International Journal of Psychophysiology 91 (2014) 186–193

Contents lists available at ScienceDirect

International Journal of Psychophysiology

journal homepage: www.elsevier.com/locate/ijpsycho

Development and validation of an unsupervised scoring system (Autonomate) for skin conductance response analysis

Steven R. Green [a,1], Philip A. Kragel [b,c,1], Matthew E. Fecteau [c], Kevin S. LaBar [b,c,*]

[a] Department of Psychology, Indiana University, Bloomington, IN, United States
[b] Department of Psychology and Neuroscience, Duke University, Durham, NC, United States
[c] Center for Cognitive Neuroscience, Duke University, Durham, NC, United States

ARTICLE INFO

Article history:
Received 9 August 2013
Received in revised form 22 October 2013
Accepted 24 October 2013
Available online 31 October 2013

Keywords:
Autonomic nervous system
Electrodermal activity
Psychophysiology
Sympathetic arousal
Skin conductance response
Computer-assisted scoring

ABSTRACT

The skin conductance response (SCR) is increasingly being used as a measure of sympathetic activation concurrent with neuroscience measurements. We present a method of automated analysis of SCR data in the contexts of event-related cognitive tasks and nonspecific responding to complex stimuli. The primary goal of the method is to accurately measure the classical trough-to-peak amplitude of SCR in a fashion closely matching manual scoring. To validate the effectiveness of the method in event-related paradigms, three archived datasets were analyzed by two manual raters, the fully-automated method (Autonomate), and three alternative software packages. Further, the ability of the method to score non-specific responses to complex stimuli was validated against manual scoring. Results indicate high concordance between fully-automated and computer-assisted manual scoring methods. Given that manual scoring is error prone, subject to bias, and time consuming, the automated method may increase the efficiency and accuracy of SCR data analysis.

1. Introduction

Measures of electrodermal response from the human periphery, such as the skin conductance response (SCR), provide insight into activation levels of the sympathetic branch of the autonomic nervous system (Boucsein, 1992). SCRs are thought to reflect increases in the electrical conductivity of the skin caused by the release of sweat from eccrine sweat glands located on the palmar surface of the hand and foot. Because these glands are innervated by sympathetic sudomotor nerves, they provide a window into the activity of multiple brain structures, such as limbic regions, basal ganglia, and frontal cortex, that regulate the autonomic nervous system (Edelberg, 1972).

Due to the specificity of the measure, ease of setup, participant tolerance, and relatively low cost, SCRs have gained increasing popularity in clinical, neuroscientific, and psychological studies of emotion and decision-making, learning and conditioning, orienting and attention, and deception. While there are multiple parameters associated with a SCR, such as response latency, rise time, and half-recovery time, the most commonly used parameter is the amplitude of the SCR relative to a post-stimulus baseline. Traditional scoring of SCR amplitude consisted of manually measuring the distance from trough to peak of responses that fit a well-defined set of criteria pertaining to the amplitude, latency, and duration of the response (Barry, 1990; Levinson and Edelberg, 1985).

Manual scoring of skin conductance data has multiple benefits, making it a historically popular choice of data analysis. Primarily, close inspection of trial-by-trial data traces ensures that individual responses are physiological signals related to an event of interest. Manual scorers need only examine data close in time to an event in order to score a SCR. Unfortunately, manual scoring has several drawbacks, even with computer-assisted graphical interfaces. Chief among them is the amount of time needed to perform the analysis. Since traditional scoring requires a trained rater to inspect each event for a response, studies with many events are highly time-consuming. Another drawback is human bias wherein a rater may inadvertently vary the stringency of the criteria for including a response. Finally, manual analysis has been known to suffer from the *scale invariance problem* in which the detection of an inflection point depends on what scale the rater uses to inspect the data. For instance, viewing the electrodermal trace at low magnifications or poor viewing angles may lead to the misidentification of subtle changes in electrodermal data.

In an attempt to overcome some of the problems associated with manual scoring, computer-based algorithms have been previously implemented to detect SCRs (Trosiener and Kayser, 1993), although not in an event-related fashion, as response latency and duration are not utilized in detection analyses. Generally, these methods identify points in the skin conductance time-series with a slope of zero. If the change in skin conductance within this range is large enough, it is identified

* Corresponding author at: Duke University, Center for Cognitive Neuroscience, B203 LSRC Building, Research Dr. Box 90999, Durham, NC 27708-0999, United States. Tel.: +1 919 681 0664; fax: +1 919 681 0815.
E-mail address: klabar@duke.edu (K.S. LaBar).
[1] These authors contributed equally to the manuscript.

0167-8760/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.ijpsycho.2013.10.015

as a SCR. While these methods can accurately extract increasing portions of a time series of skin conductance data, they do not filter out responses that are not plausibly event-related from a physiological perspective (that is, time-locked to the onset of a particular stimulus of interest). Other computer-based algorithms for peak detection have been implemented and compared to manual scoring, with favorable results for experimental designs with long inter-stimulus intervals (ISIs) that can accommodate temporal separation of individual SCR profiles from successive stimuli (Storm et al., 2000). While suitable when SCRs are distant in time and do not overlap, peak detection approaches based solely on the slope of the electrodermal trace are limited in their ability to isolate overlapping responses. If two SCRs occur within a short period of time, the skin conductance trace may not peak (have a slope of 0) before rising again.

Due to the increase in popularity of rapid, event-related experimental designs with shorter ISIs, additional methods have been developed to deal with the issue of overlapping SCRs. One graphical manual approach involves extending the baseline drift at stimulus onset to the time of a skin conductance peak, essentially linearly detrending the baseline drift (Barry et al., 1993). Approaches utilizing deconvolution (Alexander et al., 2005; Benedek and Kaernbach, 2010b; Lim et al., 1997) can be used to decompose skin conductance data into tonic and phasic activity, reducing the impact of overlapping responses. The goal of these methods is to more accurately measure SCRs by generating an estimate of phasic activity with a constant level of baseline activity. Alternatively, a general linear convolution model can be used to isolate event-related skin conductance activity (Bach et al., 2009). In solving a general linear model, this method generates parameter estimates that reflect the amplitude of task-related skin conductance activity. For researchers interested in experimental designs with short ISIs, these methods may be preferential for analyzing SCR data.

While methods estimating the SCR using mathematical models are attractive from a theoretical and procedural standpoint, one main issue complicates their use when compared to manual scoring: non-specific or spontaneous fluctuations. Changes in skin conductance that occur in the absence of stimuli can introduce error into models of electrodermal time-series. Spontaneous fluctuations have been successfully incorporated into generative models of skin conductance activity (Bach et al., 2010), although it remains unclear under what conditions assumptions about the occurrence and duration of these activations are valid. If assumptions concerning when spontaneous fluctuations are likely to occur are incorrect, the estimation of event related responses could be negatively impacted. We posit that, in the context of event-related analysis, focusing on data that is close in time to an event (i.e. the rise of the SCR) and is not dependent on characterizing spontaneous fluctuations will perform more consistently across a variety of experimental settings.

Here we present a traditional method of SCR data analysis in the context of event-related cognitive tasks that is fully-automated and does not depend on fitting data to a modeled response profile. The goal of our method is to automate manual scoring of stimulus-locked SCR amplitudes, while systematically dealing with overlapping SCRs and other common problems that introduce biases in manual scoring, such as consistency in applying response criteria. By design the software (called 'Autonomate') will apply the same criteria to each event to determine if a response occurred, thus avoiding the problem of manual raters inadvertently shifting their stringency of criteria as the data are analyzed. Furthermore, variation in the scale used to inspect the data by a manual rater (e.g. scoring under different magnifications) is not an issue for the software, as it is scale-invariant. To validate the new method in event-related paradigms, three archived datasets previously scored by two manual raters are analyzed using four software packages (Autonomate, and three methods which aim to address the issue of overlapping responses — AcqKnowledge, Ledalab, and SCRalyze), and the results are compared using standard metrics. To generalize the use of our method beyond event-related designs, we additionally validated Autonomate against manual scoring of non-specific SCRs in a fourth dataset of electrodermal responses to cinematic films. Complex datasets of this nature provide a challenging test of the software's utility as they contain more frequent and highly variable SCRs compared to event-related designs. By validating the software in a variety of experimental paradigms, we can more precisely determine under what conditions it is a suitable alternative to manual scoring.

2. Materials and methods

2.1. Automated method

Prior to analysis, data recorded at a sampling rate of 200 Hz were preprocessed using a 25 Hz finite impulse response low-pass filter and smoothed using a 3-sample moving average function. SCRs – one dimensional vectors of digitized data here denoted as **S** – were segmented into windows of **L** seconds following each stimulus. These data were down-sampled to 8 Hz using a Chebyshev Type I filter in order to reduce the effect of high frequency noise on subsequent analysis. The rises of candidate SCRs were found by searching for sections of the first order temporal derivative of the skin conductance data, **S′**, that are above the threshold of **U** μS per second for a minimum duration of **w** seconds (Fig. 1A).[2] The start and end of candidate SCRs were determined by the zero crossings of **S′**, and the amplitude was recorded as the difference in **S** between the second and first crossings.

Candidate SCRs were classified as being isolated or affected by neighboring responses by searching for patterns of inflection points (zero crossings in the second order temporal derivative). Inflection points were categorized based on whether the slope goes from increasing to decreasing (type **A**) or decreasing to increasing (type **B**) around the point. If there were three sequential inflection points within a SCR with a pattern **A**–**B**–**A**, then the center point **B** was used to split the SCR into multiple candidate responses (Fig. 1B).

Each candidate SCR must meet a specific set of criteria used in hand scoring in order for the SCR amplitude to be regarded as time-locked to the stimuli and recorded. Consistent with our prior reports (e.g., Dunsmoor et al., 2009; Huff et al., 2009; Thomas and LaBar, 2008), the following response criteria were established as default: the latency between the eliciting stimulus onset and the rise of the response, or SCR latency, must be between 1 and 4 s; the time between response start and peak, or SCR duration, must occur between 0.5 and 5 s; and the response amplitude must be greater than 0.02 μS. While these default values are recommended, these criteria are free parameters in the Autonomate software (as are the variables **U**, **L**, and **w**) and can be adapted by the user if further optimization is required. In the case that multiple SCRs meet all the criteria, the largest response within the window was recorded. For the analysis of spontaneous responses, the criterion of response latency was relaxed and all responses within a specified window were recorded. Once the final response was selected, the response amplitude was computed by finding the difference between local maxima and minima in the preprocessed (not down-sampled) data.

2.2. Validation of Autonomate software

To ensure the performance of our automated method closely matched that of the manual scoring on which it is based, subsamples of randomly selected subjects from three archived event-related studies were analyzed both manually (by two expert raters) and with the automated method. Studies using a range of stimuli and tasks were chosen to generate SCRs with variable amplitude, latency, and degree of overlap. For the event-related studies, electrodermal activity was recorded from the nondominant hand, using Ag–AgCl electrodes attached to the

[2] The term 'rise' is used here and throughout as a period of increase over baseline; not to be confused with the more specific term 'rise time,' which refers to the length of time from onset to peak for a SCR.