AMITIES: Avatar-Mediated Interactive Training and Individualized Experience System

Arjun Nagendran[*]
Synthetic Reality Lab
University of Central Florida

Remo Pillat[†]
Synthetic Reality Lab
University of Central Florida

Adam Kavanaugh[‡]
Synthetic Reality Lab
University of Central Florida

Greg Welch[§]
Synthetic Reality Lab
University of Central Florida

Charles Hughes[¶]
Synthetic Reality Lab
University of Central Florida

Abstract

This paper presents an architecture to control avatars and virtual characters in remote interaction environments. A human-in-the-loop (interactor) metaphor provides remote control of multiple virtual characters, with support for multiple interactors and multiple observers. Custom animation blending routines and a gesture-based interface provide interactors with an intuitive digital puppetry paradigm. This paradigm reduces the cognitive and physical loads on the interactor while supporting natural bi-directional conversation between a user and the virtual characters or avatar counterparts. A multi-server-client architecture, based on a low-demand network protocol, connects the user environment, interactor station(s) and observer station(s). 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. This approach has been used to create experiences designed for training, education, rehabilitation, remote presence and other-related applications.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual reality I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Animation H.5.2 [Information Interfaces and Presentation]: User Interfaces—Interaction styles C.2.3 [Computer-Communication Networks]: Distributed Systems—Client/server

Keywords: Interaction Techniques, Avatars and Virtual Humans in VR, Applications of VR (Training Systems)

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 driven by a computer program. Advances in Artificial Intelligence (such as Natural Language Processing, 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 bi-directional 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 we have developed a new system for controlling virtual characters that involves a human interactor who "inhabits" (controls) the character. This human-in-the-loop approach combines digital puppetry [Hunter and Maes 2013; Mapes et al. 2011] with basic Artificial Intelligence routines (expert systems) and a network interface to allow the human interactor to control multiple virtual characters seamlessly from a remote location. The system, illustrated in Figure 1, uses a network-efficient protocol during control, thereby minimizing the required bandwidth and hence any associated latencies. 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 supports multiple interactors and flexible assignments to support many-to-many, many-to-one, and one-to-many interactor-character scenarios. An integrated after-action review system supports trainers in the processes of tagging and commenting on events, subsequently using these to assist reflection on the part of users (trainees). 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.

We start by providing context through discussions of the basics of digital puppetry which forms the basis of the system. We then describe the architecture of the system and its individual components. We also present some of our previous user interfaces for our "interactors", individuals trained in improvisation, interactive performance and story development [Erbiceanu et al. 2013], who, with the aid of agent-based (programmatically determined) behaviors, control the virtual characters. We discuss the devices and the interaction paradigms, with an eye towards reducing the cognitive and physical demands placed on these puppeteers. Our user and inhabiter interfaces are aimed at intuitiveness and low cost, while retaining the realism of the interaction required during critical personalized training and rehabilitation.

2 Problem Description

Using virtual characters and associated environments for applications such as training, rehabilitation, and practicing inter-personal skills has several associated challenges. One challenge area is related to the technology affordances of the system ; another is related to the virtual character interaction paradigm . For the experience

---

[*] e-mail:arjun@cs.ucf.edu
[†] e-mail:rpillat@cs.ucf.edu
[‡] e-mail:akavanaugh@knights.ucf.edu
[§] e-mail:gwelch@ist.ucf.edu
[¶] e-mail:ceh@cs.ucf.edu

VRST 2013, October 6–9, 2013, Singapore.

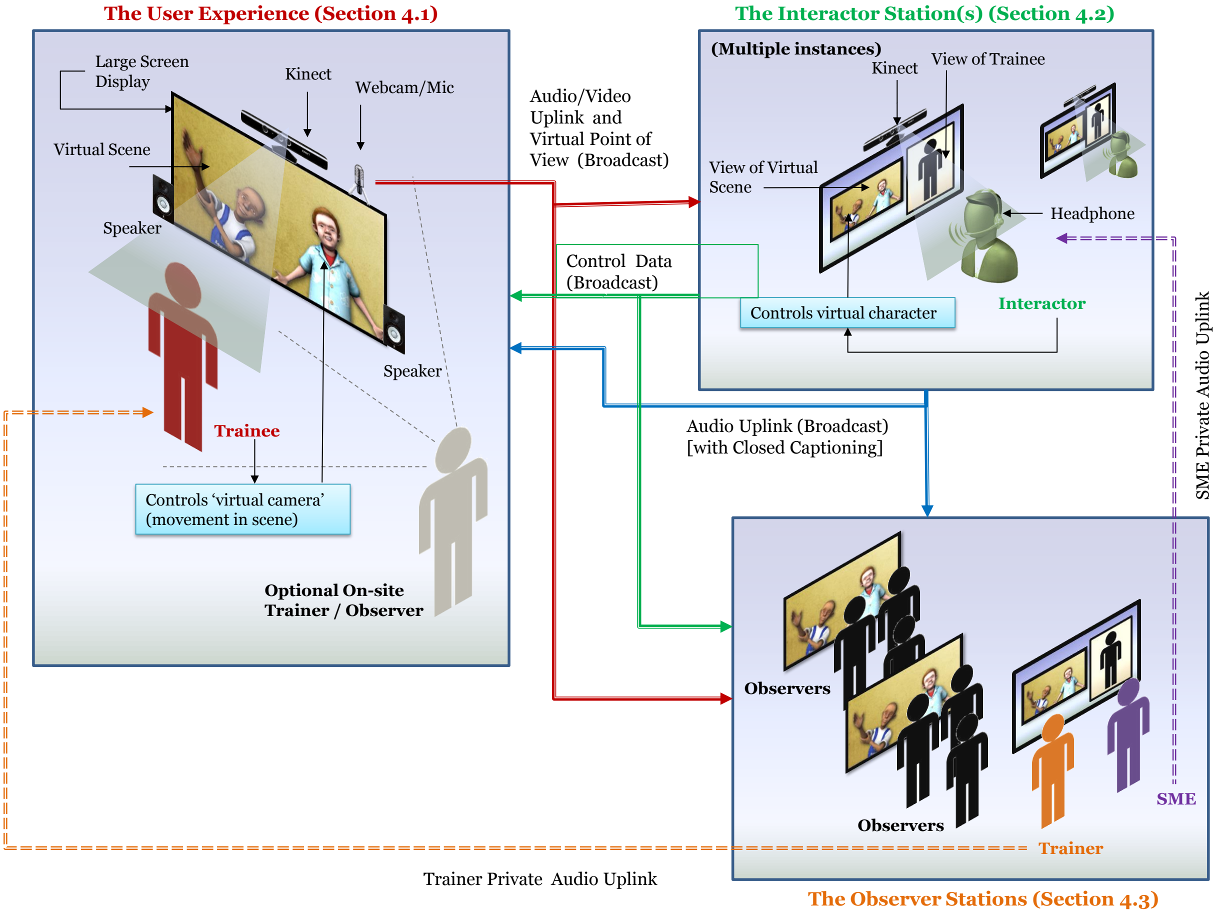

Figure 1: *The complete system showing different stations and data flow between them. The acronym SME is used to indicate a Subject Matter Expert in the above figure.*

to be effective, a user's beliefs about the validity of the scenario should be fostered, preserved, and reinforced. Explicit or implicit anomalies during bi-directional 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 comprising a human (interactor) in the loop there are several specific challenges including setting up a bi-directional 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 towards training, rehabilitation, and other applications where human interaction is critical.

3 Background

Traditionally, two terms have been used to denote manifestations of virtual humans in these experiences: avatars and agents. The distinction is based on the controlling entity, which could be either a human (avatar) or a computer algorithm (agent) [Bailenson and Blascovich 2004]. There is a rich set of literature comparing how the agency of a virtual character is perceived by human users. In general, intelligent agents [Wooldridge and 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 et al. 2012]

Several studies have provided valuable insights that both agents and avatars are effective tools in establishing a feeling of telepresence (user feels present in virtual world) and social presence (user acknowledges the presence of another intelligent mind) [Nowak and Biocca 2003] for the user. On the other hand, a recent meta study 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 and Reeves 2010].

Due to the open-ended nature of conversations in bi-directional conversations in training and rehabilitation scenarios, our AMITIES system uses human-controlled avatars. This choice of 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 used successfully employed for decades in many fields including children's education [Revelle 2003], games [Mazalek et al. 2009], and interactive networked simulations [Dieker et al. 2013].

Existing puppeteering systems mapped the full-range of captured human motion data to an avatar, e.g., [Lee et al. 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. [Shin et al. 2001] uses 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 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 micro-poses [Mapes et al. 2011; Nagendran et al. 2012].

Other recent approaches to capturing the human user employ the Kinect system, e.g. [Leite and Orvalho 2011] and [Held et al. 2012]. There are also techniques that solely concentrate on capturing a human's face with high precision [Weise et al. 2011]. Others have worked on the use of arbitrary control devices to control avatars through genetic programming [Gildfind et al. 2000], and 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]. 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. 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. In addition, the emphasis of this paper is on the general system architecture that supports efficient, human-guided avatar control, network bandwidth optimization and observing capabilities.

4 System Description

AMITIES is a mixed-reality environment designed for individualized experience creation such as training and rehabilitation, and is based on the Marionette Puppetry Paradigm. The system has evolved over a period of six years with continuous refinements as a result of constant use and evaluation. The system has the following features:

1. A low-cost, low-demand (physical and cognitive) interface for interactors that allows them to easily participate in the control of the verbal and non-verbal activities of a set of virtual characters;
2. a low-cost, unencumbered interface for users that allows them to employ natural movement and verbal/non-verbal interaction with virtual characters;
3. a network protocol that supports real-time remote interaction even when dealing with relatively poor network connections; and
4. an integrated after-action review system that supports trainers in the processes of tagging and commenting on events, subsequently using these to assist reflection on the part of users.

Figure 1 shows the architecture of the system, with the individual components and the data flow between them explained in detail in the following sections.

The environment consists of several characters (digital avatars) in a virtual environment, whose space is shared with the real world. In specific, the movement of a user in the real world controls the movement of a camera (view-point) in the virtual environment. This allows the user to walk up to specific characters in the environment to initiate a conversation with them.

The digital avatars themselves have a hybrid intelligence model. This hybrid intelligence model facilitates a more natural and unconstrained interaction on a wide variety of topics as opposed to traditional interaction techniques. A human in-the-loop (interactor) can control any of these digital avatars, allowing them to gesture, change facial expressions, and hold bi-directional conversations that are both contextual and meaningful depending on the pre-determined scenario. When not controlled by a human, the digital avatars exhibit pre-recorded idle behaviors that are executed throughout a session.

AMITIES consists of a server-client model that supports bi - directional communication. The server controls the digital 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 digital avatar (interactor-controlled) and trainee during the session. The interactor (server) also receives a video feed of the trainee, allowing them to assess body language and other non-verbal 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 IR Cameras, The Microsoft Kinect, Razer Hydra, and Keypads. This is mapped onto the digital avatars via a custom animation blending system. The movement of the trainee (at the client) in the interaction space controls the camera view during the process. This camera view is seen by the interactor, allowing them to determine the character that is in-focus.

4.1 The User Experience

Figure 2 The user, typically located at a remote site, stands or sits in front of a large display on which the current scene is visible.

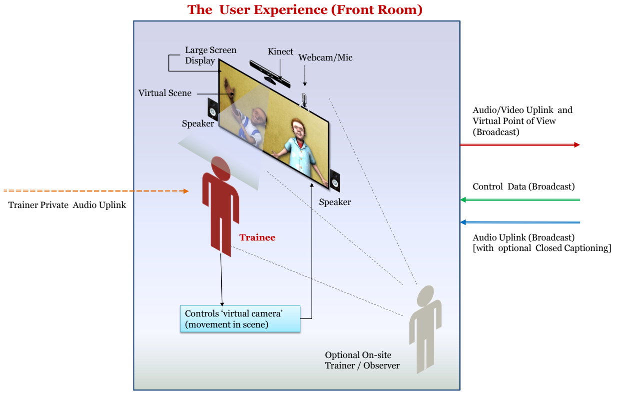

Figure 2: *The User Experience (extracted from the left-hand-side of Figure 1)*

The user's movement is tracked by a Microsoft Kinect for Xbox 360. The user's arms and head are tracked via a VICON IR Tracking System that features 10 T-40S imaging sensors. At present, the user'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 towards the display results in a corresponding movement of the virtual camera through the scene's space (Figure 3).

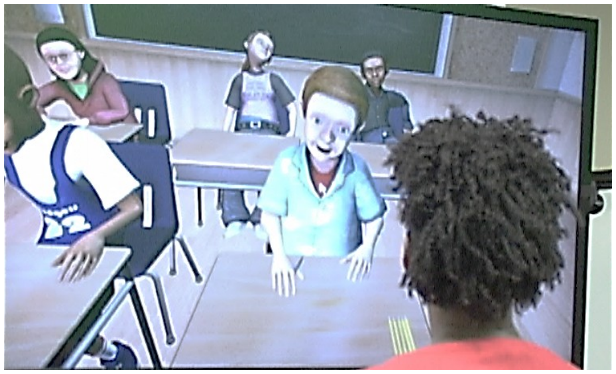

Figure 3: *User experiencing TeachLivE*^{TM}*virtual classroom*

For instance, in a classroom instruction (TeachLivE$^{\text{TM}}$), when a teacher moves towards a student, the virtual camera moves in the same direction. We typically place named pads on the floor associated with standing directly in front of each student. If the user bends down slightly, their point-of-view can result in eye-to-eye contact with the virtual character. In our classroom environment, the virtual children's eyes automatically follow the teacher, unless the student is tagged as having autistic behavior or attention deficit.

Speakers and a microphone are used to create a bi-directional audio channel, with the audio/video uplink and the 'virtual point of view' of the user being broadcast over the network. This allows external 'observers / trainers' to view the interaction passively. A private audio uplink facilitates a trainer or subject matter expert to directly communicate with the user (trainee).

4.2 The Interactor Station(s)

The AMITIES control paradigm has evolved from a very literal system based on motion-capture to a gestural one based on a marionette paradigm. Common to all the paradigms we have implemented are 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 the interactor, and all others exhibiting agent-based behaviors that are influenced by the actions of the interactor, the current avatar, and the user.

4.2.1 Previous Interactor UI Paradigms

Historically, we explored several User Interface (UI) paradigms to allow the interactors to control the virtual characters. We describe those prior implementations here, and describe the current implementation in Section 4.2.2. Initially, we used tracking cameras, with the interactor wearing a vest and baseball cap that had retro-reflective markers (Figure 4). The vest had markers on wrists, elbows and shoulders. The baseball cap had an easily recognized pattern of three retro-reflective markers in order to capture orientation as well as position. This approach had noise problems typically experienced in motion-capture, but without the opportunity for post-production, as all actions had to take effect in real time. Moreover, with capture rates of 120 fps, we were transmitting a substantial amount of network data, with the attendant issues when communicating with clients who had poor connectivity.

Figure 4: *Puppetry controlled by upper body and head motion capture. Foot buttons are used for facial expressions and Ergodex for agent behaviors. Interactor's monitors have two windows: one shows virtual content as seen by user; the other window displays a video of the user.*

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 micro-poses. These are a set of key poses that, in effect, form a basis set for the poses that an avatar is expected to perform, between which blending can occur. Some of these micro-poses are shown super-imposed on each other to view the 'motion-space' of the avatar in Figure 5.

Figure 5: *Microposes for a virtual avatar named "Sean" (a) shows Sean standing (translucent) and holding a pen (solid) (b) shows Sean leaning forward and turning (translucent) and slouching (solid) (c) depicts Sean laying on the desk (translucent) and raising his hand (solid).*

After we developed the concept of micro-poses, we experimented with a series of gestural schemes to control the selection of these micro-poses. The first involved replacing upper body tracking with a single retro-reflective ball on a golf tee-like holder (a pawn). In this version, avatar control was done by moving the pawn close to the desk on which we placed a template of characters (one per column) and their genres of behaviors (one per row). Once a selection is made, the interactor raises the pawn and it enters a virtual sphere that is populated by key-frames (also called micro-poses). As one moves the pawn (see Figure 6), the system finds the closest such poses for weighted blending and then starts a process of decaying the weights of current poses, while increasing those of the newly "selected" ones. Facial gestures, mouth movements, parameters to agent behaviors and scene transitions are controlled by a game pad. Network traffic is greatly reduced with this paradigm in that all that needs to be transmitted are the newly selected micro-poses and their weights. The actual pose blending takes place independently at the server and client sites.

Figure 6: *Puppetry controlled by a pawn (single retro-reflective marker on a post), head tracking and a game pad. Note the template on the desk. This was used in studies that employed puppetry for experiences that dealt with peer pressure and cross cultural interaction.*

When the Kinect for Xbox 360 was released in November, 2010, we briefly went back to using a literal mode of controlling avatars and, in fact, continue to use the Kinect for Windows for head orientation in applications where this is appropriate. The problem with a purely literal approach is that it makes it hard to implement some desired behaviors, such as having the avatar place her head on a desk, as we often want to do when using the system in teacher training, or going into a fetal position as required with some experiences involving social and psychological issues. Having the interactor place her head on the table or buried in her chest would make it very hard for her 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 like standing up and clicking a pen are just more natural to trigger by gestural, rather than literal movements. For these reasons, we returned to gestural as soon as we became aware of the capabilities of the Razer Hydra in spring 2011.

Figure 7: *Puppeteer controlling students in virtual classroom. Uses head tracking and Hydra. Note the window with video-feed (on the right side of the monitor) that allows the interactor to observe a user's non-verbal behaviors.*

4.2.2 Current Interactor UI Paradigm

Figure 8 shows the system architecture at the interactor station. The 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. The interactor is seated in front of a large-screen display, 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 - no video uplink of the interactor is provided to either observers or the users. This helps keep the interaction paradigm 'behind closed doors' to promote situational plausibility and belief. A Subject Matter Expert (SME) has a private audio uplink to the interactor, allowing them to prompt appropriate responses to complicated situations as required.

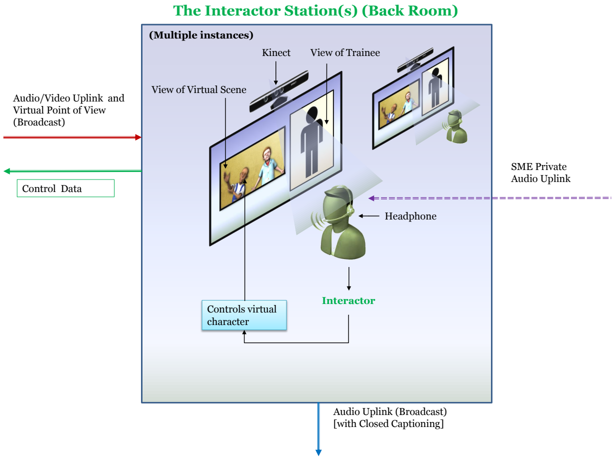

Figure 8: *The Interactor Station (extracted from the top-right-hand-side of Figure 1)*

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 hand-held 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 one millimeter and one degree. 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-specifc reactions and appearance-related features of a virtual character such as frowning, smiling and winking. As with the simple pawn scheme (Figure 6), we have a library of micro-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 minimum physical demands. This particular approach appears to provide the best balance between expressiveness and cognitive and physical requirements on the interactor.

Control of the current avatar's pose is done by gestures that are mapped to micro-poses, with variations in those gestures coming from being close to several poses and by twisting the controllers to get subtle deviations (see Figure 7). This is explained in more detail below.

The current activated virtual character is controlled using a micro-pose system with the Razer Hydra controller's 3D position and orientation input across two hand-held controllers. Every micro-pose is configured with a user specified pair of 3D coordinates, one for each controller (recorded via a calibration phase using the 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 which best matches the input coordinates. The two pose cross-fade system selects the two poses with the shortest Euclidean distance to the input and then calculates the animation blend between them, allowing for a partial poses that is the combination of the two selected poses. If the selected pose is not currently active, the system begins to transition into this 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 the 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 animations transition rates. This allows for a more natural and fluid motion between poses giving the interactor more fine-grained and direct control depending on their initial pose configurations and their 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 by rotating the controllers about their 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 active animation mixture depending on the current micro-pose mode.

The virtual character's facial expressions are controlled with the 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 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.

4.3 The Observer Station(s)

The system architecture of the observer stations is shown in Figure 9. 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 Subject Matter Experts (SME) and Trainers, allowing them to interact either with the interactor or the trainee (when appropriate) to inject their specialized opinions. The SMEs and Trainers can be thought off as special observers, who also have the option of viewing the trainee (driven by a situational need) if the latter requests / permits this. Several instances of the observer station can be simultaneously spawned, thereby supporting interaction from remote locations.

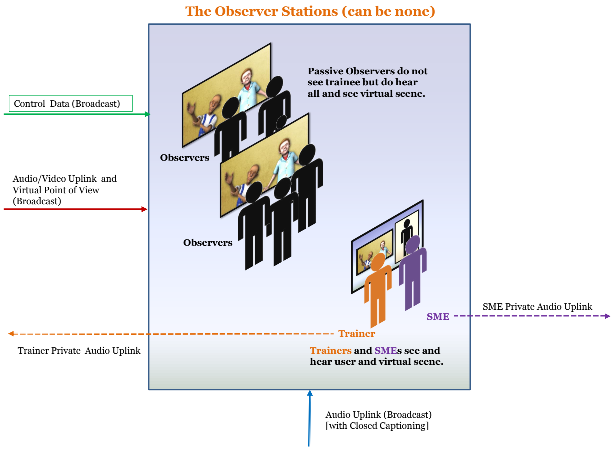

Figure 9: *The Observer Station (extracted from the bottom-right-hand-side of Figure 1)*

4.4 Recording and After-Action Review

Our system also includes a subsystem 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 hone in on sets of frames in which these behaviors are observed during a training session. For example, in teacher practice, a coder might tag frames in which the user asks high order questions (a positive attribute) or in which very little time is allowed to expire 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. 2013].

5 Using AMITIES

AMITIES provides a flexible framework for controlling expressive, avatar-mediated, human-to-human communication. However it does not inherently define character personalities nor 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 AMITIES to create an overall experience for a particular purpose.

5.1 Character and Story Design

The SREAL team has developed 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, interactors, programmers and, most importantly, the requirements of users of the resulting system. Components of this include model design and creation, and verbal and non-verbal behavior selection and implementation (puppeteered and automated).

The design process starts with meetings to refine the requirements. These meetings include a principal investigator, two to three interactors, a software developer, at least one modeler and a subject-matter expert. The interactors then rehearse the characters' (including the trainee's) behaviors (verbal and non-verbal) using a role-playing approach designed to flesh out the characters' backstories and interaction styles. 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. Latter stages of these role-playing sessions are recorded for analysis and eventually for use by the artist(s) and programmer(s).

These initial steps establish artistic/technical requirements. We then produce concept art. Once these artistic designs are accepted and the role-playing is deemed to have uncovered the collection of required gestures (facial and body), the artists do model-development, texturing and rigging of the characters. We then identify key frames (micro-poses) that support specific behaviors (those uncovered in rehearsal) as well as a broad range of dynamically created behaviors so the Interactor can react to widely varying interactions with users. Additionally, the artist/interactor/programmer team develops animations of specific behaviors like "annoy others", "look attentive" or "act bored", and transitions between scenes, as needed. This results in an operational set of puppets and scenes. With this nearly final product in-hand, our interactors perform role-playing rehearsals again, but this time using the puppetry system. The outcome of this process is then a set of characters, micro-poses, scenes, animations and decision trees that enable the final experiences.

5.2 Case Studies

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 et al. 2012], peer pressure resist/avoid strategies [Wirth et al. 2011] and communication skills development for young adults with autism. Here we describe several cases where we applied the above design processes and AMITIES for a particular application.

As shown in Figure 10, AMITIES is the foundation for the TLE TeachLivE ^{TM}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 ^{TM}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 recondite biases, so they 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 assistance to this program, as it provides objective measures for the teacher and trainer to use during reflection.

AMITIES also supports the control of physical-virtual avatars (PVA)—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 as well as physical manifestations of the same character that involves mechanical components (robotic) on, for instance, physical virtual avatar. As a part of this submission, we have included a video (screengrab shown in Figure 11) 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. In specific, multiple interactors control all the virtual characters in the scene (Section 4.2), while the PVA and the 2D Flat-screen display provide the User Experience (Section 4.1). 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.

We used AMITIES 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 (i) face-face with a human, (ii) face-face with a virtual character on a flat-screen 2D display surface and (iii) face-face with a physical manifestation of the virtual character (a Physical-Virtual Avatar [Lincoln et al. 2011]). The scenarios and virtual characters were developed to facilitate a ten minute 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.

Finally, we are using AMITIES as the underlying framework for a multi-year 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.

6 Conclusion and Future Work

We present a system for controlling virtual characters/avatars in remote environments. The system lends itself to control of character manifestations ranging from purely virtual (e.g., a 2D display) to physical (e.g., a Physical-Virtual Avatar). 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 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 on integrating an after action review 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

Figure 10: *Virtual class of five students. Left screen shows virtual students on task; right shows them distracted, except for student in front row left who is presently controlled by Interactor.*

interactions. This involves analyzing video data streams, tracking data streams and audio data during the interactions.

Acknowledgements

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) and the Bill & Melinda Gates Foundation. 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.

References

AHN, S. J., FOX, J., AND BAILENSON, J. N. 2012. Avatars. In Leadership in Science and Technology: A Reference Handbook. SAGE Publications, Inc, ch. 79.

BAILENSON, J. N., AND BLASCOVICH, J. J. 2004. Avatars. In Encyclopedia of Human-Computer Interaction, W. S. Bainbridge, Ed., 1st ed. Berkshire Publishing Group, 64–68.

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

BLASCOVICH, J., LOOMIS, J., BEALL, A. C., SWINTH, K. R., HOYT, C. L., AND 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., AND RYU, S. 2008. CoPuppet: Collaborative Interaction in Virtual Puppetry. In Transdisciplinary Digital Art. Sound, Vision and the New Screen, R. Adams, S. Gibson, and S. M. Arisona, Eds., vol. 7 of Communications in Computer and Information Science. Springer Berlin Heidelberg, 326–341.

DIEKER, L. A., LINGNUGARIS-KRAFT, B., HYNES, M., AND HUGHES, C. E. 2013. Mixed Reality Environments in Teacher Education: Development and Future Applications. In American Council for Rural Special Educators, In Press, B. Collins and B. Ludlow, Eds.

DOW, S., MACINTYRE, B., LEE, J., OEZBEK, C., BOLTER, J., AND GANDY, M. 2005. Wizard of Oz Support throughout an Iterative Design Process. IEEE Pervasive Computing 4, 4 (Oct.), 18–26.

ERBICEANU, E., MAPES, D. P., AND HUGHES, C. E. 2013. Modeling Attention and Interaction in Small Groups of Virtual Characters. In Non-Verbal Communication in Virtual Worlds, In Press, J. Tanenbaum, M. Nixon, and M. S. El-Nasr, Eds. ETC Press.

FOX, J., YEYKELIS, L., JANSSEN, J. H., AHN, S. J., SEGOVIA, K. Y., AND BAILENSON, J. N. 2010. A Meta-Analysis Quantifying the Effects of Avatars and Agents on Social Influence. In Nactional Communication Association Annual Convention (NCA).

GILDFIND, A., GIGANTE, M. A., AND AL-QAIMARI, G. 2000. Evolving Performance Control Systems for Digital Puppetry. The Journal of Visualization and Computer Animation 11, 4, 169–183.

HELD, R., GUPTA, A., CURLESS, B., AND AGRAWALA, M. 2012. 3D Puppetry: A Kinect-Based Interface for 3D Animation. In ACM Symposium on User Interface Software and Technology (UIST), ACM, New York, NY, USA, UIST '12, 423–434.

HUNTER, S., AND MAES, P. 2013. Designing Digital Puppetry Systems: Guidelines and Best Practices. In ACM SIGCHI Conference on Human Factors in Computing Systems (CHI) - Extended Abstract, ACM, Paris, France, 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 (Jan.), 26–41.

LEE, J., CHAI, J., REITSMA, P. S., HODGINS, J. K., AND POLLARD, N. S. 2002. Interactive Control of Avatars Animated with Human Motion Data. In ACM SIGGRAPH Annual Conference on Computer Graphics and Interactive Techniques, ACM, 491–500.

LEITE, L., AND ORVALHO, V. 2011. Anim-Actor: Understanding Interaction with Digital Puppetry using Low-Cost Motion Capture. In International Conference on Advances in Computer Entertainment Technology (ACE), ACM, New York, NY, USA, ACE '11, 65:1—-65:2.

LIM, S., AND 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 (Jan.), 57–68.

LINCOLN, P., WELCH, G., NASHEL, A., STATE, A., ILIE, A., AND FUCHS, H. 2011. Animatronic Shader Lamps Avatars. Virtual Reality 15, 2-3, 225–238.

Figure 11: *A screen-grab of the submission video that shows the virtual characters on the 2D screen, the physical virtual avatar and a human engaged in a conversation*

LOPEZ, A. L., HUGHES, C. E., MAPES, D. P., AND DIEKER, L. A. 2012. Cross Cultural Training through Digital Puppetry. In Advances in Design for Cross-Cultural Activities Part I, D. M. Nicholson, Ed. CRC Press, ch. 25, 247–256.

MAPES, D. P., TONNER, P., AND HUGHES, C. E. 2011. Geppetto: An Environment for the Efficient Control and Transmission of Digital Puppetry. In International Conference on Virtual and Mixed Reality: Systems and Applications, Springer-Verlag, Berlin, Heidelberg, 270–278.

MAZALEK, A., CHANDRASEKHARAN, S., NITSCHE, M., WELSH, T., THOMAS, G., SANKA, T., AND CLIFTON, P. 2009. Giving Your Self to the Game: Transferring a Player's Own Movements to Avatars Using Tangible Interfaces. In ACM SIGGRAPH Symposium on Video Games (Sandbox), ACM, vol. 1, 161–168.

MAZALEK, A., CHANDRASEKHARAN, S., NITSCHE, M., WELSH, T., CLIFTON, P., QUITMEYER, A., PEER, F., KIRSCHNER, F., AND ATHREYA, D. 2011. I'm in the Game: Embodied Puppet Interface improves Avatar Control. In International conference on Tangible, Embedded, and Embodied Interaction (TEI), ACM, New York, NY, USA, TEI '11, 129–136.

NAGENDRAN, A., PILLAT, R., HUGHES, C. E., AND WELCH, G. 2012. Continuum of Virtual-Human Space : Towards Improved Interaction Strategies for Physical-Virtual Avatars. In ACM SIGGRAPH International Conference on Virtual Reality Continuum and Its Applications in Industry (VRCAI), 135–142.

NOWAK, K. L., AND 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.

RAZER, 2013. Razer Hydra. \url{http://www.razerzone.com/gaming-controllers/razer-hydra}, May.

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 (Aug.), 119–136.

RIZZO, A., BUCKWALTER, J. G., FORBELL, E., REIST, C., DIFEDE, J., ROTHBAUM, B. O., LANGE, B., KOENIG, S., AND TALBOT, T. 2013. Virtual Reality Applications to Address the Wounds of War. Psychiatric Annals 43, 3, 123–138.

SHIN, H. J., LEE, J., SHIN, S. Y., AND GLEICHER, M. 2001. Computer Puppetry: An Importance-Based Approach. ACM Transactions on Graphics (TOG) 20, 2, 67–94.

STURMAN, D. J. 1998. Computer Puppetry. IEEE Computer Graphics and Applications 18, 1, 38–45.

WEISE, T., BOUAZIZ, S., LI, H., AND PAULY, M. 2011. Real-time Performance-Based Facial Animation. ACM Transactions on Graphics (TOG) 30, 4 (July), 77:1–77:10.

WIRTH, J., NORRIS, A. E., MAPES, D., INGRAHAM, K. E., AND MOSHELL, J. M. 2011. Interactive Performance: Dramatic Improvisation in a Mixed Reality Environment for Learning. In Virtual and Mixed Reality - Systems and Applications, R. Shumaker, Ed., vol. 6774 of Lecture Notes in Computer Science. Springer Berlin Heidelberg, 110–118.

WOOLDRIDGE, M., AND JENNINGS, N. R. 1995. Intelligent Agents: Theory and Practice. Knowledge Engineering Review 10, 2, 115–152.