IPSViz: An After-Action Review Tool for Human-Virtual Human Experiences

Andrew B. Raij*   Benjamin C. Lok†

Department of Computer and Information Science and Engineering, The University of Florida

Abstract

This paper proposes after-action review (AAR) with Human-Virtual human (H-VH) experiences. H-VH experiences are seeing increased use in training for real-world, H-H experiences. To improve training, the users of H-VH experiences need to review, evaluate, and get feedback on them. AAR enables users to review their H-VH interaction, evaluate their actions, and receive feedback on how to improve future real-world, H-H experiences.

The *Interpersonal Scenario Visualizer (IPSViz)*, an AAR tool for H-VH experiences, is also presented. IPSViz allows medical students to review their interactions with VH patients. To enable review, IPSViz generates spatial, temporal, and social visualizations of H-VH interactions. Visualizations are generated by treating the interaction as a set of signals. Interaction signals are captured, logged, and processed to generate visualizations for review, evaluation and feedback. The results of a user study (N=27) show that reviewing the visualizations helps students become more self-aware of their actions with a virtual human and gain insight into how to improve interactions with real humans.

Keywords: Virtual Humans, Information Visualization

Index Terms: H.5.1 [INFORMATION INTERFACES AND PRESENTATION]: Multimedia Information Systems—Artificial, augmented, and virtual realities; I.3.7 [COMPUTER GRAPHICS]: Three-Dimensional Graphics and Realism—Virtual Reality

1 Introduction

Human-virtual human (H-VH) interactions are immersive virtual experiences that simulate social scenarios. Existing scenarios include military leadership [14], law enforcement [9], cultural competency [2, 5], medical interview and diagnosis [16, 23, 19], and interaction between children [25]. As training in H-H social situations requires *after-action reviews (AAR)*, so does training with H-VH interactions. This paper 1) proposes AAR with H-VH interactions to enhance the efficacy of virtual social experiences, 2) presents the Interpersonal Scenario Visualizer (IPSViz), a tool for medical students to review their interactions with a VH patient, and 3) describes a user study that evaluates AAR for a real-world interaction between a medical student and a virtual human.

1.1 AAR of a Human-Virtual Human Interaction

Social interaction training is a major component of business, military, and medical education. In these fields, communication skills are taught using lectures, role-play, and situational immersion with expert-observation. A critical educational component of these methods is the AAR. In AARs, students review their social interaction, are evaluated using a combination of self, instructor, and peer-group evaluation. The evaluation serves as feedback to effectively improve social interaction skills.

Recently, VHs have been applied to simulating and educating social interaction skills. To improve skills education, we propose augmenting H-VH interactions with AAR. AAR enables students to *review* their H-VH interaction to *evaluate* their actions, and receive *feedback* on how to improve future real-world experiences.

AAR for H-VH interactions incorporates three design principles:

1. An H-VH interaction is composed of social, temporal, and spatial characteristics. These characteristics will be explored by students via AAR visualizations.
2. An H-VH interaction is a set of signals. Interaction signals are captured, logged, and processed to produce visualizations.
3. An H-VH interaction is complex. Students gain insight into this complexity by reviewing multiple visualizations, such as audio, video, text, and graphs.

1.2 Interpersonal Scenario Visualizer (IPSViz)

To enable AAR, IPSViz processes the signals characterizing an H-VH interaction to provide an array of visualizations. The visualizations are used to facilitate interpersonal skills education. Novel visualizations can be produced by leveraging the many signals that are captured in an H-VH interaction. Given an H-VH interaction, AAR is facilitated through the following visualization types:

- *Spatial visualizations* - The H-VH interaction can be 3D-rendered from any perspective, including that of the conversation partner (the virtual camera is located at the VH's eyes).

Figure 1: A user interacts with a virtual human. Afterward, the user reviews the experience in IPSViz.

*e-mail: raij@cise.ufl.edu
†e-mail: lok@cise.ufl.edu

Students are able to perceive ”what it was like to talk to them”.

- *Temporal visualizations* - Events in the H-VH interaction are visualized with respect to an interaction timeline. Students are able to discern the relationship between conversation events.
- *Social visualizations* - Verbal and nonverbal behaviors are presented in log, graph, and 3D formats. Students are able to understand how their behavior affects the conversation.

1.3 Evaluating AAR for H-VH Experiences

The impacts of IPSViz were determined in a recent user study with 27 health profession students. Students conducted a ten-minute medical interview of a VH (similar to that described in [15]). After the interaction, students used IPSViz to review, evaluate, and get feedback on their communication with the VH (Figure 1). To determine how the AAR affected the students, the students graded their interview on communication skills relevant to the scenario before and after the AAR.

The study shows it is possible to alter a person’s self-identity by combining an H-VH experience with AAR. IPSViz helped students 1) become *self-aware* of their communication skills, and 2) identify ways to *improve their communication with real humans.*

IPSViz expands the virtual human experience to include after-action review. *The virtual human experience does not end when the user leaves the virtual environment.* Rather, the impact of the virtual experience continues into the after-action review. Through AAR, self-awareness and self-directed change are possible.

2 Previous Work

2.1 After-Action Reviews

The term after-action review (AAR) was originally used to describe military reviews of events (e.g., combat missions). AAR allows ”soldiers to discover for themselves what happened, why it happened, and how to sustain strengths and improve on weaknesses.” [6]. The practice of conducting AARs has spread outside the military to business, medical and other training situations.

An important feature of military AARs is the use of training aids such as 3d models (physical or digital), maps, and video. Training aids help AAR participants review the event in context of spatial relationships. Information about key events are also collected to establish the temporal relationships between events. This allows AAR participants to understand how their actions led to specific outcomes (cause and effect). Reviewing experiences in spatial and temporal context is an important principle incorporated into IPSViz.

2.2 After-Action Review Systems

Several systems have been built to support AAR. TAARUS [28] allows military personnel to review events (e.g. troop movements) through graphs, tables and maps. DIVAARS [17] allows review of battlefield simulations practiced in virtual environments (VEs). Phloem [21] visualizes large, behavioral datasets produced from user interactions in VEs. RIAS [24] allows health professions students to review annotated video of their interaction with patients.

Review systems have also been built into core VE tools and platforms. Optitrack [22] provides tools to review motion capture data. CAVERNsoft [18], a toolkit for CAVE applications, allows recording and playback of user speech and avatar behavior. Forterra’s OLIVE supports logging and playback of virtual experiences [8].

Steed et al [29] built an AAR system to study breaks in presence (BIPs). The system rendered the user as an avatar situated in a virtual environment (VE). 3D rays highlighted the user’s head and hand orientation in the VE. User-reported BIPS were shown in the VE as yellow spheres. These visualizations led to the discovery that users focus on the source of a BIP just before reporting it.

Common features of these review systems include 1) capturing data about the experience as it occurs, 2) the ability to play back the experience (on video or as a 3D rendering) and 3) the ability to review captured data through graph and text displays.

2.3 Recording Virtual Experiences for Playback

The VR community has proposed several approaches to recording and playing back virtual experiences. Hart et al [12] propose an image-based approach. The VE is projected onto a viewer-centered environment map. The environment map is later rendered to allow review of the virtual experience.

Greenhalgh et al [11] propose *temporal links* that define a relationship between present and past events in a VE. Defining the temporal, spatial and presentational aspects of this relationship enables adding past events in a VE into a VE occurring in the present.

Friedman et al propose standardizing the recording of VEs for analyzing presence [10]. Guidelines are derived by analyzing a real presence data set. An important distinction is made between temporal data (e.g., system events, tracking, video recordings) and non-temporal data (e.g., questionnaires). The authors propose building tools that aggregate VE data to replay or summarize VEs.

IPSViz aggregates temporal data sets associated with an H-VH experience (e.g. tracking data, audio, video, speech recognition data, event logs, and virtual human behavior) to produce visualizations of the interaction. Rather than focus on presence, the visualizations characterize the communication between a virtual human and a real human. Characterizing this communication enables AAR of H-VH experiences for communication skills training.

3 Interpersonal Scenario Visualizer

3.1 Expanding AAR to H-VH Interactions

By expanding AAR to H-VH interactions, IPSViz builds upon the previous work in AAR. Unique to IPSViz is the focus on reviewing a social experience. Social-experience VEs have very different goals than spatially-driven VEs. For example, reviewing a user’s posture and speech has vastly different meanings in a H-VH interaction than an architectural walkthrough VE. Thus, new approaches and techniques need to be applied to AARs for H-VH interaction. IPSViz’s visualizations draws inspiration from the substantial body of research into interpersonal communication, medical communication, and psychology. In return, IPSViz adds to application and basic research by providing a significant value-add to simulating social situations for training and education.

3.2 Overview of IPSViz

IPSViz is based on representing a H-VH interaction as a set of signals. The signals include user speech, video, tracking data, and VH behavior. IPSViz processes and visualizes these interaction signals to enable AAR (Figure 3).

To generate visualizations of an interaction, the interaction is captured from a variety of sensors (Section 4). From a signal analysis perspective, capture is equivalent to sampling the interaction as if it were a set of continuous signals. These interaction signals characterize the interaction between a human and virtual human.

Before generating visualizations, interaction signals may undergo filtering and processing (Section 5). In both stages, a chain of digital filters is applied to one or more signals to derive new signals. Filtering and processing are separated into two stages as each solves a different problem. Filtering compensates for errors caused by sampling a continuous signal with real-world sensors (e.g. discretization error and noise). Processing manipulates and combines signals to provide new information about the interaction.

After filtering and processing, interaction signals are mapped to the visual (or other perceptual) domain to produce visualizations (Section 6). The visualizations allow users to gain new insight into H-VH communication. Sections 4-6 describe the capture, filter, process and visualize stages in more detail.

Figure 2: A screenshot of IPSViz.

3.3 A Human-Virtual Human Experience

The discussion of IPSViz is guided by a real-world H-VH experience - the interaction between a health professions (HP) student and a VH [15] (Figure 1). This experience was chosen to guide the discussion because HP students 1) take this H-VH interaction seriously [16], and 2) need review, evaluation, and feedback to improve their communication with patients [24, 3].

A typical interaction between a HP student and a VH begins with the VH complaining of a medical problem (e.g., pain). The student's goal is to determine what the problem is (diagnosis) and treat it. Effective diagnosis and treatment requires gathering accurate information from the VH. The student gathers information from the VH by asking the VH questions and examining the VH.

During the interview, the VH will also ask the student questions to learn what is happening to her and why. The VH may also ask questions in the hope that the student can relieve her anxiety about the medical problem. Typical questions a VH can ask include "Do you know what I have?" and "Do you think this could be cancer?" Students should answer these questions carefully to relieve the VH's anxiety and build rapport with the VH.

4 Capturing Human-Virtual Human Communication

Communication is the means by which the VH and user reach their goals. Hence, logging communication is crucial to understanding the interaction. As communication is mediated by the system's input and output devices, system inputs and outputs are logged to capture H-VH communication.

4.1 System Inputs

Natural speech Students wear a wireless microphone on their head. This enables talking to the VH using natural speech. Speech recognition software (Dragon Naturally Speaking 9 Pro) extracts the words spoken by the user from microphone input. Both the speech waveform and the speech recognition output are logged.

Head, hand and body lean tracking Students wear a hat and glove outfitted with reflective markers. Also, markers are attached to the back of the student's chair. Combining head and chair tracking data enables computing approximate body lean. The markers are tracked by an optical, infrared tracking system (# cameras: 2-3, # objects: up to 4, DOF: 6, Update Rate: 60Hz, Latency: 100ms, Registration: $<$ 1cm, Jitter: 10mm). This allows the system to detect user presence, head gaze, pointing gestures, and chair motion. Detected events and the 3D positions of each marker are logged.

Video Video of the interaction is also recorded for later review in IPSViz. Video is recorded because it is a standard practice for students and instructors to review videos of patient interactions.

4.2 System Outputs

Natural speech and animation When a student speaks to the VH, the speech recognition software interprets her words. A keyword matching algorithm matches speech recognition output to questions in the VH's response database. If a match is found in the database, the VH executes a corresponding vocal and gesture response. The VH responds appropriately to the question 60-70% of the time. The VH's gesture and speech responses are logged.

Visual Immersion The interactions take place in a medical exam room or a mock exam room in a controlled laboratory. A projection display (NEC VT770 Projector or equivalent, 1024x768, 60-100 degree diagonal FOV depending on distance to display surface) or a head-mounted display (EMagin Z800, 800x600, 40 degree diagonal FOV) is used to show the user the VE. The VE is rendered at life-size, such that virtual objects appear to have the same proportions as they would if they were real. The VE is rendered in real-time (35 - 45Hz) using the OGRE 3D rendering engine.

Using the head-tracking inputs, the system renders the VE from the student's perspective. This allows the student to perceive the VH's gaze behavior accurately. The head tracking data also allows the VH to respond when the student enters the room. In the HMD condition, head-tracking enables looking around the virtual room. Life-size, user-perspective rendering and VH responsive behaviors create a highly immersive experience. 3D environment paramaters, rendering parameters, and VH gaze are logged.

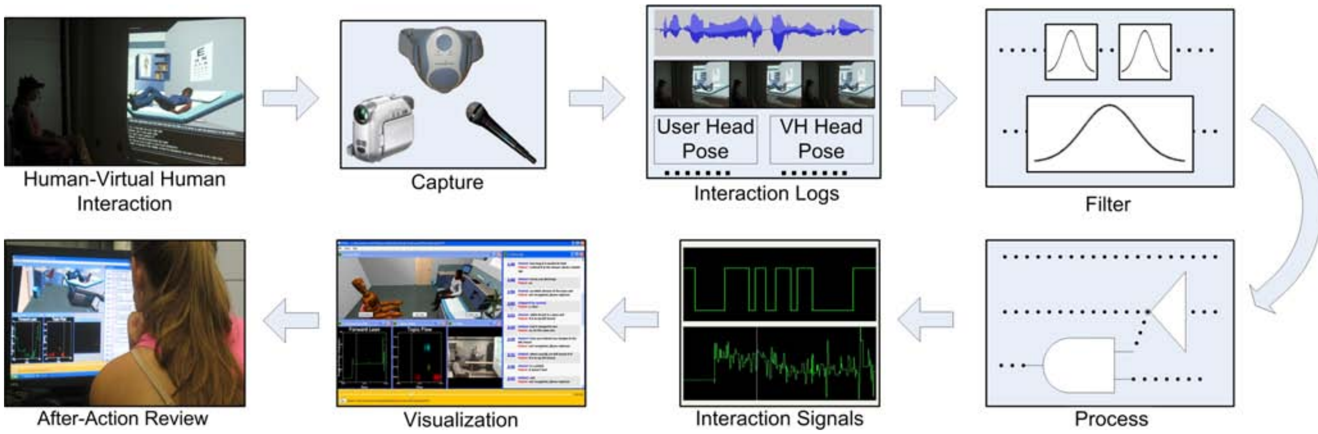

Figure 3: H-VH interaction is captured, filtered, processed, and visualized for review, evaluation and feedback.

4.3 Summary of Captured Data

Capturing a ten-minute H-VH interaction produces on average 70 megabytes of data, stored into six separate files:

Transcript: Timestamped list of events including the speech and gestures of the VH and student. As user speech and gestures are interpreted by imperfect speech and gesture recognition, this data can contain errors.

Student State: Binary file that describes the state of the student over the course of the interaction. This is primarily tracking data.

VH State: Similar to the student state log. It describes the state of the VH over the course of the interaction. VH state includes head pose and a variable indicating if the VH is talking.

Audio: Audio recording of the student's speech. This is recorded by the wireless microphone worn by the student. Note that audio of the interaction is also recorded to the video log.

Video: Video recording (360x240) of the VH and user. This is captured by a mini-DV camera.

System: System and rendering parameters.

Note that this list of data is not meant to be exhaustive, nor completely representative for all VH applications. Rather, the types of data captured were chosen because of the importance of:

- verbal and nonverbal behavior in communication, [13, 20]
- communication content and rapport-building in the medical interview, [27, 3], and
- communication content and rapport-building in H-VH experiences [14, 2, 5, 19, 25, 4, 16].

5 Filtering and Processing

The captured interaction data is interpreted as a set of signals that can be filtered and processed. Filtering and processing is necessary to 1) compensate for sensor error, 2) combine signals to form new signals, and 3) extract signals embedded in other signals.

Sensor Error Real-world sensors introduce errors into the signals they sample. Errors include noise, discretization error, information loss, and aliasing. For example, in optical trackers, these errors correspond to jitter from CCD noise and discrete pixels, data loss due to occlusion and limited tracking volumes, and aliasing when motion is faster than the Nyquist rate. Filtering cannot correct for these errors, but it does compensate for them to acceptable tolerances. Filtering prevents sensor errors from propagating into the visualizations that are presented to users. Hence, filtering is crucial for encouraging users to trust the information IPSViz provides.

Combining Signals: Body Lean One way to measure forward body lean is to measure the angle between the student's back and an up-vector (a vector perpendicular to the floor of the room). Unfortunately, it is difficult to track a student's back with an optical, marker-based tracker. Markers attached to the back could cause discomfort and be occluded when the student leans back in the chair.

To overcome these issues, an approximate body lean signal $L(t)$ is computed by combining head and chair tracking data. First, head and chair data is filtered to compensate for tracker jitter. Then the chair position is subtracted from the head position to compute a head-chair vector. The head-chair vector serves as a proxy for a body lean vector that runs along the student's spine. $L(t)$ is set to the angle between the head-chair vector and the up-vector.

$L(t)$ is a reasonable approximation of body lean because it increases when leaning forward and decreases when leaning back. This allows identification of 1) when body lean changes and 2) the direction of the change (forward or backward). For example, the student represented in Figure 4 did not change their body lean except for leaning forward and then back around 3:12.

Embedded Signals: Topics A topic signal is an example of an embedded signal. Topic signals characterize the discussion of a topic in an H-VH interaction. As the transcript log contains all the speech content of the interview, topic signals are embedded in the transcript. Topic signals can be computed by filtering the transcript log. Filtering extracts words from the transcript that are associated with a topic of interest. For example, if the topic is "family history" then words like "family", "sister", and "mother" remain after filtering. Let $F(t)$ be a mathematical representation of the topic signal. Then for all interaction times associated with the remaining words, the topic signal $F(t) = 1$, else $F(t) = 0$.

6 Spatial, Temporal, and Social Visualization

IPSViz generates visualizations to help health professions students evaluate interactions with VHs. Students evaluate themselves by asking certain kinds of questions about the interaction.

- *How much time* did I spend discussing the VH's symptoms?
- *Where* was the VH's pain?
- *When* was I empathetic to the VH?
- *Was there a moment when* my nonverbal behavior affected rapport with the VH negatively?
- Did I look *at* the VH *when* she was talking, or *elsewhere*?

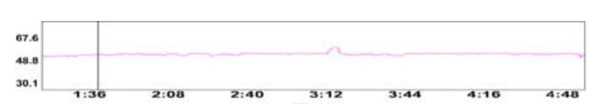

Figure 4: A student's body lean throughout an H-VH interaction.

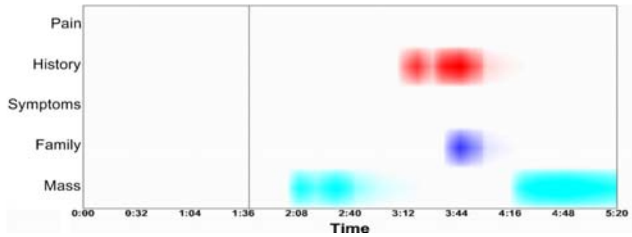

Figure 5: A visualization of five topic signals. The plot shows when the student and VH discussed important topics.

(a) (b)

Figure 6: The interaction is rendered in 3D. The student is represented by a wooden posing doll. The student can observe his behavior from multiple viewpoints, including that of the VH (a). A target is added to the environment to indicate the student's head gaze (b).

As the italicized words above show, these questions are *spatial* and *temporal* in nature. The spatial questions focus on where objects are and how they are related to each other in the space of the 3D world. The temporal questions focus on when events happen and how long they happen for. Furthermore, these questions focus on how the student behaved *socially* with the virtual human. IPSViz generates visualizations that are spatial, temporal, and social to help students gain insight into their communication with the VH.

6.1 Spatial Visualization

IPSViz renders the space of the H-VH interaction using 3D models of the student, VH, and the mixed interaction environment (real room + virtual medical exam room). Using tracking data, the poses of the VH, user, and chair models are updated to reflect their motion during the interaction. This allows students to review their behavior in the space of the interaction environment.

Multiple Viewpoints Using a 3D representation of the interaction enables rendering the interaction from multiple viewpoints. This allows students to see what their behavior looked like to an external observer or to the VH (Figure 6a). Seeing the interaction through the VH's eyes is a powerful way of demonstrating to students how their nonverbal behavior is perceived by their patients.

Augmenting the Environment By using a 3D representation of the interaction, spatial information about the student's communication can be augmented to the environment. IPSViz demonstrates this by augmenting the 3D environment and VH models with a **gaze target** (Figure 6b). The gaze target is a texture that is projected wherever the user was looking during the interaction. The gaze target allows students to become aware of where their attention was actually focused, as opposed to where they thought it was focused.

6.2 Temporal Visualization

IPSViz allows users to explore an H-VH interaction temporally through nonlinear review and scalable timelines.

Nonlinear Review Students can play back the H-VH interaction nonlinearly. While in playback mode, the visualizations (video, 3D rendering, and timeline plots) are updated to present information relevant to the current playback time. The audio of the interaction is also played back so that students can hear themselves talk to the VH. Similar in principle to nonlinear video editors (e.g. Adobe Premiere), students select moments from a timeline to instantaneously play them back. This allows students to review the interaction in any order. The timeline is represented visually in the interface by a slider (Figure 2 - bottom). Additionally, events can be selected from a transcript (Figure 2 - right) to review them.

Scalable Timelines Interaction events and signals are depicted visually on timelines. This allows users to see the temporal relationships among events. Timelines are scalable. At the global time scale, users review the entire interaction at a glance. Local features are narrowed in on to review them in detail. This is similar in spirit to Shneiderman's mantra: "Overview first, zoom and filter, then details-on-demand" [26]. The topic (Figure 5) and body lean plots (Figure 4) demonstrate the use of scalable timelines.

6.3 Social Visualization

IPSViz highlights social aspects of the interaction by visualizing verbal and nonverbal communication.

6.3.1 Verbal Communication

Verbal communication is presented explicitly in IPSViz by playing back the audio of the interaction. The transcript (Figure 2 - right) and topic plot (Figure 5) provide text and graphical representations of verbal communication.

Transcript The transcript displays everything the VH and student said to each other. By reviewing the transcript, the student learns what information the VH gave him and what important information he did not get from the VH. The transcript also serves as a table of contents into the interaction in that it helps users find important events to review.

Topic Plot The topic plot is an overview of the conversation. It filters the speech in the conversation down to a set of topics relevant to the scenario (e.g. symptoms, pain) and summarizes the use of these topics by plotting them on a timeline. Reviewing the topic plot allows students to see if they progressed from topic to topic logically, or if the flow of the conversation was confusing. They can also see if they forgot to discuss an important topic.

6.3.2 Nonverbal Communication

Nonverbal communication is represented in IPSViz through plots, the 3D-rendered motions of the VH, student, and gaze target, and through the video playback.

Posture Posture is used in interpersonal interaction to communicate friendliness, interest, and social status [20]. The student can review their posture in IPSViz by watching himself on video, looking at the body lean of the 3D model that represents them, or by looking for peaks and valleys on the body lean plot.

Paralanguage Paralanguage is the set of nonverbal vocal expressions (non-speech), such as tone of voice, volume and length of pauses. [13]. Students can review their paralanguage by hearing their voice when they play back the interaction.

Gaze Appropriate gaze builds rapport by communicating attention, friendliness, and respect [1]. The gaze target highlights where the student was looking. Also, the head pose of the student model provides another indicator of gaze behavior. Rendering from the VH's viewpoint allows the student to see how a patient interprets the student's head motion. Reviewing the interview from the VH's viewpoint shows students that patients notice gaze behavior and know when they are not paying attention.

7 Study Design

A study was conducted to determine how AAR with IPSViz affects perceptions of an H-VH interaction. Participants (N=27) in the study were health professions (medical and physician assistant) students. Each participant conducted a medical interview with a VH and then reviewed the experience in IPSViz. Before and after using IPSViz, the participant graded her interaction on how she appeared to the VH (e.g., friendly, emotional), the emotional state of the VH (e.g., scared), and patient communication skills (information gathering, rapport, and procedure). Students were also interviewed to gather comments about what they learned from using IPSViz. This is a within subjects, repeated measures design with two levels (Pre-AAR and Post-AAR). This design enabled measuring how IPSViz changed participants' views of their interaction with the VH.

In addition, two separate studies were conducted concurrently to evaluate the effect of VH skin tone (light vs. dark) and system display type (head-mounted vs. projection display) on the H-VH interaction. VH skin tone and display type are between-subjects factors that were varied randomly among the N=27 participants. The between-subjects results from these studies are reported elsewhere (in submission). While no cross interactions between VH skin tone, display type and AAR were expected, these factors were taken into account in the analysis as a precaution.

7.1 Participants

23 third-year medical and 4 first-year physician assistant students were recruited (15 male, 12 female). All participants were students at the Medical College of Georgia. Participants were paid $20.

7.2 Study Procedure

Figure 7 summarizes the study procedure.

Pre-Virtual Human Interaction Before interacting with the VHs, the participant signed a consent form and filled out a background survey. Students in the HMD subgroup wore an HMD tracked with 6 DOF. Students in the projector subgroup wore a hat tracked with 6 DOF. A wireless microphone was attached to the HMD and hat so that participants could talk to the VH. Students spent two minutes training the speech recognition software to understand her voice.

Virtual Human Interactions Participants conducted two interactions with VHs, a practice interaction and a main interaction. In the practice interaction, the VH was DIANA (Digital Animated Avatar), a 19-year-old female complaining of stomach pain. A study proctor began the practice with a tutorial on communicating with the VH. Then participants interviewed DIANA. The proctor remained in the room to assist the user. The practice interaction was no more than five minutes long.

In the main interaction, the VH was EDNA (Elderly DIANA), a 55-year-old woman who found a mass in her breast. Participants were instructed to interview EDNA as if she were a real patient. This time the study proctor did not enter the room with the student, and the student was expected to interact with EDNA without any assistance. The main interaction was between five and ten minutes long.

Interviewing a patient who found a mass in her breast is a challenging situation for an HP student. We chose to amplify this challenge and increase the realism of the scenario by having the VH challenge the student at three separate points in the interview.

- **Challenge 1:** The VH said "I found a mass in my breast, and uh, I'm really worried about it."
- **Challenge 2:** The VH said: "Could this be cancer?"
- **Challenge 3:** The VH sneezed.

In a real medical interview, the first two challenges should elicit empathetic, comforting responses from the student, and the last challenge should elicit a socially appropriate response (e.g., "bless you" or "gesundheit"). Students reviewed these challenges in IPSViz to evaluate if they responded appropriately.

Pre-AAR Survey After the main interaction, students filled out surveys to gather their perceptions of their performance in the main VH interview. The surveys are discussed in Section 7.3.

IPSViz Review Sessions Students conducted a practice AAR and a main AAR using IPSViz. In the practice AAR, students reviewed their practice VH interaction. A study proctor gave a five-minute tutorial on IPSViz, and then students were given a few minutes to become familiar with the IPSViz interface.

In the main AAR, students used IPSViz to review the main VH interaction. First, students reviewed the interaction on their own. They were encouraged to use IPSViz any way they wanted, including changing 3D viewpoints, reviewing plots, and skipping to important events. The study proctor did not interact with participants during this time. When the student was done, the study proctor conducted a short, guided review of the interaction. The guided review directed students to review their reactions to the three challenge statements (as described previously). The main AAR took approximately 10 minutes.

Post-AAR Survey and Debriefing After using IPSViz, students filled out a survey to learn how perceptions of their performance changed due to the AAR. Finally, students were debriefed to collect qualitative data about the H-VH experiences and IPSViz.

7.3 Measures

Two types of data were collected from participants, background data (e.g. demographics, educational background) and self-evaluations of the H-VH interaction. The background survey was filled out before the VH practice interaction. The self-evaluation surveys were filled out before and after using IPSViz.

7.3.1 Self-Evaluation Survey

Friendly, Natural, and Emotional Expression In previous studies, some participants asked the VH questions in a stilted, emotionless, rapid-fire manner [23]. To assess if participants recognized this improper behavior during AAR, participants evaluated themselves on naturalness, expression of emotion, and friendliness. Five "friendliness" scales were used - pleasant, cruel, cold, unfriendly, and unlikable (seven point Likert Scales, 1 = Not at all, 4 = Neutral, and 7 = Very). The friendliness measures have high internal consistency for real human interactions ($\alpha > 0.9$) [7].

State of the Virtual Human In addition to helping patient medically, doctors often need to help their patients emotionally. Hence, being able to read the emotional state of a patient is an important skill. To assess how well students read the VH's emotional state, participants ranked how scared and how friendly they thought the VH was (scale: 1 - 100, 1 = very little, 100 = very much).

Medical Interview Communication Participants evaluated their medical interview communication with the VH by filling out a survey of twenty five questions. Participants were familiar with the survey as it is commonly used by their instructors to evaluate their interviews with real patients. The survey's continued use by medical instructors indicates it is a reliable evaluation metric for this interpersonal scenario.

Questions from the survey are divided into three categories, information gathering, building rapport with the VH, and procedural aspects of the interview. Sample information gathering questions were "Rate how well you found out all complaints" and "Rate how well you elicited the story and meaning as well as biomedical facts." Sample rapport-building questions were "Did you legitimize the patient's ideas and feelings?" and "Did you use appropriate eye contact?" Samples procedure questions were "Did you use the patient's

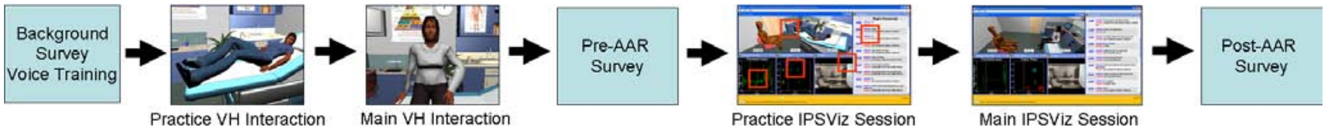

Figure 7: The Study Procedure

name appropriately?" and "Rate how well you began with open-ended questions and moved to closed-ended questions." Questions were rated on four or five point Likert scales (Never, Rarely, Sometimes, Usually, Almost Always or Poor, Fair, Good, Excellent).

8 Study Results and Discussion

8.1 After-Action Review Improves Self-Awareness

Survey results show that after action review changed participant perceptions of their interaction with the VH. Participants indicated they were less friendly ($F_{1,19} = 17.7$, $p < .001$, Pre-AAR: $M = 5.5, S_E = .19$, Post-AAR: $M = 4.6, S_E = .25$) and less natural ($F_{1,19} = 17.1$, $p < .001$, Pre-AAR: $M = 3.8, S_E = .27$, Post-AAR: $M = 2.7, S_E = .28$) towards the virtual human than they remembered. On 15 separate measures of rapport-building (e.g. non-verbal behavior, listening, and sensitivity) users indicated their rapport with the VH was worse after the AAR ($F_{1,19} = 18.4$, $p < .001$, Pre-AAR: $M = 2.54, S_E = .11$, Post-AAR: $M = 2.18, S_E = .11$) . The AAR also changed the way participants interpreted the state of the VH. The VH was perceived as being more scared after AAR ($F_{1,19} = 4.1$, $p < .06$, Pre-AAR: $M = 69, S_E = 5.0$, Post-AAR: $M = 76, S_E = 2.8$) . No differences were found on ratings of the user's emotional expression and on perceptions of the VH's friendliness.

The effect of AAR on perceptions of information gathering and procedural skills was not as clear. After AAR, participants rated themselves lower on "finding out all the patient's complaints" ($F_{1,19} = 18.3$, $p < .001$, Pre-AAR: $M = 2.4, S_E = .15$, Post-AAR: $M = 2.0, S_E = .12$) and "beginning with open-ended questions and moving to closed-ended questions" ($F_{1,19} = 4.87$, $p < .04$, Pre-AAR: $M = 1.8, S_E = .18$, Post-AAR: $M = 1.5, S_E = .17$) . In contrast, differences on "gathering the patient's story as well as biomedical facts" and "using the patient's name appropriately", were small and not statistically significant.

Participant comments support the notion that AAR helped users gain awareness of how they behaved in the VH interview. One participant said, "I like this part [IPSViz] ... actually seeing it. Seeing where my eyes were. Seeing like, having [the transcript] written out for me - she did this, and then I responded this way. It helps a lot more to know how you did."

Students mentioned specific improper behaviors that IPSViz helped them become aware of.

Gaze: "I didn't realize how much I was kinda looking around. Like, I switched it to her point of view and I was kind of looking all around instead of looking at her."

Nervous Behavior: "I think that I learned that kinda shifting in the chair can be easily recognized by the patient as a sign of maybe discomfort or something like that."

Response to Empathy Challenges: "From looking at the eye contact, I think that was an appropriate level of nonverbal stuff, but I didn't give any verbal empathetic responses on reflection, on looking back at the playback."

Information Gathering: "[It was] interesting to see that I just barely hit some of the required or desired topics - and I didn't really stick with anything for that long."

These comments highlight that the visualizations provided by IPSViz were key in making students more self-aware. The 3D rendering of the interaction - including the gaze target, the user model, and the patient's viewpoint - helped students understand how their gaze, posture, and nervous behavior is interpreted by patients. The transcript and topic plot helped students recall the structure and content of the interview, as well as determine what questions they forgot to ask the VH. Finally, video and audio rounded out the feedback provided by the other visualizations by allowing users to watch and hear their inappropriate responses to the VH.

8.2 Changing Communication with Real Humans

Students reported they would change behavior in future interviews with real patients based on the AAR with IPSViz.

Gaze: "Definitely, it will make me be more conscious of where I'm looking when I'm talking to a patient."

Response to Empathy Challenges: "I guess being more mindful about what I'm actually saying. Because when she mentioned breast mass I went directly to, you know, where is it, you know, tell me about the breast mass, rather than that must be concerning to you. And I think I learned from that."

Overall Behavior: "It was good to watch myself. See how I act, and to hear my voice, how I ask questions, the [in]tonation in my voice. And after doing it I can think about it more, you know, things I could have done. And then when I go back and view, I can see ways I could have changed."

The notion that an H-VH experience should impact future interactions with real humans is a surprising result of the study. Given the technological limitations of the system (e.g. speech recognition and understanding), it was not clear that users would see the VH experience as correlating to interactions with real patients. **AAR served as a catalyst to connecting interviews with the VH to interviews with real patients.**

8.3 AAR, VH Skin Tone, and Display type

As mentioned previously, two separate studies were run concurrently with the AAR study to evaluate the effect of display type and VH skin tone on the interaction. This section discusses interaction effects between AAR, skin tone and display type.

Display type did not affect the AAR responses, but an intriguing cross-interaction between VH skin tone and AAR was found. As shown in Figure 8, the difference due to AAR on "showing interest in the VH" (F(1,19) = 6.9, p = .02, η = .27) was affected by the VH's skin tone. After AAR, students who talked to a dark-skinned VH lowered their score on showing interest to the VH. Students who interacted with a light-skinned VH did not change their score.

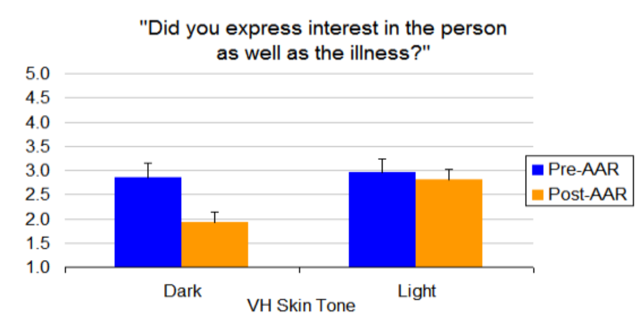

Figure 8: The interaction of VH skin tone and after-action review on showing interest in the VH.

The cross interaction between the VH Skin and AAR conditions shows that participants considered the skin tone of the VH when evaluating their behavior with IPSViz. They evaluated themselves differently on this measure depending on whether they interacted with a dark or light-skinned VH. This implies participants (consciously or subconsciously) perceived that the skin color of the dark-skinned VH biased their interaction. The discovery of one's biases is a first step towards changing those biases. We are currently exploring this interplay between after-action review and many types of biases (e.g. skin tone, ethnicity, religion, weight, age) with the goal of building virtual human experiences for diversity training.

9 Conclusions and Future Work

A fundamental question of VH research is - **to what extent (if at all) can we change or affect people using an interaction with a virtual human?** This work argues that we *can* change people by combining H-VH interactions with AAR. Guidelines for AAR for H-VH interactions and IPSViz, a tool that generates visualizations for AAR, were presented. In addition, a study demonstrated that combining H-VH interaction with AAR, at least in the short term, changes the way a user sees himself (self-identity and self-awareness). More importantly, user comments indicate this change may extend into future interactions with real humans.

The next step is to learn if using IPSViz results in measurable improvement in H-VH and eventually H-H communication skills. This will be evaluated through repeated exposures to VHs followed by AAR with IPSViz. This experience-feedback loop will enable conclusively evaluating if student comments of AAR changing their perspective is a realizable goal.

IPSViz is also being applied to two other groups, communication skills educators and VH researchers. We are developing a new tool, IPSViz^n, to characterize H-VH interactions in the aggregate. Instead of visualizing a single interaction at-a-time, IPSViz^n visualizes n interactions. For example, students can generate a portfolio of H-VH interactions for student and educator review of progression of communication skills; educators can efficiently evaluate class performance on an interaction by reviewing class-wide visualizations; researchers can review how manipulating properties of the virtual human affects users.

IPSViz represents an important progression in the application of VEs. Traditionally, the impact of a virtual experience has been viewed as a product of the experience itself - the VE is what impacts the user and makes the experience valuable. IPSViz extends this to include review and self-reflection. By enabling review and self-reflection, the VH experience continues to impact the student beyond the conclusion of the experience itself. This impact is different than that provided by the experience itself. With AAR, users have an opportunity for self-reflection, insight, and self-directed change for real-world social interactions.

Acknowledgements

The authors thank Dr. D. Scott Lind and Dr. Adeline Deladisma for access to facilities at the Medical College of Georgia and assistance in recruiting study participants. They thank Dr. Rick Ferdig for his assistance in developing the study design. They thank Kyle Johnson, Aaron Kotranza, and Brent Rossen for their assistance in running the studies. They thank Joshua Horton for his work on the topic plotter. This work was supported by NSF Grant IIS-0643557 and the University of Florida Alumni Graduate Fellowships.

References

[1] M. Argyle and M. Cook. *Gaze and Mutual Gaze*. Cambridge University Press, UK, 1976.

[2] S. Babu, T. Barnes, and L. Hodges. Can immersive virtual humans teach social conversational protocols? *IEEE International Conference on Virtual Reality 2007*, pages 215–218, March 2007.

[3] J. L. Coulehan and M. R. Block. *The Medical Interview: Mastering Skills for Clinical Practice*. F.A. Davis Co., Philadelphia, 3rd edition, 1997.

[4] S. Craig, B. Gholson, and D. Driscoll. Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology*, 94(2):428–434, 2002.

[5] E. Deaton, C. Barba, T. Santarelli, L. Rosenzweig, V. Souders, C. McCollum, J. Seip, W. Knerr, and J. Singer. Virtual environment cultural training for operational readiness (vector). *Virtual Reality*, 8(3):156–167, 2005.

[6] Department of the Army, Washington, DC. *Training Circular 25-20: A Leader's Guide To After-Action Reviews*, September 1993.

[7] J. Dovidio, K. Kawakami, and S. Gaertner. Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, 82(1):62–68, 2002.

[8] Forterra. On-line interactive virtual environment (olive): A multi-user, virtual world software platform. Technical report, Forterra Systems, December 2006.

[9] G. Frank, C. Guinn, R. Hubal, P. Pope, M. Stanford, and D. Lamm-Weisel. Just-talk: An application of responsive virtual human technology. In *In Proc. of the Interservice/Industry Training, Simulation and Education Conference*, Orlando, FL, December 2002.

[10] D. Friedman, A. Brogni, A. Antley, C. Guger, A. Steed, and M. Slater. Standardizing data analysis in presence experiments. *Presence: Teleoperators and Virtual Environments*, 15(5):599–610, October 2006.

[11] C. Greenhalgh, M. Flintham, J. Purbrick, and S. Benford. Applications of temporal links: Recording and replaying virtual environments. In *IEEE Virtual Reality 2002*, March 2002.

[12] J. C. Hart. On recording virtual environments. In *IEEE Symposium on Research Frontiers in Virtual Reality*, page 8083, October 1993.

[13] M. Hickson III, D. W. Stacks, and N.-J. Moore. *Nonverbal Communication: Studies and Applications*. Roxbury Publishing Company, Los Angeles, CA, 2004.

[14] R. W. Hill Jr., J. Gratch, S. Marsella, J. Rickel, W. Swartout, and D. Traum. Virtual humans in the mission rehearsal exercise system. *Kynstliche Intelligenz (KI Journal)*, 17(4), 2003.

[15] K. Johnsen, R. Dickerson, A. Raij, B. Lok, J. Jackson, M. Shin, J. Hernandez, A. Stevens, and D. S. Lind. Experiences in using immersive virtual characters to educate medical communication skills. In *IEEE Virtual Reality 2005*, Bonn, Germany, 2005.

[16] K. Johnsen, A. Raij, B. Lok, A. Stevens, and D. S. Lind. The validity of a virtual human experience for interpersonal skills education. In *ACM SIGCHI*, 2007.

[17] B. W. Knerr, D. R. Lampton, G. A. Martin, D. A. Washburn, and D. Cope. Developing an after action review system for virtual dismounted infantry simulations. In *The InterserviceIndustry Training, Simulation & Education Conference (I/ITSEC)*, 2002.

[18] J. Leigh, A. E. Johnson, T. A. DeFanti, M. Brown, and et al. A review of tele-immersive applications in the cave research network. In *IEEE Virtual Reality 1999*, pages 180–187, Houston, TX, March 1999.

[19] A. Manganas, M. Tsiknakis, E. Leisch, M. Ponder, T. Molet, B. Herbelin, N. Magnenat-Thalmann, D. Thalmann, M. Fato, and A. Schenone. The just vr tool: An innovative approach to training personnel for emergency situations using virtual reality techniques. *Journal on Information Technology in Healthcare*, 2(6):399–412, 2004.

[20] A. Mehrabian. *Silent messages: Implicit communication of emotions and attitudes*. Wadsworth, Belmont, California, 2nd edition, 1981.

[21] J. Morie, K. Iyer, D. Luigi, J. Williams, A. Dozois, and A. Rizzo. Development of a Data Management Tool for Investigating Multivariate Space and Free Will Experiences in Virtual Reality. *Applied Psychophysiology and Biofeedback*, 30(3):319–331, 2005.

[22] NaturalPoint. Optitrack software. `http://www.naturalpoint.com/optitrack/`, November 2007.

[23] A. Raij, K. Johnsen, R. Dickerson, B. Lok, M. Cohen, M. Duerson, R. Pauley, A. Stevens, P. Wagner, and D. S. Lind. Comparing interpersonal interactions with a virtual human to those with a real human. *IEEE Transactions on Visualization and Computer Graphics*, 13(3), May 2007.

[24] D. Roter, S. Larson, H. Shinitzky, R. Chernoff, J. Serwint, G. Adamo,

and L. Wissow. Use of an innovative video feedback technique to enhance communication skills training. *Medical Education*, 38:145–157(13), February 2004.

[25] K. Ryokai, C. Vaucelle, and J. Cassell. Virtual peers as partners in storytelling and literacy learning. *Journal of Computer Assisted Learning*, 19(2):195–208, 2003.

[26] B. Shneiderman. The eyes have it: A task by data type taxonomy for information visualizations. In *IEEE Visual Languages*, pages 336–343, College Park, Maryland 20742, U.S.A., 1996.

[27] P. D. Sloane, R. Beck, V. Kowlowitz, A. M. Blotzer, L. Wang, L. Akins, F. White-Chu, and C. M. Mitchell. Behavioral coding for evaluation of medical student communication: Clarification or obfuscation? *Academic Medicine*, 79(2):162–70, February 2004.

[28] R. Smith and G. Allen. After action review in military training simulations. In *WSC '94: Proceedings of the 26th conference on Winter simulation*, pages 845–849, San Diego, CA, USA, 1994.

[29] A. Steed, A. Brogni, and V. Vinayagamoorthy. Breaks in presence as usability criteria. In *Proc. of HCI International 2005, The 11th International Conference on Human Computer Interaction*, July 2005.