User-State Sensing for Virtual Health Agents and TeleHealth Applications

Jonathan Gratch [a,1], Louis-Philippe Morency[a], Stefan Scherer[a], Giota Stratou[a], Jill Boberg[a], Sebastian Koenig[a], Todd Adamson[b], Albert Rizzo[a]

[a]*University of Southern California, Institute for Creative Technologies*
[b]*U.S. Vets, Long Beach, California*

Abstract. Nonverbal behaviors play a crucial role in shaping outcomes in face-to-face clinical interactions. Experienced clinicians use nonverbals to foster rapport and "read" their clients to inform diagnoses. The rise of telemedicine and virtual health agents creates new opportunities, but it also strips away much of this nonverbal channel. Recent advances in low-cost computer vision and sensing technologies have the potential to address this challenge by learning to recognize nonverbal cues from large datasets of clinical interactions. These techniques can enhance both telemedicine and the emerging technology of virtual health agents. This article describes our current research in addressing these challenges in the domain of PTSD and depression screening for U.S. Veterans. We describe our general approach and report on our initial contribution: the creation of a large dataset of clinical interview data that facilitates the training of user-state sensing technology.

Keywords. Telemedicine, nonverbal behavior, virtual reality, depression

1. Introduction

It has long been recognized that nonverbal behavior plays an important role in human communication [1, 2]. Facial expressions, gaze, body gestures and vocal prosody can reveal information about a person's emotions, mental state and social intentions. Further, people use nonverbal behavior as a tool to shape social interactions. For example, expressions of anger can extract greater concessions in a negotiation [3] whereas smiles, nods and postural mimicry can establish trust and promote intimate disclosure [4, 5].

After some early neglect, the medical community is increasingly embracing the important role nonverbal communication plays in clinical settings [6, 7]. Expert clinicians use nonverbal signals to help build rapport and therapeutic alliance with their patients, allowing patients to feel greater trust and self-disclose clinically-relevant information [2, 5, 6]. Experienced clinicians are also adept at reading a patient's nonverbal behavior to inform their diagnosis [8]. The increased use of standardized patients in medical training – actors that have been trained to exhibit both the verbal and nonverbal characteristics of disease – is, in part, recognition of the important role nonverbal communication plays in heath settings.

[1] Jonathan Gratch. Institute for Creative Technologies, 12015 Waterfront Dr. Playa Vista, CA. 90094, gratch@ict.usc.edu This work is supported by DARPA under contract (W911NF-04-D-0005) and U.S. Army Research, Development, and Engineering Command and. The content does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

This trend is directly threatened, however, by the growing use of technology to mediate health-care interactions. For example, telemedicine allows wider and more efficient distribution of health care, but simultaneously disrupts many of the social cues that make face-to-face interaction so effective (e.g. [9]). While new technology can address some of these challenges (e.g., by reducing video-conference delay or correcting for the inability to establish mutual gaze), it will not address the more fundamental problem that many healthcare providers lack the basic skills effectively use or interpret nonverbal behavior. Another trend towards software-administered health presents even more severe challenges. For example, there is growing interested in "virtual health agents" that interact with patients over the web and perform basic health functions such as screening [10] or medical adherence [11]). These agents express many of the social cues found in face-to-face interactions, but to date, such technology is completely deficient when it comes to understanding and reacting to patient nonverbal behavior.

These observations make a compelling case for computer-based systems that capture and quantify nonverbal behavior in clinical settings. Such techniques could supplement information garnered within the course of traditional telemedicine. For example, a junior clinician could receive a summary of nonverbal indicators of client depression or anxiety. In training, students could receive feedback on their own behavior (e.g., how much mutual gaze did they establish with the patient). When it comes to virtual health agents, detecting and responding to the nonverbal channel could improve the efficiency and effectiveness of such techniques. This is one of the major premises of the interdisciplinary research area of affective computing that focuses on the study and development of systems and devices that can recognize, interpret, process, and simulate human affective states [12]. It also builds on research into perceptual user interfaces [13] which seek to maximize the bandwidth of communication between a user and a computer with such sensing technologies, and aims to enable a user experience with computers that is more similar to the way that people interact in the real world.

This article reports on our recent success in automatically recognizing clinically relevant nonverbal behavior. To this end, we are developing a software package known as *MultiSense* that combines audio and video information to recognize these behaviors. This module acts as a pluggable component that can inform a variety of potential health applications. Here, we report on our current success in applying this approach within the context of two methods (one using a virtual health agent and the employing using telemedicine) for performing PTSD and depression screening for U.S. We first describe these use cases and then focus on recent progress in developing MultiSense. A main initial accomplishment is the creation of a large dataset of clinical interview data that facilitates the training of user-state sensing technology. We discuss this dataset and our initial success in training techniques to recognize clinically-relevant nonverbal signals.

2. SimSensei and TeleCoach Use Cases

SimSensei explores the feasibility of virtual health agents for mental health screening. This kiosk-based system is aimed at clients resistant to seeking traditional care (Fig. 1). The system combines the advantages of traditional web-based self-administered screening [14], which allows for anonymity, with anthropomorphic interfaces, which foster some of the beneficial social effects of face-to-face interactions [5]. SimSensei builds on an earlier web-based screening tool, SimCoach [10], and engages users in a structured interview using natural language and nonverbal sensing with the aim of identify-

Figure 1. SimSensei virtual health agent (on left) and Telecoach interface concept (on right)

ing risk factors associated with depression or PTSD. MultiSence will fuse information from web cameras, Microsoft's Kinect and audio processing to identify moment-to-moment opportunities to provide supportive feedback (as in [15]) and to identify the presence of any nonverbal indicators of psychological distress.

TeleCoach explores the use of user state sensing to inform telemedicine applications. TeleCoach will also use MultiSense to recognize indicators of psychological distress, but rather than guiding intelligent software, this information is presented to a geographically-distant a human healthcare professional. Telecoach acts as a real time decision-support "dashboard" tool. Risk factors identified by MultiSense will be summarized and presented to clinicians with the aim aimed to enhance the quality of the online interaction between remotely located patients and care providers. In essence, the *TeleCoach* project aims to also take information gleaned from face/body gestures and vocal patterns that can be used to infer user state and that information is displayed to the healthcare provider as an additional "channel of insight" that might support real time clinical decision-making and patient care.

3. Nonverbal behavior and Psychological Distress

A large body of research has examined the relationship between nonverbal behavior and clinical conditions and there is general consensus on the relationship between some clinical conditions (especially depression and social anxiety) and associated nonverbal cues. These general findings inform our search for automatic nonverbal behavior descriptors, so we first review these key findings.

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 maintain significantly less mutual gaze [16], show nonspecific gaze, such as staring off into space [17] and avert their gaze, often together with a downward angling of the head [18]. The pattern for depression and PTSD is similar, with patients often avoiding direct eye contact with the clinician.

Emotional expressivity is also diagnostic of clinical state. Depressed patients frequently display flattened affect including less emotional expressivity [18], fewer mouth movements [17, 19], more frowns [18, 19] and fewer gestures [6, 18]. Further, they dynamics of these behaviors may be more important that their quantity. For example, depressed patients may frequently smile, but these are perceived as less genuine and often shorter in duration [20] than in non-clinical populations. Social anxiety and PTSD share features of depression but 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 [20], or shame [21].

Finally, certain gestures are seen with greater frequency in clinical populations. Fidgeting is often reported. This includes gestures such as tapping or rhythmically shaking hands or feet and is seen in both anxiety and depression [19]. Depressed patients also often engage in “self-adaptors” [22], such as rhythmically touching, hugging or stroking parts of the body or self-grooming, such as repeatedly stroking the hair [19].

One brewing controversy is whether categories of mental illness (e.g., depression and PTSD) are distinct categories vs. continuous differences along more general underlying dimensions. A similar controversy arises in automatic emotion recognition where broad dimensions (e.g., valence and arousal) tend to produce higher recognition rates that discrete labels. The broad dimension receiving the most support in clinical studies is the concept of *general distress*. For example, Elhai et al examine a large number of clinical diagnostic interviews and found that diagnoses of major depression and PTSD were better characterized by considering only a single dimension of general distress [23]. Several other researchers have statistically re-examine the standard scales and interview protocols used to diagnose depression, anxiety and PTSD and found they highly correlate and better seen as measuring general distress (e.g. [24]). For this reason, we will investigate if general distress may be a more appropriate concept for recognizing clinical illness in addition to the more conventional discrete categories.

4. Dyadic Interaction Dataset

The fundamental novel research challenge in this project is to endow computers with the ability to recognize clinically-relevant information from the nonverbal behavior of patients (and potentially clinicians for TeleCoach). Computer vision and audio signal processing techniques have shown growing success in identifying an number of important nonverbal cues but the limitation of state-of-the-art approaches is that they are data hungry: the require large amounts of annotated data. Thus, our initial milestone was to collect a large dataset of clinical interviews involved PTSD, social anxiety and depression and to identify and annotate nonverbal behaviors of patients and clinicians that are relevant to finding indicators of these clinical states.

Method: Participants were informed that we are interested in their experience with PTSD and depression. They completed a series of web-delivered psychometric scales to assess clinically-relevant states and traits. They then participated in a 30-minute structured interview that explored if they have been previously diagnosed and are currently experiencing symptoms of PTSD, depression and anxiety.

Participants: 167 participants were recruited from two distinct populations. 110 (54 male) were recruited from the Los Angeles general population through Craigslist, an online job posting service. 57 participants (49 male) were recruited from US Vets, a non-profit organization that helps veterans re-integrate into civilian life after deployment and has programs specifically tailored for veterans with PTSD and depression
Procedure: After obtaining consent, participants were led to a computer and, in private, completed a series of web-based scales including the PCL-C to assess PTSD [25], PHQ-9 to assess depression [26], STAI to assess state anxiety [27], PANAS to assess current mood [28] and BIDR to assess tendencies to be deceptive in such interview

contexts [29]. One of two possible interviewers then re-entered the room and conducted a structured interview consisting of three phases, as warm-up phase consisting of basic questions designed to establish rapport (e.g., “How is your day going? Where are you from”), an interview phase where participants were asked to elaborate on some of their responses to the scales (e.g., “On the survey you mention you often experience disturbing thoughts; can you tell me a little more about that?”), followed by a wind-down phase designed to return the participant to a more pleasant state of mind (e.g., “If you could travel to any destination, where would you go?”). During the interview phase, both the participant and the interviewer are fitted with a lapel microphone and are recorded with video cameras and the Kinect system to track their body posture. The video cameras and Kinect are placed between the participants. Following the interview, we assessed the quality of the interaction with measures of rapport and social presence.

Summary Statistics: Overall, 32% assessed positive for PTSD, 29% for depression, and 62% for trait anxiety. Participants at US Vets were assessed positive more often for each of the disorders (see Figure 2). The US Vets and Craigslist populations were distinct in several ways. US Vets participants were older, less educated, more male, more unemployed, and much more likely to have been a member of the armed forces. Participants with disorders were significantly higher in neuroticism and were more anxious before the interview. The two interviewers were not significantly different from each other in any of the scales used to assess them.

Consistent with the findings on general distress discussed in Section 3, we observed significant correlations ($p<0.01$) between the disorders (i.e. PTSD, anxiety, and depression). Depression correlated with PTSD (ρ 0.64, using Pearson’s correlation) and with anxiety (ρ 0.40 ; PTSD correlated with anxiety (ρ 0.43). The scalar severity measure of the three inventories showed even stronger correlations (ρ >0.8). Based on prior findings on general distress and our comorbidity observed, we conclude that automatic recognition techniques should focus on recognizing general distress rather than attempting to distinguishing individual conditions. As a result, we use factor analysis to identify a single of distress that is used in subsequent training and analysis.

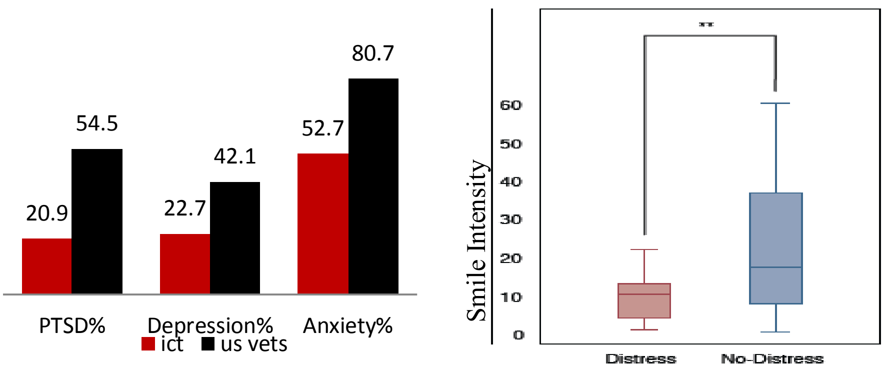

Figure 2. Chart on the left summarizes prevalence of clinical conditions in Craigslist (ict) and US Vets populations. Chart on right illustrates the different pattern of smile intensities found in the samples.

Clinical Cues: The dataset was annotated with manual and automatic techniques to identify nonverbal (audio and visual) behaviors that might be associated with generalized distress. All manual annotators were trained until their reached high inter-coder agreement. Automatic features include head orientation, gaze orientation and smile intensity and duration. Manual features include hand self-adaptors (i.e., self-touching) and leg fidgeting. Automatic features included angle of gaze, smile intensity and duration and several features related to vocal quality.

We found several statistically significant differences in the visual behavior of participants between those that scored positive for general distress and normal controls. (1) There are significant differences in the automatically estimated gaze behavior of subjects with psychological disorders. In particular, an increased overall downwards angle of the gaze could be automatically identified using two separate automatic measurements, for both the face as well as the eye gaze. (2) Using automatic measurements, we could identify on average significantly less intense smiles for subjects with psychological disorders as well as significantly shorter average durations of smiles. (3) Based on the manual analysis, subjects with psychological conditions exhibit on average longer self-touches and fidget on average longer with both their hands (e.g. rubbing, stroking) as well as their legs (e.g. tapping, shaking).

We found several significant differences in the vocal patterns of participants with general distress related to the 'coloring' of the voice when compared with normal controls (for this we only analyzed male participants to control for differences in vocal quality that arise from gender). We examined differences in vocal fundamental frequency, speech intensity, measures of monotonicity (i.e. intensity variations and spectral stationarity), and measures of the voice's breathiness (e.g. normalized amplitude quotient (NAQ)). The most promising findings are that the speech intensity variations of distressed subjects are significantly reduced and their voice quality is significantly breathier based on the observed NAQ parameter.

These results replicate findings in the psychological literature and give us confidence that these indicators can be identified automatically in real-time interactions using low-cost sensors. Current efforts seek to extend this analysis to dyadic behaviors and to train automatic recognizers that will be incorporated into MultiSense.

5. Conclusions

Integrating low-cost sensing systems to promote bi-directional engagement with both virtual health agents and telemedicine providers could drive the development of increasingly effective ways for humans to interact and benefit from such applications. In addition to enhancing the quality of interaction for healthcare applications, this research could produce generalizable knowledge/value for educational application where enhanced bi-directional interaction fidelity with virtual health agents and live trainers/educators could improve engagement, bonding and ultimate effectiveness. Any virtual agent application, where the engagement typical of a human "Mentor-Apprentice" relationship is believed to promote learning, could benefit from success in this work. Such applications might include: general education (e.g. mathematics, foreign languages, and other STEM areas), medical training (e.g. paramedic skills, diagnostic decision-making), elderly support (e.g. agents that assist in budgeting, shopping, distance education, selective reminding), and entertainment. How this affects user interaction and consequent task performance across a variety of important domains will generate many questions—the answers to which could fundamentally impact the way computational devices are used and regarded throughout the rest of the 21st century!

References

1. Darwin, C., *The Expression of the Emotions in Man and Animals*. 3rd ed2002: Oxford Uni. Press.
2. Ambady, N. and R. Rosenthal, *Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis*. Psychological Bulletin, 1992. **111**(2): p. 256-274.
3. Van Kleef, G.A., C.K.W. De Dreu, and A.S.R. Manstead, *The interpersonal effects of anger and happiness in negotiations*. Journal of Personality and Social Psychology, 2004. **86**(1): p. 57-76.
4. Tickle-Degnen, L. and R. Rosenthal, *The Nature of Rapport and its Nonverbal Correlates*. Psychological Inquiry, 1990. **1**(4): p. 285-293.
5. Kang, S.-H. and J. Gratch, *Virtual Humans Elicit Socially Anxious Interactants' Verbal Self-Disclosure*. Journal of Visualization and Computer Animation, 2010. **21**(3-4): p. 473-482.
6. Hall, J.A., J.A. Harrigan, and R. Rosenthal, *Nonverbal behavior in clinician—patient interaction*. Applied and Preventive Psychology, 1995. **4**(1): p. 21-37.
7. Philippot, P., R.S. Feldman, and E.J. Coats, eds. *Nonverbal behavior in clinical settings*. Affective Sciences, ed. R. Davidson, P. Ekman, and K. Scherer2003, Oxford University Press: New York.
8. Boice, R. and P. Monti, *Specification of nonverbal behaviors for clinical assessment* Journal of Nonverbal Behavior, 1982. **7**(2): p. 79-94.
9. Manning, T.R., E.T. Goetz, and R.L. Street, *Signal delay effects on rapport in telepsychiatry*. CyberPsychology and Behavior, 2000. **3**(2): p. 119-127.
10. Rizzo, A., et al. *SimCoach: an intelligent virtual human system for providing healthcare information and support*. in *I/ITSEC*. 2011. NTSA.
11. Bickmore, T. and L. Pfeifer, *Relational Agents for Antipsychotic Medication Adherence*, in *CHI'08 workshop on Technology in Mental Health*2008: Florence, Italy.
12. Picard, R.W., *Affective Computing: From Laughter to IEEE*. Affective Computing, IEEE Transactions on, 2010. **1**(1): p. 11-17.
13. Oviatt, S. and P. Cohen, *Perceptual user interfaces: multimodal interfaces that process what comes naturally*. Commun. ACM, 2000. **43**(3): p. 45-53.
14. Weisband,& Kiesler, *Self disclosure on computer forms: meta-analysis and implcations*, *CHI*1996.
15. Gratch, J., et al. *Virtual Rapport*. in *6th International Conference on Intelligent Virtual Agents*. 2006. Marina del Rey, CA: Springer.
16. Waxer, P., *Nonverbal cues for depression*. Journal of Abnormal Psychology, 1974. **83**(3)
17. Schelde, J.T.M., *Major Depression: Behavioral Markers of Depression and Recovery*. The Journal of Nervous and Mental Disease, 1998. **186**(3): p. 133-140.
18. Perez, J.E. and R.E. Riggio, *Nonverbal social skills and psychopathology*, in *Nonverbal behavior in clinical settings*, Philippot, et al.,, Editors. 2003, Oxford University Press: New York.
19. Fairbanks, L.A., M.T. McGuire, and C.J. Harris, *Nonverbal interaction of patients and therapists during psychiatric interviews*. Journal of Abnormal Psychology, 1982. **91**(2): p. 109-119.
20. Kirsch, A. & S. Brunnhuber, *Facial Expression & Experience of Emotions in Psychodynamic Interviews with Patients w/ PTSD in Comparison to Healthy Subjects*. Psychopath., 2007. **40**(5).
21. Menke, R., *Examining nonverbal shame markers among post-pregnancy women with maltreatment histories*, in *Clinical psychology*2011, Wayne State University.
22. Ekman, P., *Universals and cultural differences in facial expressions of emotions*, in *Nebraska Symposium on Motivation*, J. Cole, Editor 1972, University of Nebraska Press: Lincoln, Nebraska.
23. Elhai, J.D., et al., *Testing whether posttraumatic stress disorder and major depressive disorder are similar or unique constructs*. Journal of Anxiety Disorders, 2011. **25**(3): p. 404-410.
24. Marshall, G.N., T.L. Schell, and J.N.V. Miles, *All PTSD symptoms are highly associated with general distress: Ramifications for the dysphoria symptom cluster*. Journal of Abnormal Psyc., 2010. **119**(1)
25. Blanchard, E.B., et al., *Psychometric properties of the PTSD checklist (PCL)*. Behaviour Research and Therapy, 1996. **34**(8): p. 669-673.
26. Kroenke, K., R.L. Spitzer, and J.B.W. Williams, *The PHQ-9*. Journal of General Internal Medicine, 2001. **16**(9): p. 606-613.
27. Spielberger, C.D., et al., *Manual for the State-Trait Anxiety Inventory*1970: Consulting Psyc. Press.
28. Watson, D., A. Tellegen, and L.A. Clark, *Development and validation of brief measures of positive and negative affect: The PANAS scale*. JPSP, 1988. **54**: p. 1063-1070.
29. Paulhus, D., *Balanced inventory of desirable responding (BIDR)*. Acceptance and Commitment Therapy. Measures Package, 1988: p. 41.