

Considerations for designing response quantification procedures in non-traditional psychophysiological applications

Iyer, A.V.^{1,3}, Cosand, L.D.^{2,3}, Courtney, C.G.^{2,3}, Rizzo, A.A.³, and Parsons, T.D.³

¹Department of Biomedical Engineering, University of Southern California

²Department of Psychology, University of Southern California

³Institute for Creative Technologies, University of Southern California

Abstract. Psychophysiological assessment in the context of virtual environments is a promising means for benchmarking the efficacy and ecological validity of virtual reality scenarios. When applied to human-computer interaction, psychophysiological and affective computing approaches may increase facility for development of the next generation of human-computer systems. Such systems have the potential to use psychophysiological signals for user-feedback and adaptive responding. As the composition of investigating teams becomes diverse in keeping with interdisciplinary trends, there is a need to review de-facto standards of psychophysiological response quantification and arrive at consensus protocols adequately addressing the concerns of basic researchers and application developers. The current paper offers a demonstration of the ways in which such consensus scoring protocols may be derived. Electromyographic eye-blink scoring from an immersion investigation is used as an illustrative case study.

1 BACKGROUND

1.1 Psychophysiology in Human Computer Interaction research

Psychophysiology is increasingly incorporating virtual reality environments into human computer interface research [1]. The use of psychophysiological measures in studies of persons immersed in high-fidelity virtual environment scenarios offers the potential to develop current physiological computing approaches [2] into affective computing [3] scenarios. Such scenarios offer the potential for simulated environments to proffer cogent and calculated response approaches to real-time changes in user emotion, neurocognition, and motivation. The value in using virtual reality technology to produce simulations targeting emotional, neuropsychological, and motivational applications has been acknowledged by an encouraging body of research. Some of the work in this area has addressed affective processes: anxiety disorders, pain distraction and posttraumatic stress disorder [4]. Other work has assessed neurocognitive processes such as attention and executive functioning [5], [6]; memory [7], [8], [9]; and visuospatial abilities [10], [11], [12].

Thus far, the recording of psychophysiological variables while participants operate within virtual environments has produced useful results in studies examining immersion and presence [13], [14], [15]. As such, the VR assets that allow for precise stimulus delivery within ecologically enhanced scenarios appears well matched for this research. One area of increasing interest in psychophysiological assessment of persons immersed in virtual environments is startle eyeblink modification (SEM) [16], [17]. The modulatory influence of selective attention as well as generalized arousal processes on SEM, is well-established [18]. As such, the SEM is a likely source of benchmarking metrics for the efficacy and ecological validity of virtual environments in therapeutic and training applications. Unfortunately, the long tradition of use of the SEM paradigm in basic research in psychophysiology has resulted in a de facto standard of response quantification procedures [19]. This “received approach” may not be amenable to protocols that will be useful for eyeblink-actuated human computer interaction systems.

1.2 Introduction to current study

The eye-blink EMG signals used for analysis in this paper were obtained in a study which attempted the SEM paradigm in an investigation of immersion to detect differences in arousal if any, between persons placed in “low” and “high” immersion scenarios. Subjects were asked to passively view a virtual environment (VE) on two separate experimental runs consisting of both a highly immersive (HI) viewing condition and a low immersion (LI) viewing condition. In the HI condition, subjects wore a head mounted display (HMD) with full tracking capabilities and were free to explore their environment visually. The HI condition also made use of headphones and a tactile transducer floor to simulate riding in a large vehicle. The LI condition consisted of watching the same virtual Iraqi scenario on a 17 inch laptop screen while wearing headphones. During the LI condition, subjects viewed the VE from a static position.

The VE was comprised of a series of safe and combat zones in an Iraqi city. In both the HI and LI conditions, subjects viewed the VE from the perspective of the driver of a Humvee. The speed and trajectory of the vehicle was kept constant to control for time spent in each zone of the VE. Safe zones consisted mainly of a road surrounded by a desert landscape and were free of gunfire and other loud noises. The combat zones included improvised explosive devices (IEDs), gunfire, insurgents, and screaming voices. Subjects passed through 3 safe and 3 combat zones on each experimental run. The total length of each run was 210 seconds.

An acoustic startle probe was used to elicit startle eyeblink responses. The startle probe was a 110 dB white noise burst 50 ms in duration with a near instantaneous rise/fall time presented binaurally through Telephonics TDH-50P headphones. Decibel levels were measured with a Realistic sound level meter using a Quest Electronics earphone coupler. Startle probes were experienced intermittently throughout the experimental runs. A total of 4 startle probes were experienced in both the safe and combat zones in each run.

1.3 Overview of considerations for response quantification

According to the “scoring protocol” found within the “received approach” to psychophysiological experimentation, a set of signal parameters of interest and associated procedures to extract them from raw signal recordings must yield outputs that are readily amenable to further standardized analyses and expert interpretations. Due to the long tradition of use of the SEM paradigm in basic research in psychophysiology, the “received approach” represents a de facto standard of response quantification procedures [19]. The translation of psychophysiology procedures from “assessment” of persons’ responses to computer mediated information to “adaptive human computer interfaces” presents the requirement of devising scoring protocols for situations to which the existing de facto standards (i.e. received approach) may not be very suitable. It is reasonable to expect that the scoring protocols that will be useful for eyeblink-actuated systems (e.g. BCIs), may be dissimilar to the parameters currently of interest in psychophysiological assessment.

In this paper, some general considerations for designing response quantification procedures in non-traditional psychophysiological applications are presented, using eyeblink scoring for a Humvee Immersion Experiment as a case study.

1.4 Open parameters in the design of scoring protocols for electromyographic eyeblink recordings

Signal acquisition consoles currently popular for electromyographic (i.e. EMG) signal acquisition [20] typically incorporate pre-processing procedures like amplification. Further signal conditioning stages like filtering and thresholding, which characterize a response quantification procedure, were earlier performed using analog hardware implementations which are now largely being replaced by digital software implementations. For the study reported herein, filtering and thresholding were implemented using MATLAB®. The following are potential open parameters from which the investigator may choose while devising a scoring protocol for raw EMG recordings:

1.4.1 Filter parameters:

Following the “received approach”, the smoothening (low-pass filtering) requirements are determined largely by the quality of the measuring equipment and likely sources of noise. While smoothing is a necessity for noisy signals, it has the undesirable side-effect of causing phase distortions and in general, a trade-off is sought to give a high signal-to-noise ratio with minimal phase distortion. Throughout this paper, “box-car” filtering (i.e. running average filter) is implemented with the filter width (i.e. window size) as the open parameter. Box-car filtering was chosen herein for illustrative purposes due to its ease of implementation, computational speed, and near-ubiquitous use in EMG signal conditioning.

1.4.2 Threshold:

Following the “received approach”, portions of the filtered and rectified eyeblink trace within a small time-window adjacent to a startle-eliciting stimulus are treated as startle blinks if their magnitude exceeds a stipulated threshold. Threshold choice is essentially dictated by typical eyeblink EMG amplitudes observed in the population under study, and must take into account the amplitude reduction caused by the filter settings in the smoothing stage.

1.4.3 Region of interest following the startle blink:

Following the “received approach”, the length of the time window following a startle-eliciting stimulus within which a super-threshold signal will be considered a blink associated with that stimulus. This is decided mainly on psychophysiological considerations like the average latency of the population under study. Herein the length of this time window is fixed at 150 ms which is psychophysiological established as a typical time-to-peak for startle-elicited eyeblink signals.

In what follows, results obtained with divergent filter parameter and threshold settings are compared with those obtained from a “reference setting” conforming to the “received approach” to in psychophysiological assessment. This is performed to gain an increased understanding of the effect of these parameter changes and ultimately guide design of augmented scoring protocols for future applications.

1.5 Immediate motivations for the present study

In preliminary analyses, two different scoring methods were considered; 1) Received Approach: the output of this initial protocol being the superthreshold EMG of the smoothed signal in a 150 ms window adjacent to the startle probe; with the threshold at 0.0001 mV; and 2) Elementary Signal-Processing Approach (distortion-causing signal-smoothing is preferably avoided): the score for this alternative protocol being the mean amplitude of superthreshold EMG spikes of the unsmoothed signal in a 150 ms window adjacent to the startle probe, with the threshold at 0.15 mV. An analysis of variance between the EMG scores for “high” and “low” immersion did not report significant differences ($F = 0.091$, $p = 0.763$) for the scoring procedure comporting to the “received approach”. Contrariwise, the Elementary Signal-Processing Approach scoring procedure resulted in significant differences ($F = 17.412$, $p < 0.001$). This obvious mismatch between the results of two scoring protocols motivated the investigations presented in this paper to determine feasible ranges of parameter settings for eye-blink scoring protocols. The preliminary findings underscore the need for caution in the choice of scoring metrics and proper validation of the same. These considerations may be of importance to researchers interested in eyeblink analyses to be reliably employed in non-traditional applications psychophysiological data.

2 METHODS

Eyeblink EMG traces from the physiological recordings obtained from the orbicularis oculi muscle of subjects in the Humvee Immersion Experiment using Biopac MP150 were exported to MATLAB® and scored using multiple protocols to enable comparison. Each EMG trace was of duration 210 seconds, during which subjects experiencing the Humvee scenario were administered 9 acoustic startle probes via headphones at pre-determined times. The traces were smoothed using a box-car (running average) filter and rectified. Figure (2 a) shows part of a typical EMG trace filtered with running average filters of different filter-widths and rectified. Trace segments of 150 milliseconds adjacent to each startle probe were then examined for spikes whose magnitude exceeded a pre-determined threshold. These super-threshold spikes were treated as startle blinks and a time series of startle-blinks was obtained from each EMG trace. Figure (2b) shows startle-blinks obtained using this procedure, for different parameter settings. Each parameter setting corresponds to a scoring protocol.

The reference protocol adopted to represent the “received approach”, used a box-car filter window size of 100 samples (corresponding to 100 ms for a sample rate of 1 kHz) and a threshold of 0.1 microvolt, reasonably conforming to current psychophysiological practice. Alternative protocols were generated by varying the filter window-size from 10 to 200 samples and the threshold from 0.01 microvolt to 5 microvolt. The covariance of the output signals from each of these alternative protocols, with the corresponding output of the reference protocol is treated as an index to study the effect of parameter variations. The covariance between the startle trains obtained from the reference protocol and a test protocol, is of interest because it is a measure of how much the two signals ‘vary together’ i.e. how much a signal tends to be above its expected value if the other signal is known to be above its own expected value. A high positive covariance between the output signal of a particular test protocol and that of the reference protocol would suggest that there is a high degree of linear dependence between the two signals. An implication of such a high degree of linear independence is that the investigator can treat such a test protocol as practically equivalent to the reference protocol.

The results presented in Section 3 represent the mode of the observations for different subjects. The conclusions of this analysis are expected to guide the design of a scoring protocol suited to the ‘most typically encountered’ subject. Therefore the measure chosen for this analysis was the ‘mode’ which represents the most frequently occurring value in the sample.

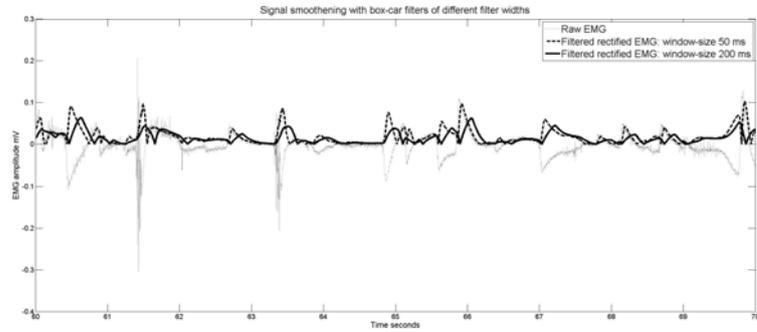


Fig. 2a. Rectification and smoothing of the raw EMG signal (dotted trace) using a running-average filter with window sizes 50 ms (dashed trace) and 200 ms (bold trace).

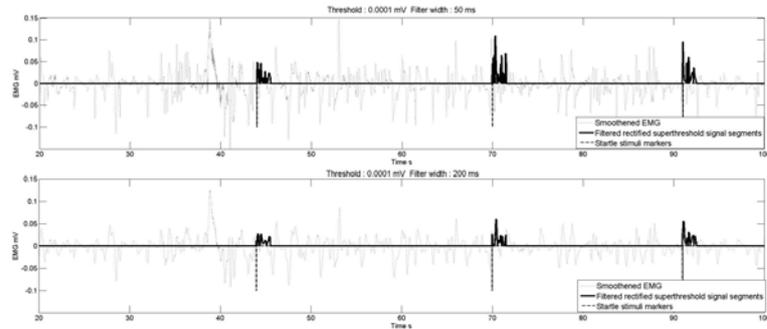


Fig. 2b. Startle blink time series (bold traces) obtained with two different settings of filter window-size and threshold, from the same raw eyeblink signal. The traces shown above are a portion of a 210-second-long electromyographic recording in the Humvee Immersion Experiment with acoustic startle probes used to elicit the blink responses. The positions of these startle probes are shown by the dashed markers.

3 RESULTS

Figure 3 summarizes the effect of two open parameters of interest, namely, boxcar filter width and EMG threshold, on the covariance matching score of test protocol outputs with the reference protocol. The electromyographic recordings used for this analysis were obtained from 7 subject volunteers of the Humvee Immersion Experiment in the high immersion condition. Each curve in the family of curves in the upper panel, shows the effect of filter width on the covariance matching score, for a fixed value of threshold (Upper curves are for lower thresholds.). Likewise, each curve in the lower panel shows the effect of EMG threshold on the covariance matching score, parametric on the filter width. Taken together, these results represent a scan of the space of open parameters to determine feasible settings. A visual

examination of the curves reveals that there is ‘plateau’ of very slow decline around the optimal values of maximum covariance outside which the covariance score shows a more rapid decline. The range of values in the ‘plateau’ regions represent parameter settings which provide outputs nearly equivalent to those of the reference protocol. This suggests that traditional parameter settings are not tight and strict, but belong to a larger range of values of open parameters which a designer can explore, while still conforming reasonably to established psychophysiological practice.

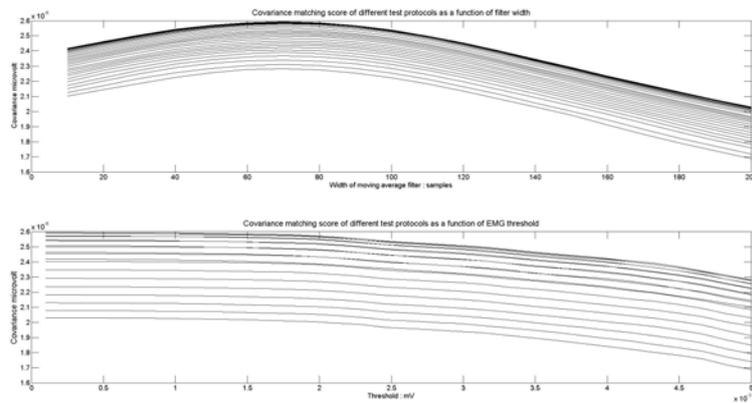


Fig. 3. Effect of filter width (upper panel) and effect of EMG threshold (lower panel) on the covariance matching score of a test scoring protocol with respect to the reference protocol. View text in Section 3 for details and explanation.

4 DISCUSSION

4.1 Consensus between requirements of basic researchers and application developers

The “received approach” to scoring protocols parameter settings need not always be viewed as a tight requirement. Instead, as the examples in this paper have illustrated, they may be viewed as a flexible constraint from which a designer may make judicious use of adjustable parameters based on other application specific demands. For example, in an experimental scenario where signal-to-noise ratio is a high priority, the designer can choose the highest filter width within the “plateau” region, thus representing a consensus of the requirements of traditional psychophysiology as well as engineering considerations. Although the method of analysis presented herein is merely suggestive, it can be readily adapted to other scenarios. Further, its variants may be employed to find feasible ranges of parameter settings. As research with psychophysiological modalities becomes more and more interdisciplinary, and the composition of investigating teams becomes more and more diverse, such analyses

will be a useful way to arrive at consensus protocols that address the concerns of career psychophysiology, signal-processing engineers, and designers of virtual environments.

A case in point, of the competing concerns of basic researchers in psychophysiology and signal processing engineers is the need for signal smoothing. Traditional response quantification procedures that are well-established in the psychophysiology research community typically involve signal smoothing of the EMG signal as a pre-processing step, and extraction of peak amplitude and response latency from the smoothed signal. Researchers interested in human computer interfaces (i.e. BCIs), however, may prefer to average raw unsmoothed signals across trials, to avoid the signal distortion and information loss caused by smoothing. A method such as that described above can be useful to arrive at a mutually acceptable parameter setting.

4.2 Emerging trends and relevance of interdisciplinary dialogue on response quantification standards

In psychophysiological assessment, data scoring is typically performed with ‘off-the-shelf’ scoring software [21], most often proprietary and provided by the manufacturers of the recording equipment. In recent years, the recognition of several new parameters of interest in EMG analysis [22] motivates the need for newer sophisticated data scoring procedures capable of dealing with a greater variety of metrics. Investigators, especially in nascent applications, will need to choose parameters tailored to the application, exploiting the additional capabilities of state-of-the-art data acquisition equipment and using flexible customizable data-scoring software, typically written in MATLAB ®. These developments will lead to a shift from an “off-the-shelf” approach to a “drawing-board” approach in designing response quantification procedures and scoring protocol design. Attempts to contribute (such as the one found in this paper) to the ensuing interdisciplinary discourse to arrive at universally accepted standards for psychophysiological response quantification, deserve due attention so that the full potential of the confluence between psychophysiology and virtual environments is attained in a genuine, well-founded manner.

5 References

1. Ferreira, A., Celeste, W.C., Cheein, F.A., Bastos-Filho, T.F., Carelli, R.: Human-Machine Interfaces Based on EMG and EEG Applied to Robotic Systems. *J of NeuroEngineering and Rehabilitation*, 5-10 (2008)
2. Allanson, J. and S.H. Fairclough, A research agenda for physiological computing. *Interacting With Computers*. 16, 857-878 (2004)
3. Picard, R.W., *Affective Computing*. Cambridge, Mass.: MIT Press. (1997)
4. Parsons, T.D., Rizzo, A.A.: Affective Outcomes of Virtual Reality Exposure Therapy for Anxiety and Specific Phobias: A Meta-Analysis. *Journal of Behavior Therapy and Experimental Psychiatry*, 39, 250-261 (2008)

5. Parsons, T.D., Bowerly, T., Buckwalter, J.G., Rizzo, A.A.: A controlled clinical comparison of attention performance in children with ADHD in a virtual reality classroom compared to standard neuropsychological methods. *Child Neuropsychology*, 13, 363-381 (2007)
6. Parsons, T.D., Rizzo, A.A.: Neuropsychological Assessment of Attentional Processing using Virtual Reality. *Annual Review of CyberTherapy and Telemedicine*, 6, 23-28 (2008)
7. Parsons, T.D., Rizzo, A.A., Bamattre, J., Brennan, J.: Virtual Reality Cognitive Performance Assessment Test. *Annual Review of CyberTherapy and Telemedicine*, 5, 163-171 (2007)
8. Parsons, T.D., and Rizzo, A.A.: Initial Validation of a Virtual Environment for Assessment of Memory Functioning: Virtual Reality Cognitive Performance Assessment Test. *Cyberpsychology and Behavior*, 11, 17-25 (2008)
9. Parsons, T.D., Silva, T.M., Pair, J., Rizzo, A.A.: A Virtual Environment for Assessment of Neurocognitive Functioning: Virtual Reality Cognitive Performance Assessment Test. *Studies in Health Technology and Informatics*, 132, 351-356 (2008)
10. Parsons, T.D., Larson, P. Kratz, K., Thiebaut, M., Bluestein, B., Buckwalter, J.G., Rizzo, A.A.: Sex differences in mental rotation and spatial rotation in a virtual environment. *Neuropsychologia* 42:555–562 (2004)
11. Parsons, T.D., Rizzo, A.A., Buckwalter, J.G.: Backpropagation and regression: comparative utility for neuropsychologists. *Journal of Clinical and Experimental Neuropsychology* 26:95–104 (2004)
12. Parsons, T.D., Rizzo, A.A., van der Zaag, C., McGee, J.S., Buckwalter, J.G.: Gender and cognitive performance: a test of the common cause hypothesis. *Aging, Neuropsychology, and Cognition* 12:78–88 (2005)
13. Macedonio, M., Parsons, T.D., Rizzo, A.A.: Immersiveness and Physiological Arousal within Panoramic Video-based Virtual Reality. *Cyberpsychology and Behavior*, 10, 508-516 (2007)
14. Parsons, T.D., Iyer, A., Cosand, L., Courtney, C., & Rizzo, A.A. Neurocognitive and Psychophysiological Analysis of Human Performance within Virtual Reality Environments. *Studies in Health Technology and Informatics*, 142, 247-252 (2009)
15. Pugnetti, L., Meehan, M., & Mendozzi, L.: Psychophysiological correlates of virtual reality: a review. *Presence*, 10, 384–400 (2001)
16. Grillon, C., Baas, J.M., Cornwell, B., Johnson, L.: Context Conditioning and Behavioral Avoidance in a Virtual Reality Environment: Effect of Predictability. *Biol. Psychiatry* 60, 752--759 (2006)
17. Baas, J. M., Nugent, M., Lissek, S., Pine, D. S., Grillon, C.: Fear Conditioning in Virtual Reality Contexts: A New Tool for the Study of Anxiety. *Biol. Psychiatry*, 55, 1056--1060 (2004)
18. Harbin, T.J., Berg, W.K., 1986.: The Effects of Age and Attention Upon Reflex Inhibition. *Biol. Psychol.* 22, 81--94 (1986)
19. Blumenthal, T. D., Cuthbert, B. N., Filion, D. L., Hackley, S., Lipp, O. V., & van Boxtel, A. Committee report: Guidelines for human startle eyeblink electromyographic studies. *Psychophysiology*, 42, 1–15 (2005)
20. Carmel, S., Macy, A.J.: Physiological Signal Processing for Biomedical Engineering Education, BIOPAC Systems, Inc., 42 Aero Camino, Goleta, CA, USA (www.biopac.com)
21. Balaban, M.T., Losito, B., Simons, R.F., Graham, F.K.: Offline Latency and Amplitude Scoring of the Human Reflex Eyeblink with FORTRAN IV. *Psychophysiology*, 23, 612 (1986)
22. Blumenthal, T.D., Melkonian D.S.: A Model Based Approach to Quantitative Analysis of Eyeblink EMG Response Components. *J of Psychophysiology*, 17, 1—11 (2003)