CHAPTER 3

# Autonomous Virtual Human Agents for Healthcare Information Support and Clinical Interviewing

**Albert Rizzo, Russell Shilling, Eric Forbell, Stefan Scherer, Jonathan Gratch and Louis-Philippe Morency**
University of Southern California Institute for Creative Technologies, Los Angeles, CA, USA

## INTRODUCTION

A virtual revolution is ongoing in the use of simulation technology for clinical purposes. When discussion of the potential use of virtual reality (VR) applications for human research and clinical intervention first emerged in the early 1990s, the technology needed to deliver on this "vision" was not in place. Consequently, during these early years VR suffered from a somewhat imbalanced "expectation-to-delivery" ratio, as most users trying systems during that time will attest. Yet it was during the "computer revolution" in the 1990s that emerging technologically driven innovations in behavioral health care had begun to be considered and prototyped. Primordial efforts from this period can be seen in early research and development (R&D) that aimed to use computer technology to enhance productivity in patient documentation and record-keeping, to deliver cognitive training and rehabilitation, to improve access to clinical care via Internet-based teletherapy, and in the use of VR simulations to deliver exposure therapy for treating specific phobias. Over the last 20 years the technology required to deliver behavioral health applications has significantly matured. This has been especially so for the core technologies needed to create VR systems where advances in the underlying enabling technologies (e.g., computational speed, 3D graphics rendering, audio/visual/haptic displays, user interfaces/tracking, voice recognition, artificial intelligence, and authoring software) have supported the creation of low-cost, yet sophisticated VR systems capable of running on commodity-level personal computers and mobile devices. In part driven by digital gaming and

*Artificial Intelligence in Behavioral and Mental Health Care.*
DOI: http://dx.doi.org/10.1016/B978-0-12-420248-1.00003-9

entertainment sectors, and a near insatiable global demand for mobile and networked consumer products, such advances in technological "prowess" and accessibility have provided the hardware and software platforms needed to produce more usable, hi-fidelity VR scenarios for the conduct of human research and clinical intervention. Thus, evolving behavioral health applications can now usefully leverage the interactive and immersive assets that VR affords as the technology continues to get faster, better, and cheaper moving into the twenty-first century.

While such advances have now allowed for the creation of ever more believable context-relevant "structural" VR environments (e.g., combat scenes, homes, classrooms, offices, markets), the next stage in the evolution of clinical VR will involve creating virtual human (VH) representations that can engage real human users in believable and useful interactions. This emerging technological capability has now set the stage for the next major movement in the use of VR for clinical purposes with the "birth" of intelligent VH agents that can serve the role of virtual interactors for a variety of clinical purposes. VHs can be used to populate fully immersive virtual reality environments or be delivered on a standard nonimmersive display for user interaction. Such VH conversational agents, whether located within a virtual reality context or simply as a pop up on a computer monitor, provide new opportunities in the area of healthcare support and advice. Imagine a user going online and interacting with a VH "coach" who can answer questions about clinical issues, discuss treatment options, help you to locate a treatment center suited to your needs, guide you through self-assessment questionnaires, and generally provide a face-to-face human-like interaction to support healthcare awareness and access. Moreover, a VH could be waiting for you to appear at a clinic and, within the space of a private kiosk, be able to conduct an initial clinical interview with you to gather information that may useful to start the early steps for creating a treatment plan that would be followed on by a live clinician. At the same time, this kiosk-based VH agent can leverage advances in sensing technology (e.g., cameras and microphones) to "observe" your facial/body gestures and vocal features and use those sensed signals to infer your psychological state from these observable behaviors to enhance the interaction or document mental status over time. These types of VH interaction possibilities are not visions anticipated to occur in a far-away time later in the twenty-first century — they in fact exist today either as early production prototypes or are on the verge of escaping the laboratory following an appropriate course of research to determine their value and to better understand the ethical boundaries for their use.

The emphasis of this chapter is on the use of such autonomous VH conversational agents to support client interaction within a healthcare context. These forms of VH agents, armed with various levels of artificial intelligence, can play a role in healthcare support in a fashion that does not aim to replace well-trained care providers. Rather, they can be designed to fill in the gaps where the economics are such that there is not now or ever likely to be a real human available to fill these roles. After a brief discussion of the history of VHs in the clinical VR domain, this chapter will detail our applications in this area where a VH can provide private online healthcare information and support (i.e., *SimCoach*) and where a VH can serve the role as a clinical interviewer (i.e., *SimSensei*). Our work using intelligent VH agents in the role of virtual patients for clinical training and for other purposes can be found elsewhere (Rizzo et al., 2011; Swartout et al., 2013; Talbot, Sagae, John, & Rizzo, 2012).

## THE RATIONALE AND BRIEF HISTORY OF THE CLINICAL USE OF VHS

Recent shifts in the social and scientific landscape have now set the stage for the next major movement in clinical VR with the "birth" of intelligent VHs. It is now technically feasible to create VH representations that are capable of fostering believable interaction with *real* VR users. This capability has been around since the 1990s, but the previous limitations in graphic rendering, natural language processing, voice recognition, and face and gesture animation made the creation of credible VHs for interaction a costly and labor-intensive process. Thus, until recently VHs existed primarily in the domain of high-end special effect studios/labs that catered to the film or game industry, far from the reach of those who thought to employ them in clinical health applications.

This is not to say that representations of human forms have not usefully appeared in clinical VR scenarios. In fact, since the mid-1990s, VR applications have routinely employed "primitive" VHs (e.g., low-fidelity graphics, nonlanguage interactive, limited face and gesture expression) to serve as stimulus elements to enhance the realism of a virtual world simply by their static presence. For example, VR exposure therapy applications for the treatment of specific phobias (e.g., fear of public speaking, social phobia) were successfully deployed using immersive simulations that were inhabited by "still-life" graphics-based characters or 2D photographic sprites (i.e., static full-body green-screen-captured photo images of a person) (Anderson, Zimand, Hodges, & Rothbaum, 2005; Klinger, 2005; Pertaub, Slater, & Barker, 2002). By simply adjusting the number and location of these VH representations, the intensity of these anxiety-provoking VR contexts could be modulated systematically with the aim of gradually habituating phobic patients to what they feared, leading to improved functioning in the real world with real people. In spite of the primitive nature of these VHs, phobic clients appeared to be especially primed to react to such representations and thus, they provided the necessary stimulus elements to be effective in this type of exposure-based cognitive behavioral treatment.

Other clinical applications have also used animated graphic VHs as stimulus entities to support and train social and safety skills in persons with high-functioning autism (Padgett, Strickland, & Coles, 2006; Parsons et al., 2012; Rutten et al., 2003) and as distracter stimuli for attention assessments conducted in a virtual classroom (Parsons, Bowerly, Buckwalter, & Rizzo, 2007; Rizzo et al., 2006). Additionally, VHs have been used effectively for the conduct of social psychology experiments, essentially replicating and extending findings from studies conducted with real humans on social influence, conformity, racial bias, and social proxemics (Bailenson & Beall, 2006; Blascovich et al., 2002; McCall, Blascovich, Young, & Persky, 2009).

In an effort to further increase the visual realism of such VHs for phobia treatment, Virtually Better Inc. (www.virtuallybetter.com), began incorporating whole video clips of crowds into graphic VR scenarios for therapeutic exposure with clients suffering from fear of public speaking. They later advanced the technique by using green-screen-captured dynamic video sprites of individual humans inserted into graphics-based VR social settings for social phobia and for cue-exposure substance abuse treatment and research applications (Bordnick, Traylor, Carter, & Graap, 2012). A large library of green-screen-captured video sprites of actors behaving or speaking with varying degrees of provocation could then be strategically inserted into the scenario with the aim of modulating the emotional state of the client by fostering encounters with these 2D video VH representations.

The continued quest for even more realistic simulated human interaction contexts led other researchers to use panoramic video capture (Macedonio, Parsons, Wiederhold, & Rizzo, 2007; Rizzo, Ghahremani, Pryor, & Gardner, 2003) of a real-world office space inhabited by hostile co-workers and supervisors to produce VR scenarios for anger management research. With this approach, VR scenarios were created using a 360-degree panoramic camera that was placed in the position of a worker at a desk. Actors then walked into the staged workspace, addressed the camera (as if it was the targeted user at work) and proceeded to verbally threaten and abuse the camera, vis-à-vis, the worker. Within such photorealistic scenarios, VH video stimuli could deliver intense emotional expressions and challenges and psychophysiological measures of users in these immersive video scenarios indicated increased arousal (i.e., skin conductance and heart rate) indicative of an emotional reaction (Macedonio et al., 2007). However, the number of clips available was limited and while the initial work demonstrated the capacity of such stimuli to activate nonclinical users, the system was never tested with actual anger management clients to determine their value in supporting the role-playing practice of more appropriate coping responses.

While others have similarly employed video-captured human content, particularly for commercial clinical training systems (cf. Simmersion, 2015), the use of such fixed video content to foster some level of *faux* interaction or exposure has significant limitations. For example, a video capture method requires the creation of a large catalog of all the possible relevant verbal and behavioral clips to support their tactical presentation to users to meet the requirements of a given therapeutic or training approach. This can be costly and time-consuming and requires significant forethought as to what needs to be captured in order to provide the necessary coverage of a specific domain (e.g., all possible settings where a user might avoid a social interaction or be engaged in a hostile conversation). This requires that a developer capture many clips, test with users, and determine what is actually needed to round out the application after initial testing with relevant users. Since this fixed content cannot be readily updated in a dynamic fashion, new clips need to be captured with the same actors in the same setting with the same clothes under the same lighting conditions. This method, while providing high-fidelity realism, is ultimately limited by its lack of flexibility in terms of upgrading or evolving the application as new information becomes available. Moreover, this process is more suited to clinical applications where the only requirement is for the VH character to deliver an open-ended statement or question that the user can react to with a response, but is lacking in allowing for any truly fluid and believable interchange following a response by the user. Consequently, the absence of dynamic bidirectional interaction with these video representations without a live person behind the "screen" actuating new clips in response to the user's behavior is a significant limiting factor for this approach. This has led some researchers to consider the use of artificially intelligent autonomous VH agents as entities for simulating human-to-human interaction.

Clinical interest in artificially intelligent VH agents designed for interaction with humans can trace its roots to the work of MIT AI researcher, Joseph Weizenbaum. In 1966, he wrote a language analysis program called ELIZA that was designed to imitate a Rogerian therapist. The system allowed a computer user to interact with a virtual therapist by typing simple sentence responses to the computerized therapist's questions. Weizenbaum reasoned that simulating a nondirectional psychotherapist was one of the easiest ways of simulating human verbal interactions and it was a compelling simulation that worked well on teletype computers (and is even instantiated on the Internet today; http://www-ai.ijs.si/eliza-cgi-bin/eliza_script). In spite of the fact that the illusion of ELIZA's intelligence soon disappears due to its inability to handle complexity or nuance, Weizenbaum was reportedly shocked upon learning how seriously people took the ELIZA program (Howell & Muller, 2000). This led him to conclude that it would be immoral to substitute a computer for a human function that "... involves interpersonal respect, understanding, and love" (Weizenbaum, 1976). While Weizenbaum's sentiment is understandable coming from the era that he worked on this, modern approaches to foster an interaction with a VH for a useful purpose but where no human is available, does not fit with the "substitute" concept that he specified.

More recently, seminal research and development has appeared in the creation of highly interactive, artificially intelligent and natural language capable VH agents. No longer at the level of a prop to add context or minimal faux interaction in a virtual world, these agents are designed to perceive and act in a 3D virtual world, engage in face-to-face spoken dialogues with real users (and other VHs) and in some cases, they are capable of exhibiting human-like emotional reactions. Previous classic work on VHs in the computer graphics community focused on perception and action in 3D worlds, but largely ignored dialogue and emotions. This has now changed. Intelligent VH agents are now being created that control computer-generated bodies and can interact with users through speech and gesture in virtual environments (Gratch et al., 2002, 2013). Advanced VHs can engage in rich conversations (Traum et al., 2008), recognize nonverbal cues (Morency, de Kok, & Gratch, 2008; Rizzo et al., 2014; Scherer et al., 2014), reason about social and emotional factors (Gratch & Marsella, 2004), and synthesize human communication and nonverbal expressions (Thiebaux et al., 2008).

Such fully embodied conversational characters have been around since the early 1990s (Bickmore & Cassell, 2005) and there has been much work on full systems that have been designed and used for training (Kenny, Rizzo, Parsons, Gratch, & Swartout, 2007; Prendinger & Ishizuka, 2004; Rickel, Gratch, Hill, Marsella, & Swartout, 2001; Talbot et al., 2012), intelligent kiosks (McCauley & D'Mello, 2006), and virtual receptionists (Babu et al., 2006). Both in appearance and behavior, VHs have now passed through "infancy" and are ready for service in a variety of clinical and research applications.

These advances in VH technology have now supported our research and development for clinical VR applications in two key domains: (i) VH healthcare support agents (i.e., SimCoach) that serve as online guides for promoting anonymous access to psychological healthcare information and self-help activities. (ii) VH agents that can serve the role as a clinical interviewer while sensing behavioral signals (e.g., face, body, and vocal parameters) that can be used by the agent to infer user state, update their behavior in real time, and quantify these behaviors over the course of a 20-min interview (i.e., SimSensei). These areas are detailed below.

## USE CASES: SIMCOACH AND SIMSENSEI

### SimCoach: A VH Agent to Support Healthcare Information Access

The inspiration for the development of the SimCoach project came from research suggesting that there was an urgent need to reduce the stigma of seeking mental health treatment in military service member (SM) and veteran populations. One of the more foreboding findings in an early report by Hoge et al. (2004) was the observation that among Iraq/Afghanistan War veterans, "*. . . those whose responses were positive for a mental disorder, only 23 to 40 percent sought mental health care. Those whose responses were positive for a mental disorder were twice as likely as those whose responses were negative to report concern about possible stigmatization and other barriers to seeking mental health care*" (p. 13). While US military training methodology has better prepared soldiers for combat in recent years, such hesitancy to seek treatment for difficulties that emerge upon return from combat, especially by those who may need it most, suggests an area of military mental health care that is in need of attention. Moreover, the dissemination of healthcare information to military SMs, veterans, and their significant others is a persistent and growing challenge. Although medical information is increasingly available over the web, users can find the process of accessing it to be overwhelming, contradictory, and impersonal.

In spite of a Herculean effort on the part of the US Department of Defense (DOD) to produce and disseminate behavioral health programs for military personnel and their families, the complexity of the issues involved continues to challenge the best efforts of military mental healthcare experts, administrators, and providers. Since 2004, numerous blue ribbon panels of experts have attempted to assess the current DOD and Veterans Affairs (VA) healthcare delivery system and provide recommendations for improvement (DOD Mental Health Task Force (DOD, 2007), National Academies of Science Institute of Medicine (IOM, 2007, 2012), Dole-Shalala Commission Report (Dole et al., 2007), the Rand Report (Tanielian et al., 2008), American Psychological Association (APA, 2007)).

For example, the American Psychological Association Presidential Task Force on Military Deployment Services for Youth, Families and Service Members (APA, 2007) presented their preliminary report that poignantly stated that they were, "... not able to find any evidence of a well-coordinated or well-disseminated approach to providing behavioral health care to service members and their families." The APA report also went on to describe three primary barriers to military mental health treatment: *availability, acceptability, and accessibility.* More specifically:

1. Well-trained mental health specialists are not in adequate supply (*availability*)
2. The military culture needs to be modified such that mental health services are more *accepted* and less stigmatized
3. Moreover, even if providers were available and seeking treatment was deemed acceptable, appropriate mental health services are often not readily *accessible* due to a variety of factors (e.g., long waiting lists, limited clinic hours, a poor referral process and geographical location).

In addition to problems with well-trained provider *availability, access* to care from distant locations and the *acceptance* of treatment as a viable solution by a segment of the population who have traditionally viewed asking for help as a sign of weakness, barriers with regard to the awareness of and anticipated benefit from treatment options complicate optimal care provision. In essence, new methods are needed to reduce such barriers to care.

The SimCoach project was designed to address this challenge by supporting users in their efforts to anonymously seek healthcare information and advice by way of online interaction with an intelligent, interactive, embodied VH healthcare guide. The primary goal of the SimCoach project is to break down barriers to care by providing military SMs, veterans, and their significant others with confidential help in exploring

**Figure 3.1** Sample SimCoach Archetypes.

and accessing healthcare content and, if needed, for encouraging and supporting the initiation of care with a live provider. Rather than being a traditional web portal, SimCoach allows users to initiate and engage in a dialog about their healthcare concerns with an interactive VH. Generally, these intelligent graphical characters are designed to use speech, gesture, and emotion to introduce the capabilities of the system, solicit basic anonymous background information about the user's history and clinical/psychosocial concerns, provide advice and support, present the user with relevant online content and potentially facilitate the process of seeking appropriate care with a live clinical provider. An implicit motive of the SimCoach project is that of supporting users who are determined to be in need, to make the decision to take the first step toward initiating psychological or medical care with a live provider.

However, it is not the goal of SimCoach to breakdown all barriers to care or to provide diagnostic or therapeutic services that are best delivered by a live clinical provider. Rather, SimCoach was designed to foster comfort and confidence by promoting users' private and anonymous efforts to understand their situations better, to explore available options, and initiate treatment when appropriate. Coordinating this experience is a VH SimCoach, selected by the user from a variety of archetypical character options (see Figure 3.1), who can answer direct questions and/or guide the user through a sequence of user-specific questions, exercises, and assessments. Also, interspersed within the program are options that allow the user to respond to simple screening instruments, such as the PCL-M (PTSD symptom checklist) that are delivered in a conversational format with results fed back to the user in a supportive fashion. These screening results serve to inform the SimCoach's creation of a model of the user to enhance the reliability and accuracy of the SimCoach output to the user, to support user self-awareness via feedback, and to better guide the delivery of relevant information based on this self-report data. Moreover, an enhancement in user engagement with a SimCoach was thought to occur if a more accurate assessment of the user's needs is derived from this process to inform the relevancy of the interaction. This interaction between the VH and the user provides the system with the information needed to guide users to the appropriate next step of engagement with the system or with encouragement to initiate contact with a live provider. Again, the SimCoach project was not conceived as a replacement for human clinical providers and experts. Instead, SimCoach was designed to start the process of engaging the user by providing support and encouragement, increasing awareness of their situation and treatment options, and in assisting individuals who may otherwise be initially uncomfortable talking to a live care provider.

Users can flexibly interact with a SimCoach character by typing text and clicking on character-generated menu options. Since SimCoach was designed to be an easily accessible web-based application that requires no downloadable software, it was felt that voice recognition was not at a state where it could be reliably used at the start of the project in 2010. The feasibility of providing the option for full spoken natural language dialog interaction is currently being explored to determine whether off-the-shelf voice recognition programs are sufficiently accurate to maintain an engaged interaction between a SimCoach and a user. The options for a SimCoach's appearance, behavior, and dialog have been designed to maximize user comfort and satisfaction, but also to facilitate fluid and truthful disclosure of clinically relevant information. Focus groups, "Wizard of OZ" studies, and iterative formative tests of the system were employed with a diverse cross section of our targeted user group to create options for SimCoach interaction that would be both engaging and useful for this population's needs. Results from these user tests indicated some key areas that were determined to be important including user-choice of character archetypes across gender and age ranges, informal dialog interaction, and interestingly, a preference for characters that were not in uniform.

Engagement is also supported by insuring that the specific healthcare content that a SimCoach can deliver to users is relevant to persons with a military background (and, of course, their significant others). This was addressed by leveraging content assets that were originally created for established DOD and VA websites specifically designed to address the needs of this user group (e.g., Afterdeployment, Military OneSource, National Center for PTSD). Our early research with this user group indicated a hesitancy to directly access these sites when users sought behavioral health information, with a common complaint being that there was a fear that their use of those sites may be monitored and might jeopardize advancement in their military careers or later applications for disability benefits (Rizzo et al., 2011). In spite of significant efforts by the DOD and VA to dispel the idea that user tracking was employed on these sites, the prevailing suspicion led many of the users in our samples to conduct such healthcare queries using Google, Yahoo, and Medscape. To address this user concern, supplemental content presented by the SimCoach (e.g., video, self-assessment questionnaires, resource links) is typically "pulled" into the site, rather than directing users away "to" those sites. Users also have the option to print out a PDF summary of the SimCoach session. This is important for later personal review and for the access to links to relevant web content that the SimCoach provided during the session. The summary may also be useful when seeking clinical care to enhance their comfort level, armed with knowledge from the SimCoach interaction, when dealing with human clinical care providers and experts.

As the system evolves, it is our view that engagement would be enhanced if the user was able to interact with the SimCoach repeatedly over time. Ideally, users could progress at their own pace over days or even weeks as they perhaps develop a "relationship" with a SimCoach character as a "go-to" source of healthcare information and feedback. However, this option for evolving the SimCoach comfort zone with users over time would require significant database resources to render the SimCoach capable of "remembering" the information acquired from previous visits and to build on that information in similar fashion to that of a growing human relationship. Moreover, the persistence of a SimCoach memory for previous sessions would also require the user to sign into the system with a username and password. This would necessitate the SimCoach system to "reside" on a high-security server, such that content from previous visits could be stored and accessed with subsequent visits. Such functionality might be a double-edged sword as anonymity is a hallmark feature to draw in users who may be hesitant to know that their interactions are being stored, even if it resulted in a more relevant, less redundant and perhaps more meaningful interaction with a SimCoach over time. Likely, this would necessarily have to be a clearly stated "opt-in" function, as the technology may support this in the future. Recent developments with the SimCoach platform now allow for database access in a separate SimCoach-based project. This will allow for a SimCoach to support "opted-in" users to access their Electronic Medical

Records in applications where the SimCoach serves as a front-end interface for VA or other medical clinic-related uses. In essence, the SimCoach architecture has evolved for application beyond the original intent and, in fact, an online SimCoach authoring tool has recently been created. This allows other clinical professionals to create SimCoach content to enhance and evolve delivery of other care perspectives with this platform (cf. https://authoring.simcoach.org/sceditor/landing.html).

Although this project represents an early (and ongoing iterative) effort in this area, it is our view that the initial clinical aims of breaking down barriers to care can be usefully addressed even with this initial version of the system. The current version of SimCoach is available online (http://www.simcoach.org/) and unidentified user interaction is regularly mined to continue to advance the development of the system by examining the spectrum of interactions that take place with actual users. We expect that SimCoach will continue to evolve over time based on data collected from these ongoing user interactions with the system and with advances in technology, particularly with improved voice recognition. Along the way, this work will afford many research opportunities for investigating the functional and ethical issues involved in the process of creating and interacting with VHs in a clinical or healthcare support context.

For more information, see a SimCoach Video at: https://www.youtube.com/watch?v = PGYUqTvE6Jo/.

## SimSensei: A VH Interviewing Agent for Detection and Computational Analysis of Psychological Signals

*SimSensei* is a VH interaction platform that is able to sense and interpret real-time audiovisual behavioral signals from users interacting with the system. The system was specifically designed for clinical interviewing and healthcare support by providing a face-to-face interaction between a user and a VH that can automatically react to the inferred state of the user through analysis of behavioral signals gleaned from the user's facial expressions, body gestures, and vocal parameters. User behavior is captured and quantified using a range of off-the-shelf sensors (i.e., webcams, Microsoft Kinect, and a microphone). Akin to how nonverbal behavioral signals have an impact on human-to-human interaction and communication, SimSensei aims to capture and infer from the user nonverbal communication to improve the engagement between a VH and a user. The system can also quantify and interpret sensed behavioral signals longitudinally to inform diagnostic assessment within a clinical context.

The development of SimSensei required a thorough awareness of the literature on emotional expression and communication. It has long been recognized that facial expressions and body gestures play an important role in human communicative signaling (Ekman & Rosenberg, 1997; Russell & Fernandez-Dols, 1997). Also, vocal characteristics (e.g., prosody, pitch variation, etc.) have also been reported to provide additive information regarding the "state" of the speaker beyond the actual language content of the speech (Pentland, Lazer, Brewer, & Heibeck, 2009). While some researchers postulate that the universal expression and decoding of face/body gestures and vocal patterns are indicative of genetic "hardwired" mammalian neural circuitry as Darwin proposed over a hundred years ago (Darwin, 2002), others have placed less emphasis on investigating underlying mechanisms and instead have focused on the empirical analysis of such implicit communication signals and what can be meaningfully derived from them. In the latter category, Pentland's MIT research group has characterized these elements of behavioral expression as "Honest Signals" (Pentland, 2008). Based on his research with groups of people interacting, he suggests: "*. . . this second channel of communication, revolving not around words but around social relations, profoundly influences major decisions in our lives—even though we are largely unaware of it.*" Pentland posits that the physical properties of this signaling behavior are constantly activated, not simply as a back channel or complement to our conscious language, but rather as a separate communication network. It is conjectured that these signaling behaviors, perhaps evolved from ancient primate nonverbal communication mechanisms, provide a useful window into our intentions, goals, values, and emotional state. Based on this perspective, an intriguing case can be made for the development of a computer-based sensing system that can capture and quantify such behavior, and from that activity data make inferences as to a user's cognitive and emotional state. Inferences from these sensed signals could then be used to supplement information that is garnered exclusively from the literal content of speech for a variety of purposes.

Recent progress in low-cost sensing technologies and computer vision methods has now driven this vision to reality. Indeed, recent widespread availability of low-cost sensors (webcams, Microsoft Kinect, microphones) combined with software advances for facial feature tracking, articulated body tracking, and voice analytics (Baltrusaitis, Robinson, & Morency, 2012; Morency, Whitehill, & Movellan, 2008; Whitehill, Littlewort, Fasel, Bartlett, & Movellan, 2009) has opened the door to new

**Figure 3.2** User with SimSensei virtual clinical interviewer.

applications for automatic nonverbal behavior analysis. This sensing, quantification, and inference from nonverbal behavioral cues can serve to provide input to an interactive VH interviewer that can respond with follow-up questions that leverage inferred indicators of user distress or anxiety during a short interview. This is the primary concept that underlies the "SimSensei" interviewing agent (see Figure 3.2).

SimSensei is one application component developed from the DARPA-funded "Detection and Computational Analysis of Psychological Signals (DCAPS)" project. This DCAPS application has aimed to explore the feasibility of creating "empathic" VH health agents for use as clinical interviewers and to aid in mental health screening. The system seeks to combine the advantages of traditional web-based self-administered screening (Weisband & Kiesler, 1996), which allows for anonymity, with anthropomorphic interfaces which may foster some of the beneficial social effects of face-to-face interactions (Kang & Gratch, 2012). When the SimSensei system is administered in a private kiosk-based setting, it is envisioned to conduct a clinical interview with a patient who may be initially hesitant or resistant to interacting with a live mental healthcare provider. In this regard, SimSensei evolves the earlier web-based screening tool, SimCoach, to engage users in a private structured interview using natural language. What SimSensei adds to the mix is real-time sensing of user behavior that aims to identify behaviors associated with anxiety, depression, or PTSD. Such behavioral signals are sensed (with cameras and a microphone), from which inferences are made to quantify user state across an interview; that information is also used in real time to update the style and content of the SimSensei follow-up questions.

The SimSensei capability to accomplish this is supported by the "MultiSense" perception system (Morency, 2010; http://multicomp.ict.usc.edu/?p=1799; Devault, Rizzo, & Morency, 2014), a multimodal system that allows for synchronized capture of different modalities such as audio and video and provides a flexible platform for real-time tracking and multimodal fusion. MultiSense enables fusion of modality "markers" to support the development of more complex multimodal indicators of user state. MultiSense fuses information from a web camera, Microsoft Kinect, and audio capture and processing hardware to identify the presence of predetermined nonverbal indicators of psychological distress. This allows for the dynamic capture and quantification of behavioral signals such as 3D head position and orientation, type, intensity, and frequency of facial expressions of emotion (e.g., fear, anger, disgust, and joy), fidgeting, slumped body posture, along with a variety of speech parameters (e.g., speaking fraction, speech dynamics, latency to respond). These informative behavioral signals serve two purposes. First, they produce the capability to analyze the occurrence and quantity of behaviors to inform detection of psychological state. Second, they are broadcast to other software components of the SimSensei Kiosk to inform the VH interviewer of the state and actions of the participant. This information is then used by the VH to assist with turn taking, rapport building (e.g., utterances, acknowledging gestures/facial expressions), and to drive and deliver follow-on questions. In-depth technical details of the MultiSense software as well as the SimSensei dialog management, natural language system, and agent face and body gesture generation methods are beyond the scope of this chapter and can be found elsewhere (Devault et al., 2014; Scherer et al., 2014).

### *Nonverbal Behavior and Clinical Conditions*

To begin to develop a corpus of automatic nonverbal behavior descriptors that MultiSense could track for the SimSensei application, we searched the large body of research that had examined the relationship between nonverbal behavior and clinical conditions. Most of this research resided in the clinical and social psychology literature and until very recently the vast majority relied on manual annotation of gestures and facial expressions. Despite at least 40 years of intensive research, there has still been surprisingly little progress on identifying clear relationships between clinical disorders and expressed behavior. In part, this is due to the difficulty in manually annotating data, inconsistencies in how both clinical states and expressed behaviors are defined across studies, and the wide range of social contexts in which behavior is elicited and observed. However, in spite of these complexities, there is general consensus on the relationship between some clinical conditions (especially depression and anxiety) and associated nonverbal cues. These findings informed our initial search for automatic nonverbal behavior descriptors.

For example, gaze and mutual attention are critical behaviors for regulating conversations, so it is not surprising that a number of clinical conditions are associated with atypical patterns of gaze. Depressed patients have a tendency to maintain significantly less mutual gaze (Waxer, 1974), show nonspecific gaze, such as staring off into space (Schelde, 1998), and avert their gaze, often together with a downward angling of the head (Perez & Riggio, 2003). The pattern for depression and PTSD is similar, with patients often avoiding direct eye contact with the clinician. Emotional expressivity, such as the frequency or duration of smiles, is also diagnostic of clinical state. For example, depressed patients frequently display flattened or negative affect including less emotional expressivity (Bylsam et al., 2008; Perez & Riggio, 2003), fewer mouth movements (Fairbanks, McGuire, & Harris, 1982; Schelde, 1998), more frowns (Fairbanks et al., 1982; Perez & Riggio, 2003), and fewer gestures (Hall, Harrigan, & Rosenthal, 1995; Perez & Riggio, 2003). Some findings suggest it is not the total quantity of expressions that is important, but their dynamics. For example, depressed patients may frequently smile, but these are often shorter in duration and perceived as less genuine (Kirsch & Brunnhuber, 2007) than what is found in nonclinical populations. Social anxiety and PTSD, while sharing some of the features of depression, also have a tendency for heightened emotional sensitivity and more energetic responses, including hypersensitivity to stimuli (e.g., more startle responses, and greater tendency to display anger) (Kirsch & Brunnhuber, 2007), or shame (Menke, 2011). Fidgeting is often reported with greater frequency in clinical populations. This includes gestures such as tapping or rhythmically shaking hands or feet and has been reported in both anxiety and depression (Fairbanks et al., 1982). Depressed patients also often engage in "self-adaptors" (Ekman & Friesen, 1969), such as rhythmically touching, hugging, or stroking parts of the body or self-grooming, such as repeatedly stroking the hair (Fairbanks et al., 1982). Examples of observed differences in verbal behavior in depressed individuals include increased speaker-switch durations and diminished variability in vocal fundamental frequency (Cohn et al., 2009), decreased speech output, slow speech, delays in delivery, and long silent pauses (Hall et al., 1995). Differences in certain lexical frequencies have been reported, including use of first-person pronouns and negatively valenced words (Rude, Gortner, & Pennebaker, 2004).

Thus, the key challenge when building such nonverbal perception technology is to develop and validate robust descriptors of human behaviors that are correlated with psychological distress. These descriptors should be designed to probabilistically inform diagnostic assessment or quantify treatment outcomes. However, no descriptor is completely diagnostic by itself, but rather may reveal "tendencies" in users' nonverbal behaviors that are informational to enhance clinical hypothesis testing and/or decision-making. As an initial step in this process, we examined a variety sources of information to identify such behaviors: (i) a review of the literature on nonverbal behaviors that are indicative of psychological conditions as reported by clinical observations; (ii) existing work on automatic analysis; (iii) qualitative analysis based on observations from the videos and consultation with experts (including trained clinicians) who looked at the videos and identified the communicative behaviors that they would use to inform a diagnosis. Once a road map of candidate behaviors was created, the next step required dual videotaping of face-to-face interviews with individuals who were likely to be in a state of psychological distress (i.e., a veteran sample that had high scores on self-report tests of anxiety, depression, and PTSD) versus those not deemed to be in distress from the same psychometric tests. This face-to-face corpus of videos was scored via manual annotation.

Following the analysis of face-to-face human interactions to identify potential emotional indicators, dialog policies, and commonality of human gestures, the development and analysis of a Wizard-of-Oz (WoZ) prototype system was required. The WoZ interaction allowed human operators to choose the spoken and gestural responses of a VH character (similar to digital puppetry) that interacted with a live research participant. The final step involved the development of a fully automatic SimSensei virtual interviewer that is able to engage users in 15–25 min interactions. A full detailing of this research methodology and results from this work can be found in Scherer et al. (2014). A brief discussion of comparisons between the face-to-face, WoZ-driven, and Automatic VH agent are presented in the next section.

## COMPARATIVE EVALUATION ACROSS INTERVIEWS: FACE-TO-FACE, WOZ, AND AUTOMATIC INTERACTION WITH THE SIMSENSEI VH AGENT

After a large number of users were psychometrically characterized and behaviorally evaluated in both live interviews and WoZ systems, the next step involved programing the sensing of candidate behavior marker for integration into an autonomous SimSensei interview. More specifically, the perception system's functionality was tuned to automatically track and recognize candidate nonverbal behaviors deemed as important for the assessment of psychological distress, as reported from the previous steps, but in the context of an interview with an Automatic Interaction (AI) VH agent (essentially, all software-driven, no human in the loop), now named as "Ellie." The key sensed behaviors associated with depression, anxiety, and PTSD were extracted live during the interview, were used to guide Ellie's interactive behavior, and the summary statistics were available automatically at the end of the interview. At this stage the focus was on the capture and analysis of such behavioral signals in the real-time system and the validation of the previous analysis of face-to-face and WoZ data on the new corpus of fully automated interactions.

Across all three interview formats, 351 participants were recruited through Craigslist, posted flyers, and from access to a sample of veterans receiving services from the US Vets program (http://www.usvetsinc.org/). Of the 120 face-to-face participants, 86 were male and 34 were female. These participants had a mean age of 45.56 (SD = 12.26). Of the 140 WoZ participants, 76 were male, 63 were female, and one did not report their gender. The mean age of this group of participants was 39.34 (SD = 12.52). Of the 91 AI participants, 55 were male, 35 were female, and one did not report their gender. They had a mean age of 43.07 (SD = 12.84).

All participants were given a series of self-report assessment instruments to index their clinical state (i.e., Patient Health Questionnaire, Posttraumatic Stress Disorder Checklist, and the State/Trait Anxiety Questionnaire). Postexperience, all participants completed a validated measure of rapport (Kang & Gratch, 2012). Additionally, participants in WoZ and AI completed nine questions designed to evaluate and compare user impressions of both VH formats to test our success in meeting specific VH design goals (see Table 3.1). Examples include questions about disclosure ("I was willing to share information with Ellie"), the

**Table 3.1** Means, standard errors, *t*-values, and effect sizes on user impression questions

| Design goals | Method | | *t*-value | *d* |
|---|---|---|---|---|
| | **WoZ** | **AI** | | |
| I was willing to share information with Ellie | 4.03 (0.83) | 4.07 (0.73) | −0.33 | 0.05 |
| I felt comfortable sharing information with Ellie | 3.92 (0.98) | 3.80 (1.07) | 0.75 | 0.12 |
| I shared a lot of personal information with Ellie | 3.97 (1.04) | 3.73 (1.14) | 1.47 | 0.23 |
| It felt good to talk about things with Ellie | 3.69 (1.02) | 3.60 (0.95) | 0.55 | 0.08 |
| There were important things I chose not to tell Ellie | 2.93 (1.19) | 2.66 (1.19) | 1.48 | 0.23 |
| Ellie was a good listener | 4.10 (0.77) | 3.56 (0.98) | 3.94* | 0.61 |
| Ellie has appropriate body language | 3.85 (0.85) | 3.84 (0.86) | 0.05 | 0.01 |
| Ellie was sensitive to my body language | 3.36 (0.72) | 3.13 (0.86) | 1.87 | 0.29 |
| I would recommend Ellie to a friend | 3.72 (1.10) | 3.47 (1.03) | 1.52 | 0.24 |
| System usability | 74.37 (13.63) | 68.68 (12.05) | 3.24* | 0.44 |
| Rapport | 80.71 (12.10) | 75.43 (11.71) | 3.28* | 0.44 |

* = <0.05

mechanics of the interaction ("Ellie was sensitive to my body language"), and willingness to recommend the system to others. All were rated on a scale from 1 (strongly disagree) to 5 (strongly agree). Note that in the WoZ condition, participants were told that the agent was autonomous and not puppeted by two people.

With regard to the design goals, most participants agreed or strongly agreed that they were achieved, whether they interacted with the Wizard-operated or AI system. For example, most people agreed or strongly agreed that they were willing to share information with Ellie (84.2% WoZ; 87.9% AI), were comfortable sharing (80.5% WoZ; 75.8% AI), and did share intimate information (79.3% WoZ; 68.2% AI). Both systems performed less well with regard to their perceived ability to sense and generate appropriate nonverbal behavior. For example, a minority of participants agreed or strongly agreed that Ellie could sense their nonverbal behavior (40.3% WoZ; 27.5% AI). However, this did not seem to

**Table 3.2** Rapport scores in the three conditions

| Face-to-face | WoZ | AI |
|---|---|---|
| 74.42 (4.89) | 80.71 (12.10) | 75.43 (11.71) |

seriously detract from the overall experience and the majority agreed or strongly agreed they would recommend the system to a friend (69.8% WoZ; 56.1% AI).

We next examined the relative impressions of the AI system when compared with the WoZ. Although this initial AI version was in no way expected to reach human-level performance, this comparison gives an insight into areas that need improvement. Surprisingly, design and usability rating results yielded only three significant differences in favor of the human-driven system (WoZ). WoZ participants reported feeling that the interviewer was a better listener than the AI participants ($t(166) = 3.94$, $P < 0.001$, $d = 0.61$), rated the system as higher in usability than AI participants ($t(229) = 3.24$, $P = 0.001$, $d = 0.44$), and felt more rapport ($t(229) = 3.28$, $P = 0.001$, $d = 0.44$).

We then examined how the WoZ and AI systems compared with the original face-to-face interviews (see Table 3.2). We conducted an ANOVA to compare ratings of rapport for the three methods. Results revealed a significant effect of method on rapport ($F(2, 345) = 14.16$, $P < 0.001$, $d = 0.52$). Interestingly, this effect was driven by the WoZ. WoZ participants felt greater rapport than AI participants ($t(345) = 3.87$, $P < 0.001$, $d = 0.42$) and compared to face-to-face participants ($t(345) = -4.95$, $P < 0.001$, $d = 0.53$). Surprisingly, AI and face-to-face participants' ratings of rapport did not differ ($t(345) = -0.77$, $P = 0.44$, $d = 0.07$).

The results of this first evaluation are promising. In terms of subjective experience, participants reported willingness to disclose, willingness to recommend and general satisfaction with both the WoZ and AI versions of the system. In terms of rapport, participants reported feelings comparable to a face-to-face interview. Unexpectedly, participants felt more rapport when interacting with the WoZ system than they did in face-to-face interviews. One possible explanation for this effect is that people are more comfortable revealing sensitive information to computers than face-to-face interviewers (Lucas, Gratch, King, & Morency, 2014; Weisband & Kiesler, 1996), though this will require further study. As expected, the initial version of SimSensei did not perform as well as human "wizards." This is reflected in significantly lower ratings of rapport and system usability. Participants also felt that the AI-controlled Ellie was less sensitive to their own body language and often produced inappropriate nonverbal behaviors. Such analyses are a central focus of current work to inform the design of the next iteration of the SimSensei system and the overall results are promising and suggest that the first pass on this system elicited positive use-intentions.

Initial results from the analysis of face, gesture, and voice in the AI condition indicated clear differences across known groups (high vs low distress) in the presence of behavioral signals that were hypothesized to represent psychological distress (Scherer et al., 2014). Automatic detection of behaviors found to reflect distress included smile dynamics, averted head and gaze, fidgeting, and self-adaptors, and at least three vocal parameters which showed strong concordance with the self-report psychometric evaluations of psychological distress (i.e., anxiety, depression, and PTSD) and significant differences across groups in the expected direction were found. Note that the detailed presentation of these data is beyond the space limits of this chapter and interested readers are referred to Scherer et al. (2014).

The SimSensei system is now in the process of being iteratively refined in validation trials within two clinical projects. In one ongoing project, US military service members prior to a combat deployment in Afghanistan, were given a full battery of psychological tests and interviewed by the automatic SimSensei (AI) interviewer. This unit returned from deployment in December 2014 and participated in a post-deployment round of SimSensei testing with follow-up psychological evaluations planned at 6 months and at 1 year post-deployment. At the time of this writing, the post-deployment data are being analyzed for comparison with the pre-deployment findings with psychometric measures of mental health. The aim of this project is to determine whether automatic behavior markers captured during a SimSensei interview at pre- and post-deployment can predict mental health status change across the deployment compared to the self-report of SMs on our measures and on the Post Deployment Health Assessment (PDHA). In line with this, one study compared the standard administration of the PDHA (computerized and directly documented in the SMs military health record) with an anonymous PDHA administration and the findings indicated that depression, PTS, suicidal ideation, and interest in receiving care were twofold to fourfold higher on the anonymous survey compared with the standard administration of the PDHA (Warner et al., 2011). Thus, it becomes important to determine whether we can get a more "honest" characterization of psychological distress from a SimSensei interview that will support better-informed recommendations for care to SMs who are struggling with mental health conditions, but who "fake-good" on standard self-report assessments to avoid stigma or negative career implications. Further validation testing of a SimSensei-conducted clinical interview is also being used as part of the assessment package within an ongoing clinical trial testing VR exposure therapy for the treatment of PTSD due to military sexual trauma. The SimSensei interview is being conducted at pre-, mid-, and post-treatment in order to compare results with a sample whose mental health status is expected to improve over the course of treatment. Results from these types of known group tests are expected to continue to inform the iterative evolution of the SimSensei VH agent. A video of a user interacting with the AI version of SimSensei is available at: http://youtu.be/Yw1c5h_p6Dc.

## CONCLUSIONS

The systematic integration of autonomous VH agents into clinical applications is still clearly in its infancy. However, the days of limited use of VHs as simple props or static elements to add realism or context to a clinical VR application are clearly in the past. Work over the last 15 years in particular has produced a wealth of scientific and practical knowledge to advance both VH technology as well as the actual application of VHs. As advances in computing power, graphics and animation, artificial intelligence, speech recognition, and natural language processing continue at current rates, we expect that the creation and use of highly interactive, intelligent VHs for such clinical purposes will grow exponentially. This chapter presents the rationale and underlying structure of two clinically oriented autonomous VH applications: SimCoach and SimSensei. Each system was created with the idea of adding value to a process by providing users access to interaction with an autonomous VH agent. Neither of these applications are intended to replace an actual person in the conduct of clinical activities, but rather to fill a gap where a real person is unlikely to be available for the intended purpose. With SimCoach, the primary aim is to provide an activity that might encourage someone who is hesitant to seek care with a live provider, to seek or at least explore treatment options with a real clinician. SimSensei takes the SimCoach concept of providing a platform for a private interaction with a VH and attempts to add value by providing a novel format for clinical interviewing that supports a concurrent automated behavioral analysis designed to detect signals of psychological distress in the user. A similar case has been made across the general field of VR simulation, an advantage for a SimSensei Kiosk framework over a human interviewer is in the implicit replicability and consistency of the spoken questions and accompanying gestures. This standardization of the stimuli allows for a more detailed analysis of user responses to precisely delivered interview questions. Another advantage is that recent results suggest that VHs can reduce stress and fear associated with the perception of being judged and thereby lower emotional barriers to disclosing information (Hart, Gratch, & Marsella, 2013, Ch. 21; Lucas et al., 2014). Advancing these concepts will require iterative, research-driven VH development and future success may no longer be rate-limited by the pace of technology, but instead by the creativity and innovation of scientists and practitioners who can envision their useful application.

## REFERENCES


American Psychological Association. (2007). Presidential task force on military deployment services for youth, families and service members. The psychological needs of U.S. military service members and their families: A preliminary report. Retrieved from: <http://www.apa.org/releases/MilitaryDeploymentTaskForceReport.pdf> Accessed 18.04.07.

Anderson, P. L., Zimand, E., Hodges, L. F., & Rothbaum, B. O. (2005). Cognitive behavioral therapy for public-speaking anxiety using virtual reality for exposure. *Depression and Anxiety*, *22*(3), 156–158.

Babu, S., Schmugge, S., Barnes, T., & Hodges, L. (2006). What would you like to talk about? An evaluation of social conversations with a virtual receptionist. In J. Gratch, et al. (Eds.), *IVA 2006, LNAI 4133* (pp. 169–180). Berlin, Germany: Springer-Verlag.

Bailenson, J. N., & Beall, A. C. (2006). Transformed social interaction: Exploring the digital plasticity of avatars. In R. Schroeder, & A.-S. Axelsson (Eds.), *Avatars at work and play: Collaboration and interaction in shared virtual environments* (pp. 1–16). Berlin, Germany: Springer-Verlag.

Baltrusaitis, T., Robinson, P., & Morency, L.-P., (2012). 3D constrained local model for rigid and non-rigid facial tracking. In *Proceedings of the IEEE computer vision and pattern recognition*. Providence, RI.

Blascovich, J., Loomis, J., Beall, A., Swinth, K., Hoyt, C., & Bailenson, J. (2002). Immersive virtual environment technology: Not just another research tool for social psychology. *Psychological Inquiry*, *13*, 103–124.

Bickmore, T., & Cassell, J. (2005). Social dialogue with embodied conversational agents. In J. van Kuppevelt, L. Dybkjaer, & N. Bernsen (Eds.), *Advances in natural, multimodal dialogue systems*. New York, NY: Kluwer Academic.

Bordnick, P. S., Traylor, A. C., Carter, B. L., & Graap, K. M. (2012). A feasibility study of virtual reality-based coping skills training for nicotine dependence. *Research on Social Work Practice*, *22*(3), 293–300.

Bylsam, L. M., Morris, B. H., & Rottenberg, J. (2008). A meta-analysis of emotional reactivity in major depressive disorder. *Clinical Psychology Review*, *28*, 676–691.

Cohn, J. F., Kruez, T. S., Matthews, I., Yang, Y., Nguyen, M. H., Padilla, M. T., et al. (2009). Detecting depression from facial actions and vocal prosody. In *Affective computing and intelligent interaction (ACII)*, September, 2009.

Darwin, C. (2002). *The expression of the emotions in man and animals* (3rd ed.) London: Oxford University Press.

Devault, D., Rizzo, A. A., & Morency, L.-P. (2014). SimSensei: A virtual human interviewer for healthcare decision support. In *The proceedings of the thirteenth international conference on autonomous agents and multiagent systems* (AAMAS).

DOD Mental Health Task Force Report. (2007). Available from: <http://www.health.mil/dhb/mhtf/MHTF-Report-Final.pdf> Accessed 15.06.07.

Dole-Shalala Commission. (2007). *Serve, support, simplify: Report of the president's commission on care for America's returning wounded warriors*. Available from: <http://www.nyshealthfoundation.org/content/document/detail/1782/> Accessed 21.11.11.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica*, *1*, 49–98.

Ekman, P., & Rosenberg, E. L. (1997). *What the face reveals: Basic and applied studies of spontaneous expressions using the Facial Action Coding System (FACS)*. New York, NY: Oxford University Press.

Fairbanks, L. A., McGuire, M. T., & Harris, C. J. (1982). Nonverbal interaction of patients and therapists during psychiatric interviews. *Journal of Abnormal Psychology*, *91* (2), 109–119.

Gratch, J., & Marsella, S. (2004). A domain independent framework for modeling emotion. *Journal of Cognitive Systems Research*, *5*(4), 269–306.

Gratch, J., Morency, L. P., Scherer, S., Stratou, G., Boberg, J., Koenig, S., et al. (2013). User-state sensing for virtual health agents and telehealth applications. In *Proceedings of the 20th annual medicine meets virtual reality conference*.

Gratch, J., Rickel, J., Andre, E., Cassell, J., Petajan, E., Badler, N. (2002). Creating interactive virtual humans: Some assembly required. In *IEEE intelligent systems*. July/August. pp. 54–61.

Hall, J. A., Harrigan, J. A., & Rosenthal, R. (1995). Nonverbal behavior in clinician-patient interaction. *Applied and Preventive Psychology*, *4*(1), 21–37.

Hart, J., Gratch, J., & Marsella, S. (2013). How virtual reality training can win friends and influence people. *Fundamental Issues in Defense Training and Simulation (Human Factors in Defence)*, pp. 235–249. Surrey: Ashgate.

Hoge, C. W., Castro, C. A., Messer, S. C., McGurk, D., Cotting, D. I., & Koffman, R. L. (2004). Combat duty in Iraq and Afghanistan, mental health problems, and barriers to care. *New England Journal of Medicine*, *351*(1), 13–22.

Howell, S. R., & Muller, R. (2000). *Computers in psychotherapy: A new prescription*. Available from: <http://www.psychology.mcmaster.ca/beckerlab/showell/ComputerTherapy.PDF> Accessed 23.04.11.

IOM (Institute of Medicine). (2007). *Treatment of posttraumatic stress disorder: An assessment of the evidence*. Washington, DC: The National Academies Press. Available from: <http://www.nap.edu/catalog/11955.html> Accessed 24.10.07.

IOM (Institute of Medicine). (2012). *Treatment for posttraumatic stress disorder in military and veteran populations: Initial assessment*. Washington, DC: The National Academies Press. Available from: <http://www.iom.edu/Reports/2012/Treatment-for-Posttraumatic-Stress-Disorder-in-Military-and-Veteran-Populations-Initial-Assessment.aspx> Accessed 15.07.12.

Kang, S.-H., & Gratch, J. (2012). Socially anxious people reveal more personal information with virtual counselors that talk about themselves using intimate human back stories. In B. Weiderhold, & G. Riva (Eds.), *The annual review of cybertherapy and telemedicine* (pp. 202–207). Amsterdam, The Netherlands: IOS Press.

Kenny, P., Rizzo, A. A., Parsons, T., Gratch, J., & Swartout, W. (2007). A virtual human agent for training clinical interviewing skills to novice therapists. *Annual Review of Cybertherapy and Telemedicine, 2007*, *5*, 81–89.

Kirsch, A., & Brunnhuber, S. (2007). Facial expression and experience of emotions in psychodynamic interviews with patients with PTSD in comparison to healthy subjects. *Psychopathology, 40*(5), 296–302.

Klinger, E. (2005). Virtual reality therapy for social phobia: Its efficacy through a control study. Paper presented at: *Cybertherapy* 2005. Basal, Switzerland.

Lucas, G. M., Gratch, J., King, A., & Morency, L.-P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, *37*, 94–100.

Macedonio, M. F., Parsons, T., Wiederhold, B., & Rizzo, A. A. (2007). Immersiveness and physiological arousal within panoramic video-based virtual reality. *CyberPsychology and Behavior*, *10*(4), 508–515.

McCall, C., Blascovich, J., Young, A., & Persky, S. (2009). Proxemic behaviors as predictors of aggression towards Black (but not White) males in an immersive virtual environment. *Social Influence*, 1–17.

McCauley, L., & D'Mello, S. (2006). A speech enabled intelligent kiosk. In J. Gratch, et al. (Eds.), *IVA 2006, LNAI 4133* (pp. 132–144). Berlin, Germany: Springer-Verlag.

Menke, R. (2011). *Examining nonverbal shame markers among post-pregnancy women with maltreatment histories* (Ph.D. thesis). Wayne State University.

Morency, L.-P. (2010). Modeling human communication dynamics. *IEEE Signal Processing Magazine*, *27*(6), 112–116.

Morency, L.-P., de Kok, I., & Gratch, J. (2008). Context-based recognition during human interactions: Automatic feature selection and encoding dictionary. In *10th International conference on multimodal interfaces*, Chania, Greece, IEEE.

Morency, L.-P., Whitehill, J., & Movellan, J. (2008). Generalized adaptive view-based appearance model: Integrated framework for monocular head pose estimation. In *Proceedings of the 8th IEEE international conference on automatic face gesture recognition* (FG08). pp. 1–8.

Padgett, L., Strickland, D., & Coles, C. (2006). Case study: Using a virtual reality computer game to teach fire safety skills to children diagnosed with Fetal Alcohol Syndrome (FAS). *Journal of Pediatric Psychology, 31*(1), 65–70.

Parsons, S., Bauminger, N., Cobb, S., Gal, E., Glover, T., Weiss, P.L., et al. (2012). Collaborative technologies for supporting social conversation and collaboration in the classroom: Overview and findings of the COSPATIAL project. In *1st international conference on innovative technologies for autism spectrum disorders. ASD: Tools, trends and testimonials*.Valencia, Spain, July 6–8, 2012.

Parsons, T., Bowerly, T., Buckwalter, J. G., & Rizzo, A. A. (2007). A controlled clinical comparison of attention performance in children with ADHD in a virtual reality classroom compared to standard neuropsychological methods. *Child Neuropsychology, 13*, 363–381.

Pentland, A. (2008). *Honest signals: How they shape our world*. Cambridge, MA: MIT Press.

Pentland, A., Lazer, D., Brewer, D., & Heibeck, T. (2009). Using reality mining to improve public health and medicine. *Studies in Health Technology and Informatics*, *149*, 93–102.

Perez, J. E., & Riggio, R. E. (2003). Nonverbal social skills and psychopathology. In P. Philpot, R. S. Feldman, & E. J. Coats (Eds.), *Nonverbal behavior in clinical settings* (pp. 17–44). London: Oxford University Press.

Pertaub, D.-P., Slater, M., & Barker, C. (2002). An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence*, *11*(1), 68–78.

Prendinger, H., & Ishizuka, M. (2004). *Life-like characters – Tools, affective functions, and applications*. New York, NY: Springer.

Rickel, J., Gratch, J., Hill, R., Marsella, S., & Swartout, W. (2001). Steve goes to Bosnia: Towards a new generation of virtual humans for interactive experiences. In *The proceedings of the AAAI spring symposium on AI and interactive entertainment*, Stanford University, CA.

Rizzo, A. A., Bowerly, T., Buckwater, J. G., Klimchuk, D., Mitura, R., & Parsons, R. D. (2006). A virtual reality scenario for all seasons: The virtual classroom. *CNS Spectums*, *11*(1), 35–44.

Rizzo, A. A., Ghahremani, K., Pryor, L., & Gardner, S. (2003). Immersive 360-degree panoramic video environments. In J. Jacko, & C. Stephanidis (Eds.), *Human-computer interaction: Theory and practice* (Vol. 1, pp. 1233–1237). New York, NY: L.A. Erlbaum.

Rizzo, A. A., Sagae, K., Forbell, E., Kim, E., Lange, B., Buckwalter, J. G., et al. (2011). SimCoach: An intelligent virtual human system for providing healthcare information and support. In *Proceedings of the interservice/industry training, simulation and education conference*.

Rizzo, A. A., Scherer, S., DeVault, D., Gratch, J., Artstein, R., Hartholt, A., et al. (2014). Detection and computational analysis of psychological signals using a virtual human interviewing agent. In *The 2014 proceedings of the international conference on disability, virtual reality and associated technologies*.

Rude, S., Gortner, E.-M., & Pennebaker, J. (2004). Language use of depressed and depression-vulnerable college students. *Cognition & Emotion*, *18*(8), 1121–1133.

Russell, J. A., & Fernandez-Dols, J. M. (1997). What does a facial expression mean? In J. A. Russell, & J. M. Fernandez-Dols (Eds.), *The psychology of facial expression* (pp. 3–30). New York, NY: Cambridge University Press.

Rutten, A., Cobb, S., Neale, H., Kerr, S., Leonard, A., Parsons, S., et al. (2003). The AS interactive project: Single-user and collaborative virtual environments for people with high-functioning autistic spectrum disorders. *Journal of Visualization and Computer Animation*, *14*(5), 233–241.

Schelde, J. T. M. (1998). Major depression: Behavioral markers of depression and recovery. *The Journal of Nervous and Mental Disease*, *186*(3), 33–140.

Scherer, S., Stratou, G., Lucas, G., Mahmoud, M., Boberg, J., Gratch, J., et al. (2014). Automatic audiovisual behavior descriptors for psychological disorder analysis. *Image and Vision Computing*, *32*(10), 648–658.

Simmersion, Inc. Available from: <www.simmersion.com> Accessed 15.04.15.

Swartout, B., Artstein, R., Forbell, E., Foutz, S., Lane, H. C., Lange, B., et al. (2013). Virtual humans for learning. *Artificial Intelligence Magazine*, *34*(4), 13–30.

Talbot, T. B., Sagae, K., John, B., & Rizzo, A. A. (2012). Sorting out the virtual patient: How to exploit artificial intelligence, game technology and sound educational practices to create engaging role-playing simulations. *International Journal of Gaming and Computer-Mediated Simulations*, *4*(3), 1–19.

Tanielian, T., Jaycox, L. H., Schell, T. L., Marshall, G. N., Burnam, M. A., Eibner, C., et al. (2008). *Invisible wounds of war: Summary and recommendations for addressing psychological and cognitive injuries*. Rand Report. Retrieved from: <http://veterans.rand.org/>.

Thiebaux, M., Marshall, A., Marsella, S., Fast, E., Hill, A., Kallmann, M., et al. (2008). SmartBody: Behavior realization for embodied conversational agents. In *Proceedings of the international conference on autonomous agents and multi-agent systems (AAMAS)*. Portugal.

Traum, D., Marsella, S., Gratch, J., Lee, J., & Hartholt, A. (2008). Multi-party, multi-issue, multi-strategy negotiation for multi-modal virtual agents. In *8th International Conference on Intelligent Virtual Agents*. Tokyo, Japan, Springer.

Virtually Better Inc. Available from: <www.virtuallybetter.com> Accessed 15.04.15.

Warner, C. H., Appenzeller, G. N., Grieger, T., Belenkiy, S., Breitbach, J., Parker, J., et al. (2011). Importance of anonymity to encourage honest reporting in mental health screening after combat deployment. *Archives of General Psychiatry*, *68*(10), 1065–1071.

Waxer, P. (1974). Nonverbal cues for depression. *Journal of Abnormal Psychology*, *83*(3), 319–322.

Weisband, S., & Kiesler, S. (1996). Self-disclosure on computer forms: Meta-analysis and implications. In *Proceedings of CHI1996*, Vol. 96, pp. 3–10.

Weizenbaum, J. (1976). *Computer power and human reason* San Francisco, CA: W. H. Freeman.

Whitehill, J., Littlewort, G., Fasel, I., Bartlett, M., & Movellan, J. (2009). Toward practical smile detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*, 2106–2111.