To Rank or to Classify?
Annotating Stress for Reliable PTSD Profiling

Christoffer Holmgård
Center for Computer Games Research, IT University of Copenhagen
Rued Langgaards Vej 7, Copenhagen, Denmark
Email: holmgard@itu.dk

Héctor P. Martínez
Institute of Digital Games, University of Malta
Msida MSD 2080, Malta
Email: hector.p.martinez@um.edu.mt

Georgios N. Yannakakis
Institute of Digital Games, University of Malta
Msida MSD 2080, Malta
Email: georgios.yannakakis@um.edu.mt

Karen-Inge Karstoft
Research and Knowledge Centre, Danish Veteran Centre
Garnisonen 1, Ringsted, Denmark
Email: kikarstoft@health.sdu.dk

Abstract—In this paper we profile the stress responses of patients diagnosed with post-traumatic stress disorder (PTSD) to individual events in the game-based PTSD stress inoculation and exposure virtual environment *StartleMart*. Thirteen veterans suffering from PTSD play the game while we record their skin conductance. Game logs are used to identify individual events, and continuous decomposition analysis is applied to the skin conductance signals to derive event-related stress responses. The extracted skin conductance features from this analysis are used to profile each individual player in terms of stress response. We observe a large degree of variation across the 13 veterans which further validates the idiosyncratic nature of PTSD physiological manifestations. Further to game data and skin conductance signals we ask PTSD patients to indicate the most stressful event experienced (class-based annotation) and also compare the stress level of all events in a pairwise preference manner (rank-based annotation). We compare the two annotation stress schemes by correlating the self-reports to individual event-based stress manifestations. The self-reports collected through class-based annotation exhibit no correlation to physiological responses, whereas, the pairwise preferences yield significant correlations to all skin conductance features extracted via continuous decomposition analysis. The core findings of the paper suggest that reporting of stress preferences across events yields more reliable data that capture aspects of the stress experienced and that features extracted from skin conductance via continuous decomposition analysis offer appropriate predictors of stress manifestation across PTSD patients.

Keywords—PTSD; stress detection; annotation; ranking

I. INTRODUCTION

This paper describes a game-based method for profiling *post traumatic stress disorder* (PTSD) using affective responses to events in a virtual environment incorporating principles of game design, called the *StartleMart* game [8], [9]. We describe how we construct a specific paradigm for eliciting and capturing affective responses to particular in-game events of 13 PTSD patients through the StartleMart game. We then use these responses to enable individual PTSD profiling and stress detection through continuous decomposition analysis [2] of *skin conductance* (SC) manifestations to in-game stressor

events. Our results indicate that participants' memories of the most stressful events correspond poorly to SC responses, but their ordering of events in terms of stressfulness corresponds strongly to these same responses.

The work presented here is inspired by previous pioneering work on the use of virtual environments and games as tools supporting affective learning in PTSD patients [15], [21] and it builds upon previous work [8], [9] in which it is demonstrated that features from affective responses to game play sessions in the StartleMart game contain useful indications of the symptom severity of PTSD patients, pointing to the relevance of simulations and games for affective learning and profiling. Here, we move from correspondence analysis at the game session level to analyzing *event*-based stress responses in order to obtain not only general indications of the patient's affective response but also to identify responses to individual events. By identifying and profiling responses to stressful situations in war veterans suffering from PTSD we create a novel and efficient method for understanding the syndrome configuration of the individual patient.

The paper is novel in that continuous decomposition analysis is applied for extracting appropriate indicators of sympathetic arousal from skin conductance in PTSD patients. Such an approach allows us to derive stress detectors such as tonic and phasic drivers of skin conductance soon after a stressful event (elicitor) is presented to users [1], [2]. Additionally we, for the first time, compare two different stress self-annotation schemes for their consistency to manifested stress via skin conductance. Our results further validate evidence and observations in the literature [13], [14], [24], [25] suggesting that rank-based (compared to class-based or rating-based) annotation yields better approximators of the ground truth of experienced emotion.

II. BACKGROUND: PTSD, GAMES, AND AFFECT

In this section we describe the PTSD syndrome, the relationship between games and PTSD treatment, PTSD's links to

human physiology and its affective manifestations.

A. Post Traumatic Stress Disorder

Post Traumatic Stress Disorder (PTSD) is a psychiatric diagnosis describing an often severely disabling syndrome that is sometimes developed after being exposed to highly stressful situations. Veterans from military operations are a high-risk group for developing this syndrome [6].

Two well-known treatment approaches for PTSD — favored because of strong evidence for their therapeutic efficacy — are the *cognitive-behavioral therapy* techniques of exposure therapy and stress inoculation training. In exposure therapy, the therapist confronts the patient with anxiety provoking stimuli in a controlled setting in order to extinguish reactions to the stimuli and/or allow the patient to reprocess the memories cued by the stimuli. Three common variations are the use of real life stimuli i.e. *in vivo*, representing stimuli via media i.e. *mediated*, or having the patient imagine the stress provoking situations and thus self-generate the stimuli i.e. *imaginal* [5]. In stress inoculation training, the therapist exposes the patient to stimuli and situations that are not directly linked to the original trauma of the patient, but that cause problematic anxiety responses that are difficult for the patient to cope with [5].

B. Games for PTSD treatment

Games and game-like worlds have successfully been used as mental health interventions by appropriating commercial games [7] and by developing specialized solutions [10]. Among the possible ways of treating PTSD, computer games and virtual environments have a particular potential for eliciting stress in a controlled, graded fashion and can provide an immersive and rich medium for PTSD treatment [19], [19], [21], [22]. Earlier research has demonstrated the usefulness of virtual environments for treating veterans' PTSD with *virtual reality therapy*, an extension of exposure therapy [15], [23]. Some implementations of virtual reality therapy have focused on exposing the patient to the original stressful, traumatizing situation, in the vain of classic exposure therapy. Notable examples are the *Virtual Iraq* and *Virtual Afghanistan* applications that show promising results in clinical testing [18], [19]. Other implementations have focused on appropriating principles from stress inoculation training [21]. StartleMart implements a hybrid of exposure therapy and stress inoculation training.

The conversation between the user and the therapist is central for treatment efficacy [5]. The user's perception of which events are stressful, and the user's actual physiological stress responses to the same events, are important touchstones in this conversation. Therefore, we see a need for self-report schemes that allow the user to accurately report which events were perceived as stressful. We address this by comparing two schemes for self-reporting experiences of stressful events, classification and ranking, with physiological responses as the ground truth.

C. Physiology of PTSD

When placed in mediated stimulus exposure paradigms, PTSD-patients exhibit physiological responses to stressful visual and auditive stimuli that are significantly different from the responses of non-patients [16]. Their responses are generally characterized by high sympathetic activity as measured by SC. Responses that have been found to be robust indicators of PTSD conditions via experimental studies include slower SC habituation, elevated resting SC, and greater SC responses to startling stimuli [17]. In general, higher base levels of arousal and heightened sensitivity to stress seem to characterize the physiological manifestations of the disorder. It has been suggested that these differences could be used to support diagnostic differentiation between PTSD patients and non-patients as well as between different degrees of PTSD symptom severity [3] guiding treatment strategies or allowing for adaptive treatment tools [23].

Although the general symptomatology of PTSD is consistent across sufferers, every instance of the syndrome includes idiosyncratic aspects related to the particular individual and the instigating stressful experience. Which particular events elicit the strongest stress-response or trigger flashbacks vary across individuals and research has shown the strength of the response to be contingent upon the stimulus relation to the original trauma [12]. This characteristic of the PTSD syndrome proves a challenge when developing any treatment or diagnostic tool as we assume that greater efficacy and precision comes at a cost of lower generalizability. For virtual environments, this is particularly challenging as convincing environments are time-consuming and expensive to develop. In order for a virtual environment for PTSD to be as generally useful and cost-effective as possible, it should apply to as wide a range of PTSD symptomatologies as possible.

A design solution to this challenge is presented in the stress-inoculation training approach taken in this study. By choosing a stimulus delivery environment that is predominantly related to the everyday strivings of PTSD patients [11], rather than the original trauma, some specificity is sacrificed, but a greater relevance across patients is attempted. The general relevance of the approach has been demonstrated in [8] and [9] which focus on feature extraction from full length sessions, while here we move to an event-level analysis. Simultaneously, the everyday environment supports a patient/therapist conversation about the stimuli present in everyday life that are most stressful to the particular patient, supporting the stress inoculation process.

In the following section we outline the key features of the StartleMart game; for the interested reader the game is described in greater detail in [8], [9].

III. The StartleMart Game for PTSD Treatment

The task of shopping in a supermarket is a common situation that is severely challenging to many patients suffering from PTSD [11]. Supermarkets are highly stimulating environments with social interactions and unpredictable auditive and visual experiences which PTSD patients find stressful; some to the

Fig. 1. The three flashbacks of the game (left column) and the immediately preceding supermarket scenes (right column). Elements of the supermarket bleed into the flashbacks, simulating re-experience.

extent that they avoid going shopping or only do so with a helper present for emotional support. Consequently, the game is built to primarily take place in a virtual supermarket (see Fig. 1). The supermarket environment includes a number of stressors that aim at eliciting stress in the player. These are designed around three typical symptoms of PTSD, namely *fear-avoidance behavior*, *hyper-arousal* (i.e. heightened startle response), and *re-experiencing* of traumatic events triggered by an outside stimulus or general stress [5]. Stressors targeting *fear-avoidance behavior* include the layout of the supermarket which is designed to include hidden angles and preventing the player from attaining a full overview of the location and Non-Player-Characters (NPCs) that provide socially stressful experiences. Stressors targeting *hyper-arousal* include a dog barking at the entrance to the supermarket and the sound of crashes and glass breaking suddenly playing at random locations in the supermarket. Stressors targeting *re-experiencing* are included in the form of three different flashbacks.

IV. EXPERIMENTAL PROTOCOL AND DATA COLLECTION

In this section we provide details about the participants of our experiment and the experimental protocol followed. Thirteen male PTSD patients, veterans from Danish military operations in Afghanistan, are included in the study presented in this paper. The participants are in psychiatric treatment for PTSD and qualify for the PTSD diagnosis. All subjects

are medicated with *Selective Serotonin Re-uptake Inhibitors (SSRI)* which is known to generally lower sympathetic activity and in particular SC [20]. This clearly adds a challenge to the detection of SC stress responses to game stimuli since patients are expected to manifest responses that are pharmacologically suppressed to an unknown degree. Each patient participates in the experiment twice, engaging in a total of 6 game play sessions, 3 per participation (11 patients have participated in both sessions, while 3 participated in the first session only), progressing through low stress intensity, medium stress intensity, and high stress intensity. The experimenters, trained psychologists, welcome the participant, complete a diagnostic interview with the patient and collect various instances of demographic and background data from either the patient himself or the patient's medical records. The participant is introduced to the experimental setup and seated in front of the controls and monitor. The biofeedback device is attached to the participant's fingertips (see more details in Section IV-A), and a brief introduction to the game rules and how to control the game is given. Following a short waiting period, collecting baseline SC data, the participant is asked to play three sessions of the game. Subjective data (self-reports) is collected over the course of the experiment. Finally, the experimenter debriefs the participant, responding to any concerns or issues the patient might have.

A. Physiological Sensors and Setup

For continuous measurement of SC the IOM biofeedback device[1] is used. The IOM biofeedback device samples SC at a rate of 300 Hz and down-samples them to 30 Hz in firmware before transmitting them to the recording computer. The device's measuring electrodes are attached dryly to the distal phalanges of the little and middle fingers of the patient's non-dominant hand. A sensor measuring blood volume pulse is attached to the ring finger, but is not used for the study and analysis presented here. Since frustration with the control scheme of the game might introduce unwanted variation and artifacts to the results of the experiment [26] the game is configured to use standard controls for first-person-perspective computer games which should be familiar to most patients. The mouse, operated with the patient's dominant hand, controls the perspective and the keyboard controls movement. To minimize the risk of movement artifacts in the physiological readings, patients operate the keyboard (W, A, S, D or arrow keys) with only the index finger of their non-dominant hand, keeping the other fingers still.

B. Game Logging

During game play, a number of features are logged constantly. A screen-shot from the player's perspective is logged every second to allow for reconstruction of the events of the game play session and for subsequent identification of the most stressful experiences. Stressor stimulus presentations are logged as *game events* whenever they occur. The four types of

[1]http://www.wilddivine.com/

events are labeled as: 1) *sound* events, when sudden sounds of crashes and glass breaking are played, 2) *pickup* events, when the player obtains one of the items on the shopping list, 3) *social* events, when the player is close to one or more of the NPCs in the supermarket and 4) *flashback* events, when the flashback of the session is presented.

V. Self-reports and SC Feature Extraction

In this section we describe the data collected from the two stress self-report schemes (Section V-A) and the SC features extracted via continuous composition analysis (Section V-B).

A. Stress Self-reports

At the end of each session, the patient is asked to indicate which singular event during the session was considered the most stressful, if any. This self-report is subsequently reduced to one of the four event categories: social, sound, pickup, and flashback. The player is then presented with a *4-alternative-forced-choice* (4AFC) survey comprised of series of pairs constructed from all events that happened during the game. The two screen-shots that were taken closest in time to each event are presented side by side as a representation of each event. The player is asked to express a preference for which of the two events that was most stressful, if both were equally stressful, or if neither of the events were stressful. The player completes this process for pairings of all logged events producing a global ordering of the events in terms of stressfulness.

B. Skin Conductance Event Response Features

The trough-to-peak analysis of skin conductivity response (SCR) amplitude, area or similar measures, can be subject to super-positioning of phasic and tonic activity. This may necessitate the subtraction of baseline measures or other forms of signal correction [4]. It has been suggested that even with such corrections one may still confound phasic and tonic SC which is undesirable in a study focusing predominantly on event-related activation [2].

To address this potential issue, features of the player's SC at the time of the event are extracted using Continuous Decomposition Analysis (CDA) as described in [2]. The method allows for the decomposition of phasic and tonic electrodermal activity. It initially separates super-positioned phasic and tonic components of the raw SC signal. Subsequently it adapts a general model of the human skin's impulse response function (representing the basic SCR shape that would result from a unit impulse) to the phasic activity by sampling the tonic component around the event response to establish a local baseline and fitting the general impulse response function to the shape of the phasic component. The result is expressed in a *phasic driver* measured in μS that approximates the phasic response affecting the signal within the event window. As such, the phasic driver across the event window can be interpreted as a locally baseline-corrected measure of the patient's SC response to the event. As a result of the decomposition procedure the phasic driver value can take on negative values.

A detailed example from the CDA process is provided in Fig. 2. More details about the CDA method can be found in [2]. A $1 - 4$ s after-event response window is applied, meaning that only activation occurring with this window is considered relevant to the event (see Fig. 2). A minimum phasic driver threshold value of 0.05 μS is used, meaning that only events with a phasic driver value exceeding this threshold are considered significant and counted as SCRs.

From the CDA result, four skin conductance response features are extracted for each event in the game: the mean phasic driver within the skin conductance response window (S_p), the integral of the phasic driver within the response window (i_p), the mean tonic SC (S_t) within the window, and the global mean within the window (S_g). The literature suggests that features based on the phasic driver of skin conductance are supreme detectors of heightened sympathetic arousal [2]. We therefore trust that the four features extracted are appropriate indicators of stress elicited around the stressful events provided by the game [1], [4].

VI. Results

In this section we present a descriptive overview of the SC responses of the patients of the study to the included events demonstrating the individual differences in responses to the events in the StartleMart and the complex relationship between physiological responses and subjective experience. Moreover, we investigate the relation between the two self-report schemes and the extracted SC features. First, in Section VI-A we correlate patient classifications of the most stressful event with the extracted skin conductance features. Then, in Section VI-B we correlate preferences of the stressfulness of events with the same features.

58 sessions were deemed free of protocol and uncorrectable sensor artifacts and thus usable for analysis. Sessions ranged in length from 66 to 202 seconds, with an average length of $144.8s$ and a standard deviation of $36.3s$. Together, these sessions provided 479 events over the defined phasic driver threshold value, an average of 8.3 events per session, with a range from 2 to 11 above threshold events per session and a standard deviation of 2.2 events.

A. Classifying Stress

Our first approach to obtaining self-reports from patients was to ask them to pick of the most stressful event during the game session. All events in each session of the dataset are annotated with this classification: labeling an event as either the most stressful or not. A descriptive overview of chosen events and S_p values for each patient and each session is provided in Fig. 3. The figure shows the differences between individuals in physiological responses to events, differences in self-reports of experienced stress, and discrepancies between the physiological responses to events and self-reports of which events were experienced as most stressful. Most patients exhibit a tendency to report the flashbacks of StartleMart as the most stressful type of experience; however, their physiological responses across events indicate that responses to other types
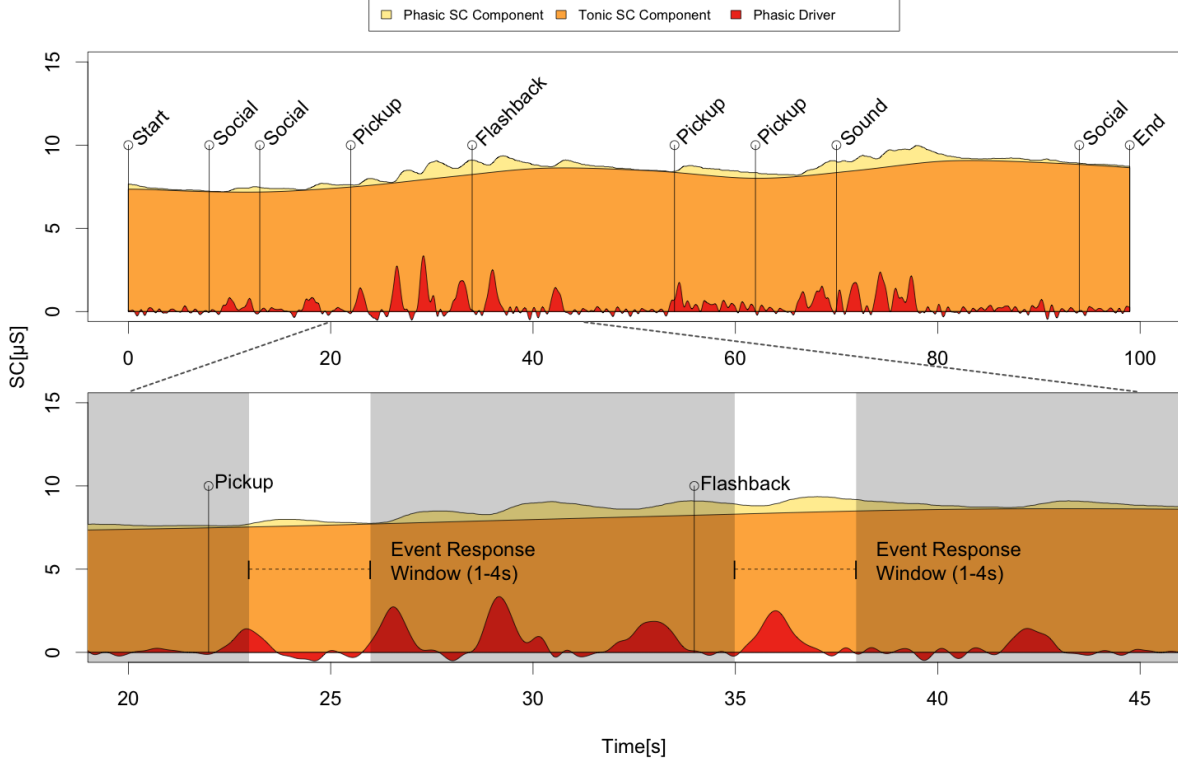
Fig. 2. Continuous Decomposition Analysis of SC of Player 5, Session 3. The top graph shows the full game session whereas the bottom graph shows a detailed view of an excerpt from that session. Both graphs depict three components extracted from the raw SC signal: Phasic activity (yellow), tonic activity (orange), and the phasic driver (red) of SC. SC features are extracted within the event response windows as illustrated in the bottom graph.

of events, on average, are stronger. Figure 3 indicates the need of more complex, non-linear, models for mapping the physiological responses of patients to the perceived experience of stress from interacting with the StartleMart game. Further it demonstrates illustratively the inconsistency between the most stressful events reported by the patients and heightened sympathetic arousal (via the mean phasic driver feature).

To investigate the relationship between the events classified as most stressful and their corresponding SC features values, the annotated events are correlated to the four extracted SC features using the binomially-distributed pairwise correlation described in [24] which is calculated as: $c(z) = \sum_i \{z_i\}/N$; where N is the sample size, and z_i is 1 if there is an agreement between the annotation (e.g. most stressful) and the corresponding SC feature and -1 otherwise. The $c(z)$ values are calculated by considering annotation agreements with the most stressful event as determined by physiology (i.e. the corresponding SC feature). To further explore potential effects between reported and manifested stress we also calculate potential agreements between the reported event and the three, five and ten most stressful events as manifested by SC features. The results reported in Table I showcase significant negative correlations when the single, three and five most stressful events are considered. This already indicates a poor consistency between class-based self-reports and physiological indications of stress. Only when the ten most stressful events

are considered we start observing a minor positive correlation which demonstrates that some of the largest physiological responses are captured in the class (most stressful event reports) only when 10 of those events are considered.

B. Ranking Stress

Our second approach to obtaining reliable stress self-reports from PTSD patients uses the fully ordered preference-pairs constructed from the 4AFC selections. These are pairwise-correlated to the extracted SC also using the $c(z)$ test statistic. The p-value, in this case, is obtained from the normal distribution as the binomially-distributed $c(z)$ approximates the normal distribution when large samples are considered. As shown in Table I significant positive correlations are found between reported ranks of stress and all four extracted SC features. This suggests that reported stress preferences expressed through 4AFC yield event orderings that are highly consistent with the orderings of the physiological responses as measured via any of the four features. In particular, S_p and i_p demonstrate the highest correlation values with reported stress preferences. Based on the obtained results it appears that preference-based annotation of stress is a reliable self-report measure of stress elicitation from events in PTSD patients. It is also obvious that classifying events as most stressful is an annotation practice that fails to capture the idiosyncratic nature of both PTSD and its physiological manifestations.
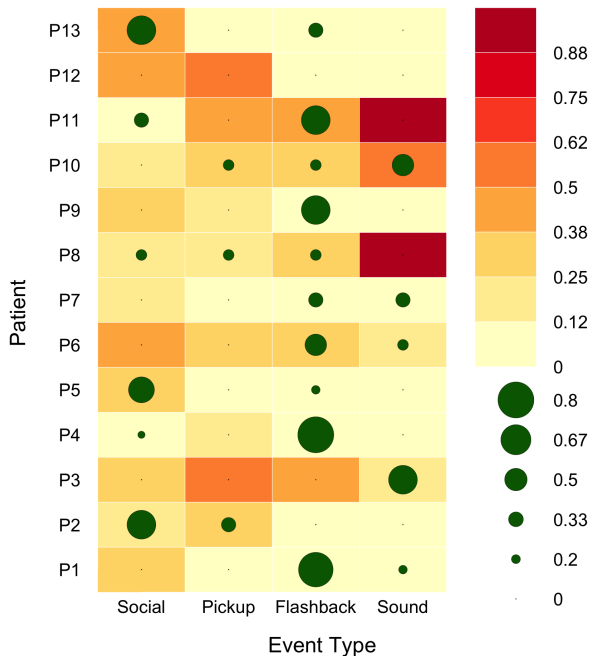
Fig. 3. Individual event response patterns: Patients and event types are cross tabulated. The color of each cell represents the patient's mean normalized S_p response to the event type, across all sessions, extracted using Continuous Decomposition Analysis [2]. The size of each green dot indicates the number of times the patient picks the corresponding type of event as the most stressful. For fair comparison purposes, both the event response levels and the event preference frequencies are normalized into $[0, 1]$ within each individual, across all sessions.

TABLE I

PAIRWISE RANK-CORRELATIONS ($c(z)$) AND CORRESPONDING P-VALUES (p) BETWEEN THE MOST STRESSFUL EVENT REPORTED (CLASS) OR THE REPORTED RANKINGS OF STRESSFUL EVENTS (PREFERENCE) AND THE FOUR SKIN CONDUCTANCE FEATURES.

		S_p	i_p	S_t	S_g
Class	$c(z)$	-0.86	-0.86	-0.76	-0.76
(Most stressful event)	p		< 0.01		
Class	$c(z)$	-0.45	-0.45	-0.24	-0.31
(3 most stressful events)	p	< 0.01	< 0.01	0.02	< 0.01
Class	$c(z)$	-0.17	-0.17	-0.17	-0.17
(5 most stressful events)	p		0.04		
Class	$c(z)$	0.14	0.14	0.14	0.14
(10 most stressful events)	p		0.06		
Preference	$c(z)$	0.43	0.43	0.33	0.37
	p		< 0.01		

VII. DISCUSSION AND CONCLUSIONS

In this paper, we have presented a set of key findings on PTSD profiling and stress detection via games from a sample of 13 clinical PTSD patients interacting with the StartleMart game. Earlier work has demonstrated that this game has an ability to instigate and detect stress in patients to a degree that scales with PTSD symptom severity [8]. The work presented here extends that study by characterizing the individual patient symptoms from subjective experience, event-based skin conductance responses, and game logging data. Further, we introduce continuous decomposition analysis for extracting features of skin conductivity near the stressful in-game events that are appropriate indicators of stress responses.

The study demonstrates the challenges of fusing interactive environments with a stimulus exposure approach, since the agency afforded to the player significantly impacts the experimenter's or therapist's control over the flow of events. Our analysis shows a high degree of inter-subject variability in terms of which events the patients responded strongest to as measured via physiology, and which events they reported as being subjectively most stressful. This finding is not surprising given the general literature on PTSD, but underlines that PTSD patients' responses to events in virtual environments exhibit the same variation across subjects as responses in more controlled stimulus-exposure paradigms or in everyday life. When asking patients to recall the most stressful event in a session and comparing this to physiological responses around the corresponding events, no significant effect is observed. Speculatively, this may be due to memory effects where patients attribute the experienced stress to the most salient and heterogeneous event in the session. However, when patients are presented with a memory cue of each event individually and rank all of them in relation to each other, a strong correlation emerges between self-reports of stress and physiological responses. This suggests that using preference-based ranking paradigms, such as the 4AFC, provides better support for recalling the experienced stress in response to individual events even though it comes with an additional effort of comparing across all possible combinations of experienced events. It is important to note that we did not consider the comparison between rating-based annotation against the rank-based and the class-based approaches in this study. Ratings are ordinal values which are already obtained via the reported preferences of the PTSD patients [13]. Further, this study is based on a relatively narrow, homogeneous sample of patients and generalizability to other conditions or groups is an open question.

The core findings of the paper indicate that rank-based stress annotation to a series of events (i.e. rank two or more stressful events) is a beneficial method for detecting stress compared to class-based (i.e. what is the most stressful event) annotation which further validates the observations of earlier studies in affect annotation and modeling [13], [14], [24], [25]. Results also suggest that the phasic driver (as obtained from continuous decomposition analysis) is a highly reliable predictor of PTSD severity manifested via skin conductance responses to in-game events. It is possible that these findings may extend to other affective disorders with arousal components, such as e.g. anxiety, though this remains an open question for future work.

REFERENCES

[1] Dominik R Bach and Karl J Friston. Model-based analysis of skin conductance responses: Towards causal models in psychophysiology. *Psychophysiology*, 50(1):15–22, 2013.

[2] Mathias Benedek and Christian Kaernbach. A continuous measure of phasic electrodermal activity. *Journal of neuroscience methods*, 190(1):80–91, 2010.

[3] Jens Blechert, Tanja Michael, Paul Grossman, Marta Lajtman, and Frank H. Wilhelm. Autonomic and respiratory characteristics of post-traumatic stress disorder and panic disorder. *Psychosomatic Medicine*, 69(9):935–943, 2007.

[4] W. Boucsein. *Electrodermal activity*. Springer Verlag, 2011.

[5] Edna B. Foa, Terence M. Keane, Matthew J. Friedman, and Judith A. Cohen. *Effective treatments for PTSD, practice guidelines from the International Society for Traumatic Stress Studies*. New York: Guilford Press, second edition, 2009.

[6] Charles W Hoge, Carl A Castro, Stephen C Messer, Dennis McGurk, Dave I Cotting, and Robert L Koffman. Combat duty in iraq and afghanistan, mental health problems, and barriers to care. *New England Journal of Medicine*, 351(1):13–22, 2004.

[7] Emily A Holmes, Ella L James, Thomas Coode-Bate, and Catherine Deeprose. Can playing the computer game "tetris" reduce the build-up of flashbacks for trauma? a proposal from cognitive science. *PLoS One*, 4(1):e4153, 2009.

[8] Christoffer Holmgård, Georgios N. Yannakakis, Karen-Inge Karstoft, and Henrik Steen Andersen. Stress detection for ptsd via the startlemart game. In *Proceedings of ACII*, 2013.

[9] Christoffer Holmgård, Georgios N Yannakakis, Héctor P Martínez, Karen-Inge Karstoft, and Henrik Steen Andersen. Multimodal ptsd characterization via the startlemart game. *Journal on Multimodal User Interfaces*, pages 1–13, 2014.

[10] Mohammed E Hoque, Joseph K Lane, Rana El Kaliouby, Matthew Goodwin, and Rosalind W Picard. Exploring speech therapy games with children on the autism spectrum. 2009.

[11] Todd B Kashdan, William E Breen, and Terri Julian. Everyday strivings in war veterans with posttraumatic stress disorder: Suffering from a hyper-focus on avoidance and emotion regulation. *Behavior therapy*, 41(3):350–363, 2010.

[12] Israel Liberzon, Stephan F Taylor, Richard Amdur, Tara D Jung, Kenneth R Chamberlain, Satoshi Minoshima, Robert A Koeppe, and Lorraine M Fig. Brain activation in ptsd in response to trauma-related stimuli. *Biological psychiatry*, 45(7):817–826, 1999.

[13] Hector Martinez, Georgios Yannakakis, and John Hallam. Don't classify ratings of affect; rank them! *IEEE Transactions on Affective Computing*, 2014.

[14] Angeliki Metallinou and Shrikanth Narayanan. Annotation and processing of continuous emotional attributes: Challenges and opportunities. In *Automatic Face and Gesture Recognition (FG), 2013 10th IEEE International Conference and Workshops on*, pages 1–8. IEEE, 2013.

[15] T.D. Parsons and A.A. Rizzo. Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: a meta-analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, 39(3):250–261, 2008.

[16] C.H. Perala. Galvanic skin response as a measure of soldier stress. Technical report, DTIC Document, 2007.

[17] Nnamdi Pole. The psychophysiology of posttraumatic stress disorder: A meta-analysis. *Psychological Bulletin*, 133(5):725, 2007.

[18] G.M. Reger, K.M. Holloway, C. Candy, B.O. Rothbaum, J.A. Difede, A.A. Rizzo, and G.A. Gahm. Effectiveness of virtual reality exposure therapy for active duty soldiers in a military mental health clinic. *Journal of traumatic stress*, 24(1):93–96, 2011.

[19] A. Rizzo, G. Reger, G. Gahm, J.A. Difede, and B.O. Rothbaum. Virtual reality exposure therapy for combat-related ptsd. *Post-Traumatic Stress Disorder*, pages 375–399, 2009.

[20] Martin Siepmann, Jens Grossmann, Michael Mück-Weymann, and Wilhelm Kirch. Effects of sertraline on autonomic and cognitive functions in healthy volunteers. *Psychopharmacology*, 168(3):293–298, 2003.

[21] B.K. Wiederhold and MD Wiederhold. Virtual reality for posttraumatic stress disorder and stress inoculation training. *Journal of Cybertherapy & Rehabilitation*, 1(1):23–35, 2008.

[22] Brenda K Wiederhold and Mark D Wiederhold. A review of virtual reality as a psychotherapeutic tool. *CyberPsychology & Behavior*, 1(1):45–52, 1998.

[23] Dennis Patrick Wood, Jennifer Webb-Murphy, Robert N McLay, Brenda K Wiederhold, James L Spira, Scott Johnston, Robert L Koffman, Mark D Wiederhold, Jeff Pyne, et al. Reality graded exposure therapy with physiological monitoring for the treatment of combat related post traumatic stress disorder: a pilot study. *Studies in health technology and informatics*, 163:696, 2011.

[24] Georgios N Yannakakis and John Hallam. Ranking vs. preference: a comparative study of self-reporting. In *Affective Computing and Intelligent Interaction*, pages 437–446. Springer, 2011.

[25] Georgios N. Yannakakis and Hector P. Martinez. Ratings are overrated! *Frontiers on Human-Media Interaction*, 2015.

[26] Georgios N Yannakakis, Héctor P Martínez, and Arnav Jhala. Towards affective camera control in games. *User Modeling and User-Adapted Interaction*, 20(4):313–340, 2010.