Characterization of Fear Conditioning and Fear Extinction by Analysis of Electrodermal Activity

Rose T. Faghih,* *Member, IEEE*, Patrick A. Stokes, Marie-France Marin, Rachel G. Zsido, Sam Zorowitz, Blake L. Rosenbaum, Huijin Song, Mohammed R. Milad, Darin D. Dougherty, Emad N. Eskandar, Alik S. Widge, Emery N. Brown, *Fellow, IEEE*, Riccardo Barbieri, *Senior Member, IEEE*

***Abstract*— Electrodermal activity (EDA) is a measure of physical arousal, which is frequently measured during psychophysical tasks relevant for anxiety disorders. Recently, specific protocols and procedures have been devised in order to examine the neural mechanisms of fear conditioning and extinction. EDA reflects important responses associated with stimuli specifically administrated during these procedures. Although several previous studies have demonstrated the reproducibility of measures estimated from EDA, a mathematical framework associated with the stimulus-response experiments in question and, at the same time, including the underlying emotional state of the subject during fear conditioning and/or extinction experiments is not well studied. We here propose an ordinary differential equation model based on sudomotor nerve activity, and estimate the fear eliciting stimulus using a compressed sensing algorithm. Our results show that we are able to recover the underlying stimulus (visual cue or mild electrical shock). Moreover, relating the time-delay in the estimated stimulation to the visual cue during extinction period shows that fear level decreases as visual cues are presented without shock, suggesting that this feature might be used to estimate the fear state. These findings indicate that a mathematical model based on electrodermal responses might be critical in defining a low-dimensional representation of essential cognitive features in order to describe dynamic behavioral states.**

I. INTRODUCTION

Human fear conditioning models have brought significant new insights into the pathophysiology of psychiatric disorders [10], [9]. In particular, studies using a two-day fear conditioning and extinction procedure demonstrated that patients diagnosed with schizophrenia and posttraumatic stress disorder (PTSD) exhibit deficits in fear extinction memory recall [8], [14] and impaired extinction retention compared to trauma-exposed normal controls [13]. More recently, a study considering electrodermal activity (EDA) to measure conditioned response during fear conditioning and fear extinction responses on day 1, and during fear extinction recall and fear renewal on day 2, found that a skin conductance response (SCR) during conditioning and extinction recall are not significantly different across time and are correlated within subjects [18].

Changes in EDA have been proposed as neurophysiologic arousal measures and as estimates of sympathetic nervous system activity [16]. When an outgoing sympathetic nervous burst occurs, resulting from temporal and spatial summation of spikes triggered by sudormotor nerve, SCR is generated. Generally, an increase of the frequency or amplitudes in SCR signal is interpreted as an increase of sympathetic nervous system activity level [11]. The sudomotor nerve stimulating the sweat glands' activity triggering EDA is composed of separate, discrete and temporally short bursts. Importantly, SCRs recorded on skin surface are not always distinguishable due to overlapping responses. Several investigators have tried to decode SRC by overcoming the overlapping problem [12], [1], [2]. The main concept behind these approaches implies a deconvolution technique relying on the existence of a stereotyped and stable impulse response function (IRF) that can be estimated in order to reveal the underlying bursts generating the sudomotor signal.

In the present work we describe a novel approach based on a model where the input is a shock event generated within a behavioral paradigm. We use the sparsity of shocks (i.e., there are a small number of stimulations compared to the experiment time) and recover the timing and amplitude of individual stimulations using compressed sensing techniques. For compressible signals, where only a small number of coefficients are large (i.e., most coefficients are small or zero), the signal can be approximated by a sparse representation and recovered using optimization or greedy algorithms [3]. We then use a coordinate descent approach based on [5] to recover stimulus (visual cue or mild electric shock) and parameters.

II. METHODS

A. Experiment

To test our model and estimation algorithm, we used SCR data of 8 healthy subjects. A detailed description of the experiment and characteristics of the participants are in [18]. None of the participants had medical conditions, or neurological disorders, none were using psychoactive drugs

* Corresponding Author's Email: rfaghih@mit.edu

Rose T. Faghih, Patrick A. Stokes, Alik S. Widge, Emery N. Brown, and Riccardo Barbieri are with Massachusetts Institute of Technology and Massachusetts General Hospital.

Marie-France Marin, Rachel G. Zsido, Sam Zorowitz, Blake L. Rosenbaum, Huijin Song, Mohammed R. Milad, Darin D. Dougherty, and Emad N. Eskandar are with Massachusetts General Hospital.

PAS, SZ, DDD, ENE, ASW, ENB, RB were supported by the Defense Advanced Research Projects Agency (DARPA) under contract number W911NF-14-2-0045. RTF and ENB were supported by the 2013 Harvard/MIT Joint Research Grants Program in Basic Neuroscience. The opinions presented are those of the authors and not of DARPA or the institutions.

978-1-4244-9270-1/15/$31.00 ©2015 IEEE

or medications. A previously validated fear conditioning and extinction procedure was used, in which conditioning and extinction took place on Day 1 and extinction recall took place on Day 2. Skin conductance was recorded during these experiments to measure conditioned responses, and the conditioned stimulus was a visual cue while the unconditioned stimulus was a mild electric shock to the fingers (highly annoying but not painful). During the experiments, an unlit lamp was shown before being switched on to visual cue (i.e. one of three colors: blue, red, or yellow) as the conditioned stimuli. The visual cue was followed by shock only on selected trials of the Conditioning phase. In the Extinction phase, the color being extinguished was presented without being followed by a shock. On Day 2, the Extinction Recall phase was presented. The experimental procedures involving human subjects described in this paper were approved by the Partners Healthcare Human Research Committee and written informed consent was obtained from all participants.

B. Model

Based on the second order differential equation model by Alexander et al. [1], our model describes the changes in the skin conductance as a function of the activity of the sudomotor nerve:

$$\frac{dx_1(t)}{dt} = -\theta_1 x_1(t) + u(t) \tag{1}$$

$$\frac{dx_2(t)}{dt} = \theta_1 x_1(t) - \theta_2 x_2(t) \tag{2}$$

where x_2 is the skin conductance levels and x_1 is a hidden state variable stimulating the skin conductance levels. θ_1 and θ_2 are time constants in the model. $u(t) = \sum_{i=1}^{N} q_i \delta(t - \tau_i)$ is an abstraction of the stimulation where q_i represents the level of stimulation (color of light or shock) initiated at time τ_i (q_i is zero if stimulation did not occur at time τ_i). N corresponds to the length of the input based on the time resolution considered for modeling the input and length of the experiment. The novelty of our model compared to the model by Alexander et al. [1] is considering a finite number of stimuli (i.e. sparse stimuli compared to the SCR data). This type of characterization makes it possible to link timing and amplitude of the stimuli to subjective fear state. Let y_{t_k} be the observed skin conductance data:

$$y_{t_k} = x_2(t_k) + v_{t_k} \tag{3}$$

where v_{t_k} represents the measurement error. We assume a Gaussian density as a reasonably good approximation of the probability density of the measurement error, and by using a least squares approach in our estimation algorithm, we model the noise as a Gaussian random variable. Using the skin conductance data (x_2), we would like to estimate θ_1 and θ_2, and obtain the timing, and amplitude of the underlying stimuli. Our goal is to recover the subject's inherent response to stress by hypothesizing a state-space model for the emotional state. We assume that $x_1(0) = 0$, and solve for y_{t_k}:

$$y_{t_k} = a_{t_k} y_0 + b'_{t_k} \mathrm{u} + v_{t_k} \tag{4}$$

where $b_{t_k} = [\frac{\theta_1}{\theta_1-\theta_2}(e^{-\theta_2 k} - e^{-\theta_1 k}) \quad \frac{\theta_1}{\theta_1-\theta_2}(e^{-\theta_2(k-1)} - e^{-\theta_1(k-1)}) \quad \ldots \quad \frac{\theta_1}{\theta_1-\theta_2}(e^{-\theta_2} - e^{-\theta_1}) \quad \underbrace{0 \cdots 0}_{N-k}]'$, $a_{t_k} = e^{-\theta_2 k}$, and u represents the entire input over the entire experiment (elements of u take values q_i for $i = 1, ..., N$). Let $\mathrm{y} = \begin{bmatrix} y_{t_1} & y_{t_2} & \cdots & y_{t_M} \end{bmatrix}'$, $\theta = \begin{bmatrix} \theta_1 & \theta_2 \end{bmatrix}'$, $A_\theta = \begin{bmatrix} a_{t_1} & a_{t_2} & \cdots & a_{t_M} \end{bmatrix}'$, $B_\theta = \begin{bmatrix} b_{t_1} & b_{t_2} & \cdots & b_{t_M} \end{bmatrix}'$, and $v = \begin{bmatrix} v_{t_1} & v_{t_2} & \cdots & v_{t_M} \end{bmatrix}'$. We can represent this system as:

$$\mathrm{y} = A_\theta y_0 + B_\theta \mathrm{u} + v. \tag{5}$$

C. Estimation

To estimate the model parameters and the stimuli, we formulate an optimization problem as follows:

$$\min_{\substack{\mathrm{u} \geq 0 \\ C\theta \leq b}} \quad J_\lambda(\theta, \mathrm{u}) = \tfrac{1}{2} \|\mathrm{y} - A_\theta y_0 - B_\theta \mathrm{u}\|_2^2 + \lambda \|\mathrm{u}\|_p^p \tag{6}$$

where the ℓ_p-norm is an approximation to the ℓ_0-norm ($0 < p \leq 2$) and λ is chosen such that there is balance between fitting the data and capturing noise, and C and b are selected such that the problem is identifiable. Then, by using a coordinate descent approach, this optimization problem can be solved iteratively through the following steps (for $l = 0, 1, 2, ...$) until convergence is achieved:

1)
$$\mathrm{u}^{(l+1)} = \operatorname*{argmin}_{\mathrm{u} \geq 0} \quad J_\lambda(\theta^{(l)}, \mathrm{u}) \tag{7}$$

2)
$$\theta^{(l+1)} = \operatorname*{argmin}_{C\theta \leq b} \quad J_\lambda(\theta, \mathrm{u}^{(l+1)}) \tag{8}$$

The optimization problem in (7) can be solved using the Focal Underdetermined System Solver (FOCUSS) algorithm [7]. We use an extension of the FOCUSS algorithm called GCV-FOCUSS+ [5], [4]. The GCV-FOCUSS+ algorithm is based on FOCUSS+ [15], and uses the generalized cross-validation (GCV) technique [6] for estimating the regularization parameter. In particular, GCV-FOCUSS+ is closely related to a version of the FOCUSS algorithm by Zdunek et al. [17] which uses the GCV technique [6] for updating the regularization parameter for the FOCUSS algorithm through singular value decomposition. We first initialize the algorithm randomly, and find a good initialization for the unknowns while using FOCUSS+ for sparse recovery and interior point method for finding the model parameters. Then, we find a good initial condition by comparing the estimates obtained in the previous step, and select the θ and u values that minimize the cost function. Finally, we use a coordinate descent approach to estimate the unknowns until convergence is achieved.

III. Results and Discussion

Figure 1 shows that, for the conditioning phase, the proposed model and deconvolution algorithm can identify when the recovered stimuli are in agreement with the known timing of the mild electric shocks given to the subject. Of note, using the recovered stimuli, it is possible

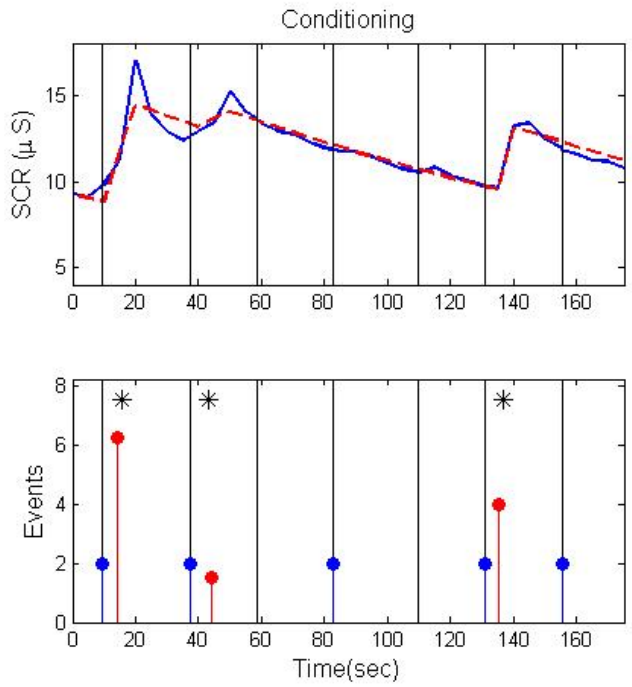

Fig. 1. **Estimated Deconvolution of the Experimental Skin Conductance Response During Conditioning in a Healthy Subject.** The top panel shows the measured SCR (blue solid line), and the estimated SCR levels (red dashed line). The bottom panel shows presentation of light color associated with shock (blue vertical lines with dots), shock (black asterisks), and estimated stimulation (red vertical lines with dots) using the SCR data.The solid black lines partition different trials by the timing of presentation of light (a visual cue that can corresponds to a shock or a non-shock color).

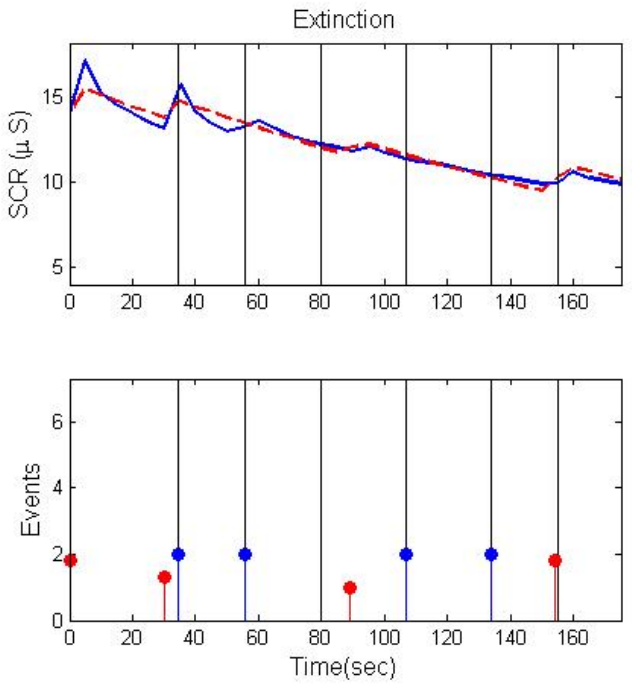

Fig. 2. **Estimated Deconvolution of the Experimental Skin Conductance Response During Extinction in a Healthy Subject.** The top panel shows the measured SCR (blue solid line), and the estimated SCR levels (red dashed line). The bottom panel shows presentation of light color associated with shock (blue vertical lines with dots), and estimated stimulation (red vertical lines with dots) using the SCR data. The solid black lines partition different trials by the timing of presentation of light (a visual cue that can corresponds to a shock or a non-shock color).

to reconstruct the SCR data as shown in Figure 1. Figure 2 shows the extinction phase where the recovered stimuli are in agreement with the impulse response behavior of the SCR data (observed peaks) except for missing one of the peaks in the SCR. Some of the responses are related to the trial and external factors while some of the recovered pulses are more correlated with the color associated with shock. The recovered stimuli in the extinction phase (Figure 2) are much lower than the conditioning phase (Figure 1). Figure 3 shows that the responses are more scattered in the recall phase than in the conditioning phase (Figure 1) and the extinction phase (Figure2) and weakly relate to stimulation with more scattered delays.

We further analyzed the response amplitudes and delay times of the estimated stimulation points. Each stimulation point was identified with the trial within which its estimated time occurred. Of those estimated points, only those occurring within a window of two seconds before and 14 seconds after light presentation were selected. Figure 4 shows the mean response amplitudes for each presentation of the different stimuli of the conditioning phases: shock color with shock (red), shock color without shock (blue), and non-shock color (green). The block structure of the conditioning phase is such that all presentations of one shock color, either extinguished or non-extinguished, occur before the presentations of the second shock color. As expected, response amplitudes are noticeably greater for shock stimuli than for no-shock stimuli. Of note, presentation of shock colors without receipt of shock are not greater than response to presentation of the non-shock color, and the former do not increase as more shocks are received. However, this comparison is limited by a small number of subjects, a small number of shock color without shock trials, and a small number of stimulation points identified for those trials.

Figure 5 shows the response amplitudes and delay times for each presentation of the extinguished color during the extinction phase. In the left plot, there appears, qualitatively, to be little change in the response amplitudes over the course of the extinction phase. No significant change in the response amplitudes over the course of the extinction phase can be observed (left plot). On the other hand, the delay times of the estimated responses (right plot) appear to increase over the course of the extinction phase, suggesting increase in response delay may be indicative of extinction.

In this paper, we presented preliminary results for modeling SCR data and identified characteristic features of reaction to the stimulation in form of shock or visual cue (by considering an impulse input with finite impulses); the reaction to visual cues depends on the state of the subject (i.e. phase of the experiment) and the reaction changes (see

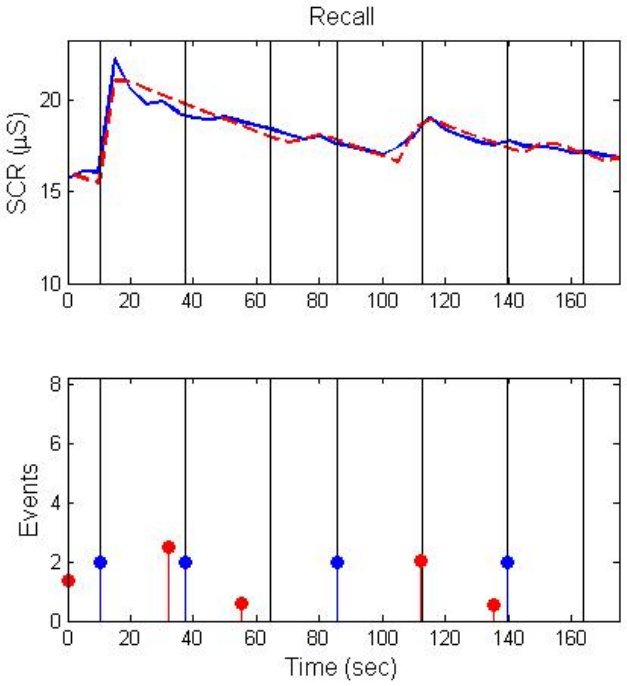

Fig. 3. **Estimated Deconvolution of the Experimental Skin Conductance Response During Recall in a Healthy Subject.** The top panel shows the measured SCR (blue solid line), and the estimated SCR levels (red dashed line). The bottom panel shows presentation of light color associated with shock (blue vertical lines with dots), and estimated stimulation (red vertical lines with dots) using the SCR data. The solid black lines partition different trials by the timing of presentation of light (a visual cue that can corresponds to a shock or a non-shock color).

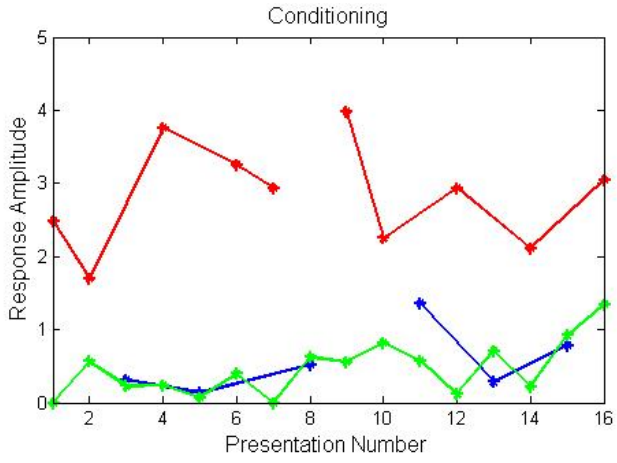

Fig. 4. **Comparison of Response Amplitudes for Shock Colors vs. Non-Shock Color During the Conditioning Phase Averaging Over 8 Healthy Subjects.** The estimated response amplitudes for the identified stimuli, averaged over subjects, for each presentation of shock color with shock (red), shock color without shock (blue), and non-shock color (green) during the conditioning phase of the experiment. The eight subjects chosen had comparable experimental sequences. The block structure of the experiment is such that one shock color is presented in the first half of the conditioning phase and the other shock color is present in the second half.

extinction versus recall). Considering that we recovered lower amplitudes for visual cue without electrical shock,

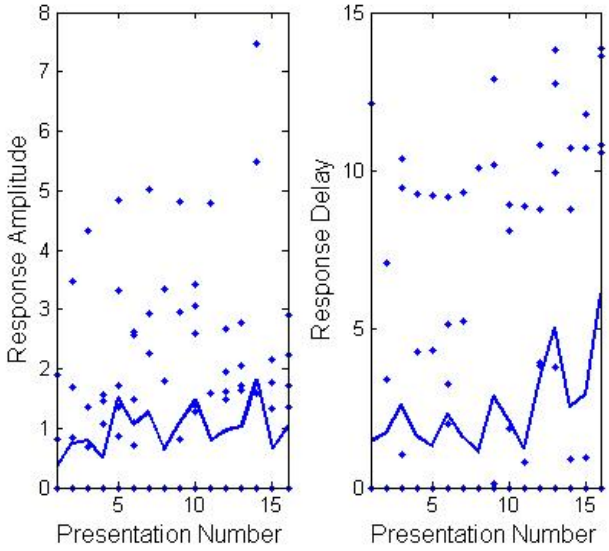

Fig. 5. **Estimated Response Amplitudes and Delay Times for Extinguished Color During the Extinction Phase.** The left panel shows the estimated SCR amplitudes for the identified stimuli corresponding to presentations of the extinguished color during the extinction phase of the experiment. The right panel shows the estimated SCR delay times ($t_{stim} - t_{light}$) for the identified stimuli corresponding to presentations of the extinguished color during the extinction phase of the experiment. The blue solid lines show the average response amplitude/delay over subjects for each presentation. The individual response amplitudes/delay for each subject for each presentation are shown as blue dots.

we can interpret the SCR to deduce when the subject is experiencing fear as a "virtual shock." Our model identifies impulses that represent the reaction of the subject to stimuli which can be potentially used in estimating the fear state of the subjects.

It is important to note that in SCR data, one observes small peaks that correspond to noise and by sparse recovery using GCV-FOCUSS+ which uses generalized cross-validation for finding the regularization parameter, we balance between fitting the data and capturing noise for different individuals and different experimental conditions.

IV. CONCLUSIONS AND FUTURE WORK

In this paper, we propose a novel sparse recovery approach for analyzing EDA. This approach allows for recovering the timings and amplitudes of the stimuli (visual cue, mild electrical shock, emotional response). This is a first step for building a state-space model for estimating the fear state in human subjects. The proposed approach for analyzing SCR data can be used to analyze different types of stress related disorders such as PTSD to distinguish between healthy subjects and different types of disease and also investigate whether treatment can be predicted by SCR data during conditioning, extinction, and recall. This can potentially be used as a predictor for changes in brain functions and improvement of clinical symptoms.

AUTHOR CONTRIBUTIONS

RTF and ENB developed the estimation algorithm. RTF, PAS, and RB analyzed the data and wrote the paper. MFM, RGZ, SZ, BLR, HS, and MRM collected or cleaned data. RB, ENB, ASW, DDD, and ENE initiated the project.

REFERENCES

[1] Alexander, D. M., et al., Separating individual skin conductance responses in a short interstimulus-interval paradigm, Journal of neuroscience methods 146(1):116-123, 2005.

[2] Benedek, M., Christian K., Decomposition of skin conductance data by means of nonnegative deconvolution, Psychophysiology 47(4): 647-658, 2010.

[3] Boufounos P., Duarte M., Baraniuk R., Sparse signal reconstruction from noisy compressive measurements using cross validation, In: IEEE Workshop on Statistical Signal Processing, 299303, 2007.

[4] Faghih R.T. , System Identification of Cortisol Secretion: Characterizing Pulsatile Dynamics. Ph.D. thesis, Massachusetts Institute of Technology 2014.

[5] Faghih R.T. et al., Deconvolution of serum cortisol levels by using compressed sensing, PLOS ONE 9(1): e85204, 2014.

[6] Golub G.H. et al., Generalized cross-validation as a method for choosing a good ridge parameter, Technometrics, 21(2), 215-223, 1979.

[7] Gorodnitsky I.F., Rao B.D., Sparse signal reconstruction from limited data using FOCUSS: a re-weighted minimum norm algorithm, IEEE Transactions on Signal Processing, 45(3):600-616, 1997.

[8] Holt Daphne J., et al., Extinction memory is impaired in schizophrenia, Biological psychiatry 65(6): 455-463, 2009.

[9] Jovanovic T., et al., Impaired fear inhibition is a biomarker of PTSD but not depression, Depression and anxiety 27(3): 244-251, 2010.

[10] Jovanovic T., et al., Posttraumatic stress disorder may be associated with impaired fear inhibition: relation to symptom severity, Psychiatry research 167(1): 151-160, 2009.

[11] Lidberg L., Wallin B.G., Sympathetic skin nerve discharges in relation to amplitude of skin resistance responses, Psychophysiology 18(3): 268-270, 1981.

[12] Lim C.L., et al., Decomposing skin conductance into tonic and phasic components, International Journal of Psychophysiology 25(2): 97-109, 1997.

[13] Milad M.R., et al., Neurobiological basis of failure to recall extinction memory in posttraumatic stress disorder, Biological psychiatry 66(12): 1075-1082, 2009.

[14] Milad M.R., et al., Presence and acquired origin of reduced recall for fear extinction in PTSD: results of a twin study, Journal of psychiatric research 42(7): 515-520, 2008.

[15] Murray J.F., Visual recognition, inference and coding using learned sparse overcomplete representations, Ph.D. thesis, University of California, San Diego, 2005.

[16] Wallin B.., Sympathetic nerve activity underlying electrodermal and cardiovascular reactions in man, Psychophysiology 18(4): 470-476, 1981.

[17] Zdunek R., Cichocki A., Improved M-FOCUSS algorithm with overlapping blocks for locally smooth sparse signals, IEEE Trans. Signal Process, 56(10):4752-4761, 2008.

[18] Zeidan M.A., et al. Testretest reliability during fear acquisition and fear extinction in humans, CNS neuroscience & therapeutics 18(4): 313-317, 2012.