require a control paradigm that can be used to modulate their avatars' behaviors. The AMITIES control paradigm has evolved from a very literal system based on motion-capture to a gestural one based on a marionette paradigm (Mapes et al., 2011). Common to all the paradigms we have implemented is support for switching avatars, and triggering agent-based behaviors for those characters not presently under direct control. In effect, there can be many virtual characters, with the current avatar being completely controlled by an interactor, and all others exhibiting agent-based behaviors that are influenced by the actions of the interactor, the current avatar, and the user. In this section, we highlight the evolution of some of these control paradigms at an interactor station—that is, the remote location from which a student avatar in the virtual classroom is being inhabited.

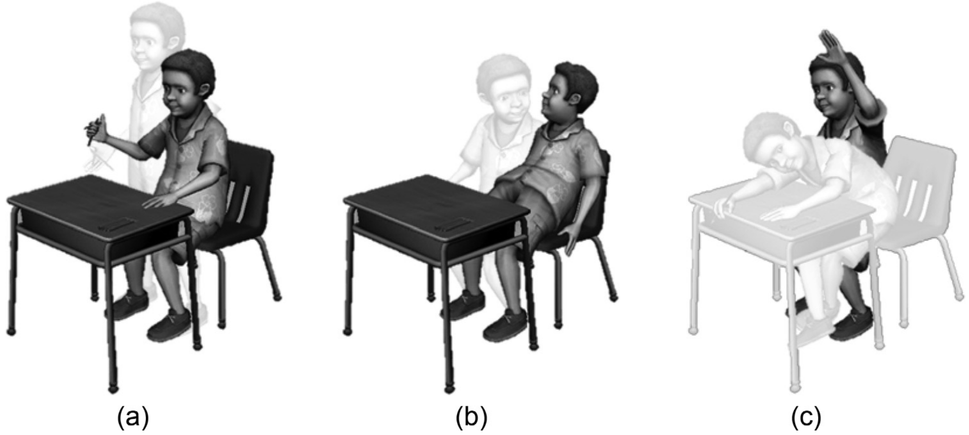


Figure 10. *Microposes for a virtual avatar named Sean. (a) Sean is standing (translucent) and holding a pen (solid). (b) Sean is leaning forward and turning (translucent) and slouching (solid). (c) Sean is laying on the desk (translucent) and raising his hand (solid).*

5.3.1 Previous Interactor User Interface Paradigms. Historically, we explored several user interface (UI) paradigms to allow the interactors to control the virtual characters. Our first approach, motion capture, had noise problems typically experienced with this approach, but without the opportunity for postproduction, as all actions had to take effect in real time. Moreover, with capture frequencies of 120 Hz, we were transmitting a substantial amount of network data, with attendant issues when communicating with clients who had poor connectivity.

To address the problems introduced above, a number of variants of the paradigm were developed, investigating each one in the context of its effect on noise, network traffic, the quality of the experience at the receiver end, and the cognitive and physical demands reported by interactors. The first and, we feel, most critical decision, was to develop the notion of *microposes*. Microposes are components that make a pose. In some cases, they are the only observed final poses, as we do not perform pose blending. However, when we do perform pose blending, we rarely render a micropose; rather, we render a blend of microposes to create a pose. In a very real sense, microposes are basis sets for the poses that an avatar is expected to perform from which all rendered poses are formed using linear coefficients (blending weights). Some of these microposes are shown superimposed on each other to view the motion-space of the avatar in Figure 10.

After we developed the concept of microposes, we experimented with a series of gestural schemes to control the selection of these microposes.

When the Kinect for Xbox 360 was released in November, 2010, we briefly went back to using a literal mode of controlling avatars. The problem with a purely literal approach is that it makes it hard to implement some desired behaviors, such as having the avatars place their heads on a desk, as we often want to do when using the system in teacher training.

Having the interactors place their head on the table would make it very hard for them to keep track of what is happening at the other end, as the video-viewing window is hard to see from that position. Other actions such as standing up and clicking a pen are more natural to trigger by gestural, rather than literal, movements (see Figure 11). For these reasons, we returned to gestural schemes as soon as we became aware of the capabilities of the Razer Hydra in the spring of 2011.

5.3.2 Current Interactor User Interface Paradigm. Figure 12 shows the system architecture at the interactor station (*Inhabiter experience*). The

Figure 11. *Puppeteer controlling students in virtual classroom, using head tracking and Hydra. Note the window with the video feed (on the right-hand side of the monitor) that allows the interactor to observe a user's nonverbal behaviors.*

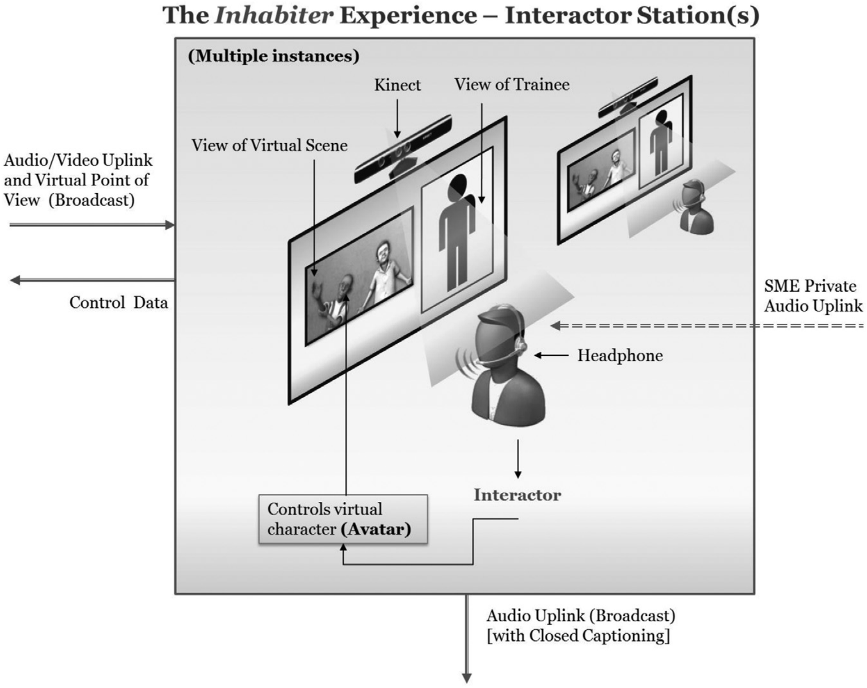


Figure 12. *The interactor station (extracted from the top right-hand side of Figure 9).*

current interactor UI paradigm supports spawning of multiple instances of the interactor station. This allows several interactors to simultaneously control the virtual characters in a scene. Our paradigm can even support multiple interactors controlling a single avatar, a feature we use in remote training of new interactors (think of one as a driving instructor and the other as the driver). In all cases, the interactor is seated in front of a large-screen display (or multiple screens if preferred), and can view the scene as well as a video feed of the remote location (where the user is located). Control of the virtual character can occur via one of several mechanisms listed above. This control data, along with an audio uplink, is broadcast to the user as well as to any observers that are in the system. Video from the user is received at the interactor station; but no video uplink of the interactor is provided to either observers or users. This helps keep the interaction paradigm "behind closed doors," to promote situational plausibility and belief (WOZ effect). An SME has a private audio uplink to the interactor, allowing him or her to prompt appropriate responses to complicated situations as required. A trainer can have have private audio uplinks to the user (training instructions) and the interactor (desired scenario branching).

In the current system, we use the Razer Hydra (Razer, 2013). This device uses a magnetic field to detect absolute position and orientation of two handheld controllers. So long as the controllers are in front of the magnetic sensor and within a six-foot radius, the device operates with a reasonably accurate precision of 1 mm and 1°. Each controller has five digital buttons, one analog stick/button, one bumper button and one analog trigger. We use the left controller for character selection, zooming, and mouth movement; we use the right controller for agent behaviors and facial gestures. These buttons can be configured to trigger situation-specific reactions and appearance-related features of a virtual character, such as frowning, smiling, and winking. They can also trigger group and individual agent-based genres of behaviors. As with all our micropose-based paradigms, we have a library of poses unique to each virtual character.

The precise mapping of an interactor's gesture to character pose can be personalized by each interactor based on what he or she feels is cognitively easiest to remember and places a minimum of physical demands on the interactor. This particular approach appears to provide the best balance between a high level of expressiveness and a low level of cognitive and physical requirements on the interactor. The decoupling of gesture from pose also allows us to localize the rendering at the user side in a manner that is appropriate to regional customs.

5.3.3 Microposes and Avatar Control. Control of the current avatar's pose is done by gestures that are mapped to microposes, with variations in those gestures coming from being close to several poses, and by twisting the controllers to get subtle deviations (see Figure 11). This is explained in more detail below.

The current activated virtual character is controlled using a micropose system with the Razer Hydra controller's 3D position and orientation input across two handheld controllers. Every micropose is configured with a user-specified pair of 3D coordinates, one for each controller (recorded via a calibration phase using the Razer Hydra). During runtime, the system then attempts to match the current position of the controllers with the predefined configurations to animate the puppets.

The system supports three modes: best match, two-pose cross-fade, and High Definition (HD) poses. Best match simply selects the pose that best matches the input coordinates. The two-pose cross-fade system selects the two poses with the shortest Euclidean distance from the input, and then calculates the animation blend between them, allowing for an interpolated pose that is the weighted combination of the two selected poses. If the selected pose is not currently active, the system begins to transition into the new pose while transitioning out of the previous active one. The rate of transition into and out of poses is customizable, allowing for longer animation between transitions as necessary.

The third pose mode is the HD poses system, which works by applying inverse distance weighting across all available poses with respect to the input coordinates to find an appropriate mixture of all poses in the system. Animating the poses in this mode is a direct mapping based on the mixtures and the movement speed of the user, without consideration of individual animation transition rates. This allows for a more natural and fluid motion between poses, giving the interactor more fine-grained and direct control depending on the initial pose configurations and movement speed.

Each pose in the system provides additional levels of control between three animation key frames. Control of the position within the animation itself is handled

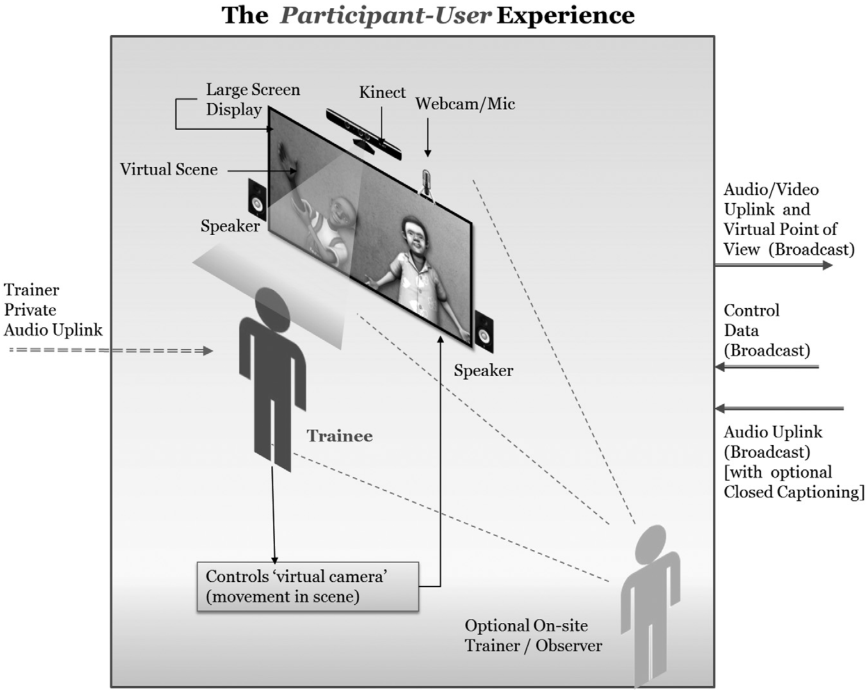


Figure 13. *The user experience (extracted from the left-hand side of Figure 9).*

by rotating the controllers about the longest side. This translates into a simple rotation of the hand, allowing for ease of use and fine-grained control, while still providing access to the other buttons. The system computes the sum of rotation of each controller and generates a rotation angle that is bounded by a configurable maximum and minimum angle. This value is then normalized such that it can be used to interpolate between the different key frames of the active animation or animation mixture. The final result translates rotational motion of the two controllers into fine-grained control of the active animation or an active animation mixture depending on the current micropose mode.

The avatars' facial expressions are controlled with the Razer Hydra's analog joystick input. This input provides a pair of values indicating the joystick's horizontal and vertical position, which is interpreted as a single angle value along a circle around the maximum extent of the joystick's range of motion. For example, if the analog joystick is pushed to the far right, this pair of values is interpreted as an angle of 0° degrees. Using this abstraction, all of the possible face morphs of the virtual character are mapped to angular arcs around the perimeter of the joystick's range of motion. The facial expression mapping is customizable to group similar facial expressions together in order to allow smooth transitions between expressions that are related. At runtime, the system simply interprets the analog joystick's position as an angle and then selects the facial expression whose predefined angular arc mapping matches the input. Once a new face morph has been selected, the system begins transitioning into the new pose and out of the previous one using customizable transition or ramp rates.

Equipped with this interface, the interactors control multiple avatars and their behaviors during the interactions to create realistic and responsive behaviors within the given virtual environment.

5.4 The Participant-User Experience

Figure 13 illustrates a teacher trainee's (participant-user's) experience. The trainees are typically located at a remote site and stand in front of a large

Figure 14. *User experiencing TeachLivE virtual classroom.*

display on which the virtual classroom is visible. Their movement is tracked by a Microsoft Kinect for Xbox 360. Where appropriate, their arms and head are tracked via a VICON IR Tracking System that features 10 T-40S imaging sensors—note that this is not employed in TeachLivE as it would negatively affect the desired scalability of that system. At present, the trainee's eye orientation is not tracked, although this is observable by the interactor through a live video feed via a webcam. Movement of the user toward the display results in a corresponding movement of the virtual camera through the scene's space (see Figure 14).

In our classroom environment, the students' eyes automatically follow the teacher, unless the student is tagged as exhibiting autistic behavior or attention deficit. We previously produced a short video demonstrating the use of the AMITIES system with TeachLivE in training a middle school teacher for a math lesson (SREAL, 2013b).

5.5 The Participant-Observer Experience

The system architecture of the observer stations involving a participant-observer is shown in Figure 15. For the purposes of maintaining anonymity and privacy, observer stations are not permitted to view the user (trainee), but can hear and see the entire visual scene, allowing them to gather the gist of the interaction in complete detail. This includes receiving the control data that is broadcast by the interactor station. Private audio uplinks are provided to SMEs and trainers, allowing them to interact either with the interactor or the trainee (when appropriate), in order to inject their specialized opinions. The SMEs and trainers can be thought of as special observers who also have the option of viewing the trainee (driven by a situational and study-approved need) if the latter requests/permits this. Several instances of the observer station can be simultaneously generated, thereby supporting interaction from remote locations.

5.6 Activity Storage and Retrieval Module for After-Action Review

The TeachLivE system also utilizes the Activity Storage and Retrieval (ASR) module for recording live sessions. This supports coding of events during and after these sessions. Event types can be created based on a given scenario's needs and used to home in on sets of frames in which these behaviors are observed during a training session. For example, in teacher practice, a coder tags frames in which the user asks high-order questions (a positive attribute), or in which very little

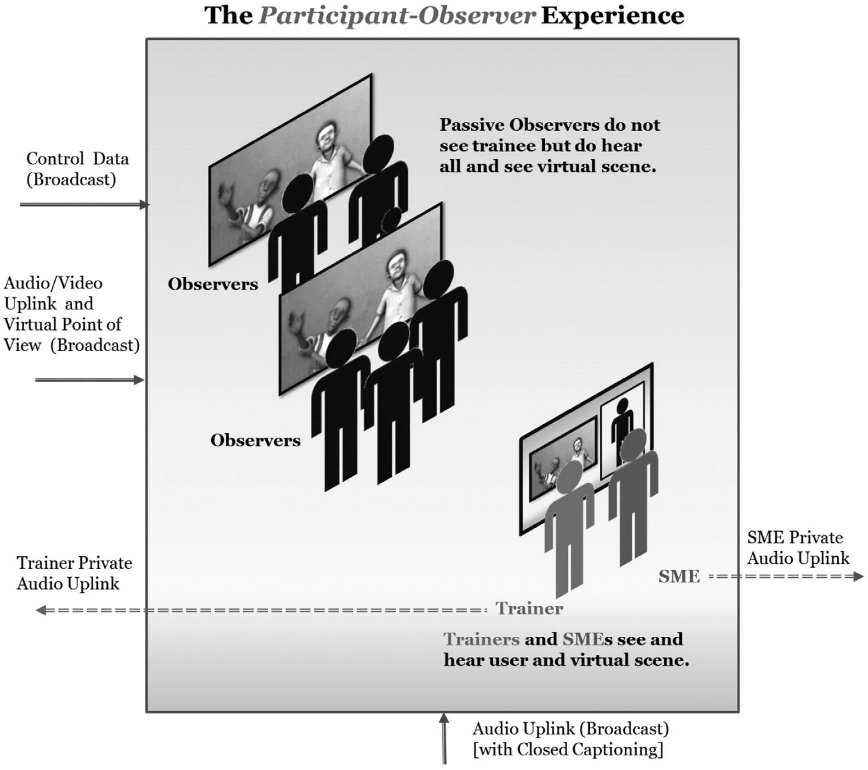
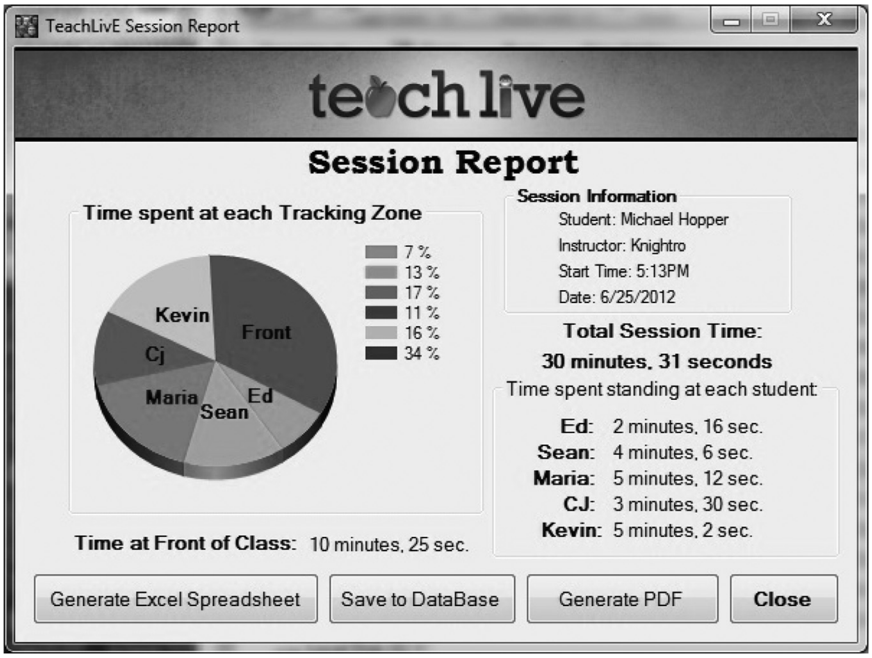


Figure 15. *The observer station (extracted from the bottom right-hand side of Figure 9).*

time is allowed to pass before the teacher goes on to another question (a negative attribute). Data recorded during a session can be exported to Comma Separated Values (CSV) files for entry into databases and spreadsheets. All decisions about when to record such events must be initiated at the receiver (client) end, where confidentiality and appropriateness of recording and coding is best made—the interactor has no integrated facilities to initiate such recording and event logging. Such a capability facilitates after-action review, reflection, and documentation of a user's progress, while following good practices of informed consent and confidentiality. This same feature can also be used to seed new automated behaviors, since the codes provide semantic labeling of user actions (Erbiceanu et al., 2014). At the end of a training session, performance statistics are reported by the system. This includes quantitative measures such as "time spent in front of each student" and "conversational times" obtained via real-time tagging (see Figure 16).

6 Other Instantiations of AMITIES

AMITIES also supports the control of Physical-Virtual Avatars (PVAs; Lincoln et al., 2011)—avatars that have physical manifestations—and the associated robotic components. While this may not appear particularly relevant to the topic of this paper, it is important to note the flexibility of the use of this system to multiple modalities: the system supports the control of virtual characters on a 2D screen, a head-worn display, as well as physical manifestations of the same character that involves mechanical components (robotic) on, for instance, a PVA. We also produced a video (screen capture shown in Figure 17) of the paradigm being used to control a virtual character manifested as a PVA and three virtual characters being controlled in a classroom setting, engaged in a conversation with a human (SREAL, 2013a). In particular, for this demonstration, one interactor controls the PVA and another controls all the virtual characters in the scene (Section 5.3),



TeachLivE Session Report

Session Report

Time spent at each Tracking Zone

Session Information

Student: Michael Hopper
Instructor: Knightro
Start Time: 5:13PM
Date: 6/25/2012

Total Session Time:
30 minutes, 31 seconds

Time spent standing at each student:

Ed: 2 minutes, 16 sec.
Sean: 4 minutes, 6 sec.
Maria: 5 minutes, 12 sec.
CJ: 3 minutes, 30 sec.
Kevin: 5 minutes, 2 sec.

Time at Front of Class: 10 minutes, 25 sec.

Figure 16. *Example performance statistics presented to a teacher trainee after a session in TeachLivE.*

Figure 17. *A screen capture of the submission video that shows the virtual characters on the 2D screen, the PVA and a human engaged in a conversation.*

while the PVA and the 2D flat-screen display provide the user experience (Section 5.4). The video showcases the interactor's interface (display and controls), the user experience (multiple modalities of virtual character display), and the natural-flowing conversation between all the users (virtual characters and the human) which is difficult to achieve with traditional AI-based control. It should be noted that a single

Figure 18. *A screen capture of debriefing session—virtual only.*

interactor can control characters having different manifestations, such as some PVAs and some purely virtual avatars.

AMITIES has also been used in a proof-of-concept with members of the Veterans Health Administration (VHA) Simulation Learning, Education and Research Network (SimLEARN). Our collaboration is in support of their mandate to train trainers who then go back to their home hospitals or clinics with improved skills. All such training focuses on team communication as well as technical skills, using simulated scenarios. Experience has shown that the most volatile skills are those associated with the debriefing process that follows each scenario. The AMITIES framework was used to recreate the standard situation that a trainer faces—a conference room populated by team members who just experienced the simulated scenario (Figure 18 shows a snapshot of this environment from a user's perspective). These simulations can include a wide variety of professionals, such as nurses, ER physicians, surgeons, and anesthesiologists. Hierarchies may already have been established and conflicting opinions about the value of simulations may already exist. Moreover, the actual events of the scenario may have led to tension among team members. The job of an effective training is to debrief with good judgment, a process described in Rudolph et al. (2007). The goal of the VA scenario we developed on top of AMITIES is to allow trainers at distributed sites to practice these skills, reflect on their performance, and have the option of being observed by SMEs in order to receive constructive feedback. We also produced a short edited video of the scenario with a participant-user interacting with the avatars (SREAL, 2013c).

We have used the AMITIES framework in an exploratory study aimed at the use of virtual characters to help prepare teens with autism and/or intellectual delays for their first job or college interviews. The subjects were exposed to three conditions in a repeated measures counterbalance design: (1) face-to-face with a human; (2) face-to-face with a virtual character on a flat-screen 2D display surface; and (3) face-to-face with a physical manifestation of the virtual character (a PVA). The scenarios and virtual characters were developed to facilitate a 10-min conversation with the subject, while several dependent variables were measured. The level of engagement was measured by analyzing several metrics, such as the frequency of initiated and reciprocal responses, latency of response times, and duration of the responses during the entire interaction. The results indicated that all participants had more engaging conversations, and interacted better, with the virtual characters than with the human. Although that result may not be surprising in itself, the significance comes in the willingness of the participants to initiate, and not just reciprocate, conversation when in the presence of purely virtual avatars.

Finally, we are using AMITIES as the underlying framework for a multiyear effort to explore various psychological and computational aspects of human–virtual-human social interaction. We will be examining the beliefs, behaviors, physiology, thoughts, and trust of human users/subjects when interacting with virtual humans in controlled scenarios. By substituting *real* intelligence (a real human) for the more common *artificial intelligence*, we hope to isolate other psychological and computational aspects of human–virtual-human social interaction such as the effects of shape, proxemics, kinesics, and other behaviors.

7 Conclusions and Future Work

In this paper, we have presented a framework for controlling virtual characters/avatars in remote environments. Critical components of the framework have been identified and their roles in enhancing individualized avatar-based interactions have been highlighted. The framework includes an activity storage and retrieval system to record all avatar-based interactions, allowing participants to reflect on their performance. The system lends itself to control of character manifestations ranging from purely virtual (e.g., a 2D display) to physical (e.g., a PVA). The associated architecture employs animation-blending routines, network communication with multiple server-client models, human-in-the-loop communication, and a control protocol that exhibits low latency, functioning effectively while using minimal network bandwidth. The resulting system is flexible enough to support personalized avatar-mediated experiences in applications including education, training, rehabilitation, and remote presence. We have successfully used it for several such experimental scenarios, each demonstrating natural interactions between people and their avatar counterparts.

In the future, we plan to improve the level of integration and automation of after-action review that is built into the system. Additionally, we want to develop several tools that support the analysis of the interactions, both online and offline, and use this data to alter the behavioral traits of the virtual characters during the interactions. This involves analyzing video data streams, tracking data streams, and processing audio data during the interactions.

Acknowledgments

The material presented in this publication is based on work supported by the Office of Naval Research Code 30 (N00014-12-1-0052 and N00014-12-1-1003), the National Science Foundation (CNS1051067, IIS1116615) and the Bill & Melinda Gates Foundation. This project also received partial support from the Veterans Health Administration (SimLEARN). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsors. In addition to directly cited authors, we would like to acknowledge the invaluable contributions of several members of the AMITIES team. In particular, we would like to thank (in no particular order): (1) our team of artists, Eric Imperiale (character modeling lead), Chip Lundell, and Alex Zelenin (environment modeling lead) for the creation of the avatars and virtual environments; (2) our team of software, audio, and hardware developers, Darin Hughes, Brian Kelly, Kangsoo Kim, Myungho Lee, and Greg Wiatroski; (3) our interactor team, consisting of Felicia Graybeal, Kate Ingraham, Kenneth Ingraham, and Morgan Russell, who are the Wizards of Oz responsible for breathing life into the avatars; (4) Lisa Dieker and Michael Hynes, who are the PIs (with Charles Hughes) on the TeachLivE project and also our long-term collaborators in the College of Education at UCF; (5) Tracy St. Benoit, Yasuharu Okuda, Carrie Straub, and Eleazar Vasquez (PI on autism-based study reported in this paper), who are subject matter experts working closely with us on scenario development.

References

Ahn, S. J., Fox, J., & Bailenson, J. N. (2012). Avatars. In *Leadership in Science and Technology: A Reference Handbook* (Chap. 79). Thousand Oaks, CA: SAGE.

Bailenson, J. N., & Blascovich, J. J. (2004). Avatars. In W. S. Bainbridge (Ed.), *Encyclopedia of Human–Computer Interaction* (pp. 64–68). Great Barrington, MA: Berkshire Publishing Group.

Baylor, A. L. (2011). The design of motivational agents and avatars. *Educational Technology Research and Development*, *59*(2), 291–300.

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, *13*(2), 103–124.

Bottoni, P., Faralli, S., Labella, A., Malizia, A., Pierro, M., & Ryu, S. (2008). CoPuppet: Collaborative interaction in virtual puppetry. In R. Adams, S. Gibson, & S. M. Arisona (Eds.), *Transdisciplinary digital art: Sound, vision and the new screen* (pp. 326–341). Berlin: Springer.

Dieker, L. A., Lingnugaris-Kraft, B., Hynes, M., & Hughes, C. E. (2013). Mixed reality environments in teacher education: Development and future applications. In B. Collins & B. Ludlow (Eds.), *American Council for Rural Special Educators*. Lexington, KY: ACR SE.

Dow, S., MacIntyre, B., Lee, J., Oezbek, C., Bolter, J., & Gandy, M. (2005). Wizard of Oz support throughout an iterative design process. *IEEE Pervasive Computing*, *4*(4), 18–26.

Erbiceanu, E., Mapes, D. P., & Hughes, C. E. (2014). Modeling attention and interaction in small groups of virtual characters. In J. Tanenbaum, M. S. El-Nasr, & M. Nixon (Eds.), *Nonverbal communication in virtual worlds*. Pittsburgh, PA: ETC Press.

Faller, J., Müller-Putz, G., Schmalstieg, D., & Pfurtscheller, G. (2010). An application framework for controlling an avatar in a desktop-based virtual environment via a software SSVEP brain–computer interface. *Presence: Teleoperators and Virtual Environments*, *19*(1), 25–34.

Fox, J., Yeykelis, L., Janssen, J. H., Ahn, S. J., Segovia, K. Y., & Bailenson, J. N. (2010). A meta-analysis quantifying the effects of avatars and agents on social influence. *Proceedings of the National Communication Association Annual Convention, NCA*.

Garau, M., Slater, M., Pertaub, D.-P., & Razzaque, S. (2005). The responses of people to virtual humans in an immersive virtual environment. *Presence: Teleoperators and Virtual Environments*, *14*(1), 104–116.

Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., & Sasse, M. A. (2003). The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 529–536.

Gildfind, A., Gigante, M. A., & Al-Qaimari, G. (2000). Evolving performance control systems for digital puppetry. *The Journal of Visualization and Computer Animation*, *11*(4), 169–183.

Gross, D. C., Stanney, K. M., & Cohn, L. J. (2005). Evoking affordances in virtual environments via sensory-stimuli substitution. *Presence: Teleoperators and Virtual Environments*, *14*(4), 482–491.

Held, R., Gupta, A., Curless, B., & Agrawala, M. (2012). 3D puppetry: A Kinect-based interface for 3D animation. *ACM Symposium on User Interface Software and Technology (UIST)*, 423–434.

Hughes, C., Dieker, L., Nagendran, A., & Hynes, M. (2013). *Semi-automated digital puppetry control*. US Provisional Patent, SL: 61/790,467, Date filed: 03/15/2013.

Hunter, S., & Maes. P. (2013). Designing digital puppetry systems: Guidelines and best practices. *ACM SIGCHI Conference on Human Factors in Computing Systems, CHI; Extended Abstract*, 2821–2822.

Kelley, J. F. (1984). An iterative design methodology for user-friendly natural language office information applications. *ACM Transactions on Information Systems*, *2*(1), 26–41.

Lee, J., Chai, J., Reitsma, P. S. A., Hodgins, J. K., & Pollard, N. S. (2002). Interactive control of avatars animated with human motion data. *ACM SIGGRAPH Annual Conference on Computer Graphics and Interactive Techniques*, 491–500.

Leite, L., & Orvalho, V. (2011). Anim-actor: Understanding interaction with digital puppetry using low-cost motion capture. *Proceedings of the International Conference on Advances in Computer Entertainment Technology, ACE*. doi:10.1145/2071423.2071505.

Lim, S., & Reeves, B. (2010). Computer agents versus avatars: Responses to interactive game characters controlled by a computer or other player. *International Journal of Human–Computer Studies*, *68*(1–2), 57–68.

Lincoln, P., Welch, G., Nashel, A., State, A., Ilie, A., & Fuchs, H. (2011). Animatronic shader lamps avatars. *Virtual Reality*, *15*(2–3), 225–238.

Lopez, A. L., Hughes, C. E., Mapes, D. P., & Dieker, L. A. (2012). Cross cultural training through digital puppetry. In D. M. Nicholson (Ed.), *Advances in design for cross-cultural activities, Part I*, (Chap. 25, pp. 247–256). Boca Raton, FL: CRC Press.

Mapes, D. P., Tonner, P., & Hughes, C. E. (2011). Geppetto: An environment for the efficient control and transmission of digital puppetry. *Proceedings of the International Conference* *on Virtual and Mixed Reality: Systems and Applications*, 270–278.

Mazalek, A., Chandrasekharan, S., Nitsche, M., Welsh, T., Clifton, P., Quitmeyer, A., Peer, F., Kirschner, F., & Athreya, D. (2011). I'm in the game: Embodied puppet interface improves avatar control. *Proceedings of the International Conference on Tangible, Embedded, and Embodied Interaction, TEI*, 129–136.

Mazalek, A., Chandrasekharan, S., Nitsche, M., Welsh, T., Thomas, G., Sanka, T., & Clifton, P. (2009). Giving your self to the game: Transferring a player's own movements to avatars using tangible interfaces. *Proceedings of the ACM SIGGRAPH Symposium on Video Games (Sandbox)*, Vol. 1, 161–168.

Nagendran, A., Pillat, R., Hughes, C. E., & Welch, G. (2012). Continuum of virtual-human space: Towards improved interaction strategies for physical-virtual avatars. *Proceedings of the ACM SIGGRAPH International Conference on Virtual Reality Continuum and Its Applications in Industry, VRCAI*, 135–142.

Nagendran, A., Pillat, R., Kavanaugh, A., Welch, G., & Hughes, C. (2013). AMITIES: Avatar-mediated interactive training and individualized experience system. *Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology*, 143–152.

Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, *12*(5), 481–494.

Nunez, D., & Blake, E. (2003). Conceptual priming as a determinant of presence in virtual environments. *Proceedings of the 2nd International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa*, 101–108.

Qu, C., Brinkman, W.-P., Wiggers, P., & Heynderickx, I. (2013). The effect of priming pictures and videos on a question-answer dialog scenario in a virtual environment. *Presence: Teleoperators and Virtual Environments*, *22*(2), 91–109.

Razer. (2013). Razer Hydra. Retrieved from http://www.razerzone.com/gaming-controllers/razer-hydra

Revelle, G. L. (2003). Educating via entertainment media: The Sesame Workshop approach. *Computers in Entertainment (CIE)*, *1*(1), 1–9.

Riek, L. (2012). Wizard of Oz studies in HRI: A systematic review and new reporting guidelines. *Journal of Human–Robot Interaction*, *1*(1), 119–136.

Rizzo, A., Buckwalter, J. G., Forbell, E., Reist, C., Difede, J., Rothbaum, B. O., . . . Talbot, T. (2013). Virtual reality applications to address the wounds of war. *Psychiatric Annals*, *43*(3), 123–138.

Rudolph, J. W., Simon, R., Rivard, P., Dufresne, R. L., & Raemer, D. B. (2007). Debriefing with good judgment: Combining rigorous feedback with genuine inquiry. *Anesthesiology Clinics*, *25*(2), 361–376.

Semwal, S. K., Hightower, R., & Stansfield, S. (1998). Mapping algorithms for real-time control of an avatar using eight sensors. *Presence: Teleoperators and Virtual Environments*, *7*(1), 1–21.

Shin, H. J., Lee, J., Shin, S. Y., & Gleicher, M. (2001). Computer puppetry: An importance-based approach. *ACM Transactions on Graphics (TOG)*, *20*(2), 67–94.

Slater, M. (2009). Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3549–3557.

SREAL. (2013a). Video of AMITIES used to control both virtual and physical manifestations of avatars. Retrieved from https://srealserver.eecs.ucf.edu/~arjun/system_overview_video_2.mp4

SREAL. (2013b). Video of TeachLivE math lesson. Retrieved from https://srealserver.eecs.ucf.edu/~arjun/TeachLive_Math_Lesson_Short.mp4

SREAL. (2013c). Video of Veteran's Administration debriefing scenario. Retrieved from https://srealserver.eecs.ucf.edu/~arjun/va_debrief.mp4

Stanney, K. M., Mourant, R. R., & Kennedy, R. S. (1998). Human factors issues in virtual environments: A review of the literature. *Presence: Teleoperators and Virtual Environments*, *7*(4), 327–351.

Sturman, D. J. (1998). Computer puppetry. *IEEE Computer Graphics and Applications*, *18*(1), 38–45.

Weise, T., Bouaziz, S., Li, H., & Pauly, M. (2011). Realtime performance-based facial animation. *ACM Transactions on Graphics (TOG)*, *30*(4), 1–10.

Wirth, J., Norris, A. E., Mapes, D., Ingraham, K. E., & Moshell, J. M. (2011). Interactive performance: Dramatic improvisation in a mixed reality environment for learning. In R. Shumaker (Ed.), *Virtual and mixed reality: Systems and applications. Lecture notes in computer science*, Vol. 6774 (pp. 110–118). Berlin: Springer-Verlag.

Wooldridge, M., & Jennings, N. R. (1995). Intelligent agents: Theory and practice. *Knowledge Engineering Review*, *10*(2), 115–152.