Under Review: *Studies in Healthcare Technology*.

Automatic Behavior Analysis During a Clinical Interview with a Virtual Human

Albert RIZZO[a, 1], Gale LUCAS[a], Jonathan GRATCH[a], Giota STRATOU[a], Louis-Philippe MORENCY[b], Kenneth CHAVEZ[c], Russ SHILLING[d], Stefan SCHERER[a],
[a]*University of Southern California, Institute for Creative Technologies*
[b]*Carnegie Mellon University,* [c]*Colorado National Guard,* [d]*U.S. Dept. of Education*

Abstract. SimSensei is a Virtual Human (VH) interviewing platform that uses off-the-shelf sensors (i.e., webcams, Microsoft Kinect and a microphone) to capture and interpret real-time audiovisual behavioral signals from users interacting with the VH system. The system was specifically designed for clinical interviewing and health care 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. Akin to how non-verbal behavioral signals have an impact on human-to-human interaction and communication, SimSensei aims to capture and infer user state from signals generated from user non-verbal communication to improve engagement between a VH and a user and to quantify user state from the data captured across a 20 minute interview. Results from of sample of service members (SMs) who were interviewed before and after a deployment to Afghanistan indicate that SMs reveal more PTSD symptoms to the VH than they report on the Post Deployment Health Assessment. Pre/Post deployment facial expression analysis indicated more sad expressions and few happy expressions at post deployment.

Keywords. Virtual Humans, Posttraumatic Stress Disorder, virtual reality, Automatic Behavior Analysis, SimSensei, Multisense

1. Introduction

Over the last 20 years, a virtual revolution has taken place in the use of Virtual Reality simulation technology for clinical purposes. Recent shifts in the social and scientific landscape have now set the stage for the next major movement in Clinical Virtual Reality with the "birth" of intelligent virtual human (VH) agents. This has been driven by seminal research and development leading to the creation of highly interactive, artificially intelligent and natural language capable VHs that can engage real human users in a credible fashion. Virtual humans can now be designed to perceive and act in a virtual world, engage in face-to-face spoken dialogues, and exhibit believable human-like emotional reactions during interactions with real humans. Recent advances in the technology needed to create VH systems is now driving application development across a number of fields, from education to clinical training to providing clinical assessment and healthcare support. For example, VH's can conduct clinically-oriented interviews

[1] Albert Rizzo. Institute for Creative Technologies, 12015 Waterfront Dr. Playa Vista, CA. 90094, rizzo@ict.usc.edu. This work is supported by DARPA, the U.S. Army Research, Development, and Engineering Command and the Oakley Infinite Hero Foundation. The content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

within a safe non-judgmental context which may encourage learning or honest disclosure of important information (1). The healthcare field, in particular, may benefit from this latter potential advantage of VH, as failure to provide honest responses in medical interviews can result in serious negative consequences for patient health. This paper will detail our efforts in the creation of a VH who can serve in the role of a clinical interviewer (i.e., *SimSensei*) while using camera and audio sensors to automatically detect behavioral signals from which user state may be inferred.

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 health care 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 non-verbal behavioral signals have an impact on human-to-human interaction and communication, SimSensei aims to capture and infer user state from signals generated from user non-verbal communication to improve engagement between a VH and a user. The system also can quantify and interpret sensed behavioral signals longitudinally for use 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 expression and body gestures play an important role in human communicative signaling (2). As well, vocal characteristics (e.g., prosody, pitch variation, etc.) have been reported to provide additive information regarding the "state" of the speaker beyond the actual language content of the speech (3). Pentland has characterized these elements of behavioral expression as "Honest Signals" (4). 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 non-verbal communication mechanisms, provide a useful window into our intentions, goals, values and emotional state. From 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 using that 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.

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 (5-7) has opened the door to new applications for automatic nonverbal behavior analysis. This sensing, quantification and inference from nonverbal behavioral cues can serve to provide input to an interactive virtual human 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 1). The SimSensei capability to accomplish this is supported by the "MultiSense" perception system (8-10), a multimodal system that allows for real-time synchronized capture, tracking, and fusion of behavioral markers of different modalities such as audio as well as video. MultiSense's fusion enables the analysis of complex behavioral indicators of user states across multiple modalities. Within SimSensei, MultiSense fuses information from a web camera, Microsoft Kinect and audio capture to identify the presence of predetermined nonverbal indicators of psychological distress. Dynamic capture and quantification of behavioral signals are used 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, latency to respond). These informative behavioral signals serve two purposes. First, they produce the capability of analyzing 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.

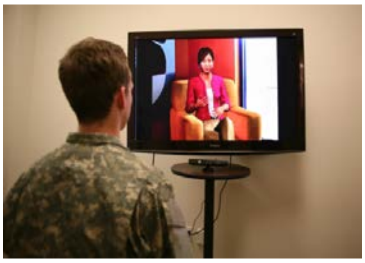

Figure 1. User with SimSensei virtual clinical interviewer

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” virtual human 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 (10), which allows for anonymity, with anthropomorphic interfaces which may foster some of the beneficial social effects of face-to-face interactions (11). 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 health care provider. SimSensei’s real time sensing of user behavior aims to identify behaviors associated with anxiety, depression or PTSD. Such behavioral signals are sensed 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. Technical details of the Multisense software as well as the SimSensei dialog management, natural language system, and agent face/body gesture generation methods are beyond the scope of this paper and can be found elsewhere (9-10). Instead, we focus on the usefulness of SimSensei in collecting *honest* health information during a clinical interview with active duty Service Members (SMs) prior to and immediately following a 9-month deployment to Afghanistan.

2. Methods and Procedure

Subjects

Twenty nine (2 female) active duty members of the Colorado National Guard volunteered for this study prior to embarking on a 9-month deployment to Afghanistan. They were a diverse sample regarding age (Mean=41.46, Range=26 to 56) and previous deployments (Number of combat deployments Mean=2.00, Range=1 to 7).

Assessment Instruments

The study compared the endorsement of Posttraumatic Stress (PTS) symptoms in three formats: 1) standard administration of the Post-Deployment Health Assessment (PDHA) upon return from deployment; 2) an anonymized version of the PDHA; 3) parallel SimSensei interview questions. Participants signed releases to access their actual PDHA that they submitted to the National Guard upon return from this deployment.

On the PDHA, participants are asked “Have you ever had any experience that was so frightening, horrible, or upsetting that, in the past month, you: A) have had nightmares about it or thought about it when you did not want to?, B) tried hard not to think about it or went out of your way to avoid situations that remind you of it?, and C) were constantly on guard, watchful, or easily startled?” These questions assess whether the SM is experiencing the core DSM-4TR diagnostic symptoms for PTSD (intrusive recollections; avoidance/numbing; hyperarousal). Participants selected “yes” or “no” on each item of the PDHA and our anonymized version.

The questions SimSensei asked on these topics were worded slightly differently to embed them in the interview without having the VH simply recite the PDHA. Participants were asked: “Can you tell me about an experience you had in the past few months that challenged you on an emotional level?" (trauma event criterion), followed by “Can you tell me about any bad dreams you’ve had about your experiences, or times when thoughts or memories just keep going through your head when you wish they wouldn't?” (intrusive recollection criterion), “Tell me about any times you found yourself actively trying to avoid thoughts or situations that remind you of past events,” (avoidance/numbing criterion) and “Can you tell me about any times recently when you felt jumpy or easily startled?” (hyperarousal). As described, MultiSense quantifies affect levels (positive, worry/fear), ranging from 0 to 100, where 0 is the absence of the emotion and 100 a strong emotion. Self-reports of these emotions were also elicited: “I am happy” and “I worry too much” were rated on 4 point scales from *never* to *always*.

Procedure

Participants completed measures both before and after deploying (PDHA by definition was only available post-deployment). After giving consent, participants completed demographic questions as well as a number of measures described elsewhere (DeVault et al., 2014). The confidentiality of all these measures was stressed. Participants then engaged in an interview with the SimSensei VH who conducts a semi-structured screening interview with a user via spoken language. The interview is structured around a series of agent-initiated questions organized into phases: initially there is a rapport-building phase where the agent asks general introductory questions (e.g., “Where are you from originally?); this is followed by a clinical phase where the agent asks a series of questions about symptoms (e.g., “How easy is it for you to get a good night’s sleep?”), which include the naturally embedded PDHA questions; finally, the agent ends with questions designed to return the patient to a more positive mood (e.g., “What are you most proud of?”). At each phase, the agent can ask follow-up questions (e.g., “Can you tell me more about that?”), provide empathetic feedback (e.g., “I’m sorry to hear that”), and produce nonverbal behaviors (e.g., nods, expressions) for active listening. Participants' answers to the three PDHA questions during the interview were coded by two blind coders as to whether the participant had this experience in the last month. These coders had 100% agreement, and codes served as "yes" or "no" answers.

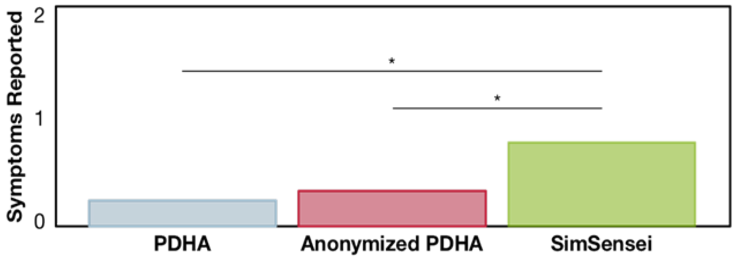

Figure 2. Results of comparison between assessment types

3. Results

To test whether responses to the three versions of the PDHA (official PDHA, Anonymized PDHA, and SimSensei) differed, scores were created for each version by counting the number of "yes" answers to the three questions, which could range from 0 to 3. To compare these scores, we conducted a repeated-measures ANOVA using the 24 participants who successfully completed all three measures. There was a significant effect of assessment type, F(2, 23) = 4.29, p = .02 (see Figure 2). Follow-up contrasts revealed that participants reported more symptoms of PTSD (responded "yes" on more questions) when asked by SimSensei (M = 0.79, SE = 0.23) than when reporting on the official PDHA (M = 0.25, SE = 0.15)), F(1, 23) = 7.38, p = .01, or even when reporting on our anonymized version of the PDHA (M = 0.33, SE = 0.16), F(1, 23) = 4.84, p = .04. Furthermore, unlike in other studies where anonymity increased reporting of symptoms (14), our analysis of this small sample did not reveal differences between official and anonymized versions of the PDHA, F (1, 23) = 0.19, p = .66.

Moreover, MultiSense facial expression analysis identified pre-to-post changes in positive affect (M = 0.32, SE = 0.05 to M = 0.15, SE = 0.02, F(1, 22) = 16.33, p = .001) and worry/fear (M = 0.01, SE = 0.002 to M = 0.04, SE = 0.01, F(1, 22) = 8.41, p = .008), whereas self-reports did not show less happiness (M = 3.32, SE = 0.13 to M = 3.28, SE = 0.15) or greater worry/fear (M = 1.56, SE = 0.12 to M = 1.60, SE = 0.14) pre- to post-deployment (Fs < 0.11 , ps > .74).

4. Conclusions and Future Work

The present study suggests that SMs following a deployment to Afghanistan were more likely to report symptoms of PTSD when interviewed by a VH than on both the standard and the anonymized version of the PDHA. This result is in line with our previous work that indicated that users felt less concerned about being evaluated and displayed more sadness in an interview with a VH agent compared to one where they believed a VH avatar was being operated by a human-in-the-loop "Wizard of Oz" controller (1). These results are part of a growing body of research that is suggesting that VH interviews may reduce stigma by providing a safe context where users may reveal more honest assessment information in contrast to situations where users are concerned about negative judgments on the part of a human interviewer. Additionally, the automatic behavior detection seems to provide a more informed window into the emotional state than self-report. We are currently running a replication of this study with a larger sample of US Veterans that aims to investigate whether these results can 1) be replicated, 2) be found within another SM population, and 3) are not the product of confounds due to slight wording differences in the assessment questions between the questionnaire and the VH. Specifically, in this study with veterans, the wording is equivalent across conditions. The data collection is ongoing, but initial results suggest replication and final results from that study will be presented at the conference. Finally, we are using the SimSensei interviewer as a PTSD assessment method within a clinical trial evaluating the use of virtual reality exposure therapy for PTSD due to Military Sexual Trauma. SimSensei interviews are being conducted at Pre-treatment, Mid-treatment, and at Post-Treatment. Data acquired from the capture and analysis of non-verbal behavior emitted by the patients during the VH interview process is being compared/correlated with: 1) the Clinician Administered PTSD Scale (CAPS) structured interview, 2) self-report screening measures of PTSD (PCL-M5) and other clinical measures (PHQ-9—Depression, etc.), and 3) a learning theory-based psychophysiological "startle response" conditioning/extinction protocol. This will allow for a better understanding of the value of VH interview assessment in a situation where stigma due to the nature of sexual trauma may be high and thus, interview questions delivered by a VH may provide a safer context for honest reporting to support the aims of treatment.

References

1. Lucas, G.M., Gratch, J., King, A., and Morency, L.-P., (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, 37, 94–100.
2. Ekman, P., & Rosenberg, E.L. (1997). *What the face reveals: Basic and applied studies of spontaneous expressions using the Facial Action Coding System (FACS)*, Oxford University Press, New York.
3. Pentland, A., Lazer, D., Brewer, D., and Heibeck, T. (2009). Using reality mining to improve public health and medicine. *Studies in Health Technology and Informatics*, 149, 93-102.
4. Pentland, A. (2008). *Honest signals: How they shape our world.* MIT Press, Cambridge, MA.
5. Baltrusaitis, T., Robinson, P., and Morency, L.-P., (2012). 3D constrained local model for rigid and non-rigid facial tracking, *Proc. of The IEEE Computer Vision and Pattern Recognition*. Providence, RI.
6. Morency, L.-P., de Kok, I. Gratch, J. (2008). Context-based Recognition during Human Interactions: Automatic Feature Selection and Encoding Dictionary. *10th International Conference on Multimodal Interfaces*, Chania, Greece, IEEE.
7. Whitehill, J., Littlewort, G., Fasel, I., Bartlett, M., and Movellan, J., (2009). Toward practical smile detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 31, 2106–2111.
8. Morency, L.-P. (2010). Modeling Human Communication Dynamics. IEEE Signal Processing Magazine. 27(6): 112-116.
9. Devault, D., Rizzo, A.A., and 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
10. Scherer, S., Stratou, G., Lucas, G., Mahmoud, M., Boberg, J., Gratch, J., Rizzo, A.A. & Morency, L. P. (2014). Automatic audiovisual behavior descriptors for psychological disorder analysis. *Image and Vision Computing*, 32(10), 648-658.
11. Weisband, S., and Kiesler, S., (1996). Self-disclosure on computer forms: Meta-analysis and implications. In *Proceedings of CHI1996*, 96, 3–10.
12. Kang, S.-H., and 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 and G Riva (Eds.), The Annual Review of Cybertherapy and Telemedicine. IOS Press, Amsterdam, The Netherlands. 202–207.
14. Warner, C.H., Appenzeller, G.N., Grieger, T., Belenkiy, S., Breitbach, J., Parker, J., Warner, C.M. & Hoge, C. (2011). Importance of Anonymity to Encourage Honest Reporting in Mental Health Screening After Combat Deployment. *Arch Gen Psychiatry*. 68(10), 1065-1071.