

## Automated Artifact-Detection Procedure for Vaginal Photoplethysmography

Carey S. Pulverman, Cindy M. Meston & J. Gregory Hixon

To cite this article: Carey S. Pulverman, Cindy M. Meston & J. Gregory Hixon (2018): Automated Artifact-Detection Procedure for Vaginal Photoplethysmography, Journal of Sex & Marital Therapy, DOI: [10.1080/0092623X.2018.1436627](https://doi.org/10.1080/0092623X.2018.1436627)

To link to this article: <https://doi.org/10.1080/0092623X.2018.1436627>



Accepted author version posted online: 05 Feb 2018.  
Published online: 06 Mar 2018.



[Submit your article to this journal](#) 



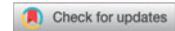
Article views: 9



[View related articles](#) 



[View Crossmark data](#) 



## Automated Artifact-Detection Procedure for Vaginal Photoplethysmography

Carey S. Pulverman, Cindy M. Meston, and J. Gregory Hixon

Department of Psychology, The University of Texas at Austin, Austin, TX, USA

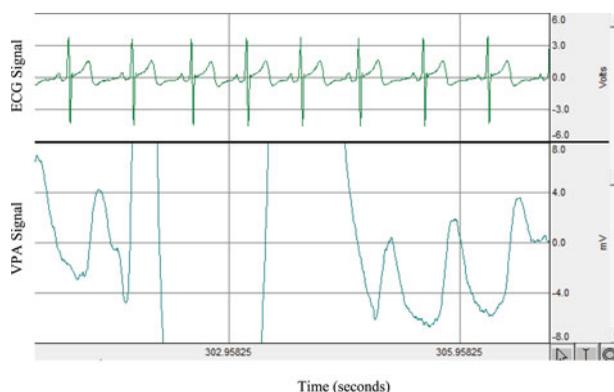
### ABSTRACT

Vaginal photoplethysmography is the most commonly used method of assessing women's genital sexual arousal. Raw photoplethysmograph data consist of a series of peaks and troughs, and movement by the participant results in artifacts indicated by unusually high or low peak-to-trough amplitudes. The gold-standard approach to artifact detection involves visual inspection by a trained experimenter and manual removal of artifacts from the data set, however, this process is time-consuming and subject to human error. We present an automated data-processing procedure that uses a series of smoothing regression splines to model the data and identify outliers. The automated procedure was applied to a set of neutral data and sexual-arousal response data, and artifacts identified were compared to artifacts identified by the standard approach of visual inspection. The automated method showed acceptable accuracy in terms of sensitivity and specificity comparable to the manual-processing method. The automated procedure could reduce human error and data-processing time for studies using vaginal photoplethysmography.

Vaginal photoplethysmography provides a physiological assessment of women's genital sexual-arousal response, and is thought to measure the vasoengorgement of blood in the vaginal walls. Although a number of other measures for assessing female genital arousal exist such as genital thermography (Kukkonen, Binik, Amsel, & Carrier, 2010) and clitoral ultrasonography (Berman et al., 1999; Kukkonen et al., 2006), vaginal photoplethysmography is the most commonly used measure of women's genital sexual arousal. The photoplethysmograph was developed by Sintchak and Geer in 1975. Photoplethysmograph data are typically analyzed as two filtered signals: the direct current and the alternating current. The direct current (DC) signal is believed to measure slow changes in vaginal blood flow, termed vaginal blood volume (VBV). The alternating current (AC) signal reflects phasic changes in blood flow that occur with each heartbeat, called vaginal pulse amplitude (VPA). Of the two components, VPA has been identified as the more sensitive measure of female genital sexual arousal (Laan & Everaerd, 1995). VPA has also been identified as a sexual arousal-specific response to the presentation of sexual stimuli; in other words, VPA only increases in response to the presentation of sexual stimuli and does not change in response to the presentation of other types of affective stimuli such as anxiety-provoking stimuli (Laan, Everaerd, & Evers, 1995; Suschinsky, Lalumière, & Chivers, 2009). Vaginal photoplethysmography provides continuous readings of VPA, and therefore a rich set of data; however, one drawback is its sensitivity to movement by the participant.

**CONTACT** Cindy M. Meston  [meston@psy.utexas.edu](mailto:meston@psy.utexas.edu)  Department of Psychology, The University of Texas at Austin, 108 East Dean Keaton Street, Stop A8000, Austin, TX 78712, USA.

Color versions of one or more of the figures in the article can be found online at [www.tandfonline.com/usmt](http://www.tandfonline.com/usmt)



**Figure 1.** Example of VPA artifact in raw electrocardiogram and vaginal photoplethysmograph data. ECG = electrocardiography; VPA = vaginal pulse amplitude.

Like other types of physiological assessment such as electrocardiography (ECG) and electroencephalography (EEG), vaginal photoplethysmography data collection is interrupted when the participant physically moves during the experiment. In addition to gross bodily movement, contraction of the pelvic muscles can also create movement artifacts in vaginal photoplethysmography (Laan et al., 1995). VPA is measured by the amplitude from the peak to the trough (or the trough to the peak) of each waveform. When artifacts occur, the normal peak-to-trough pattern is interrupted by abnormally high or low amplitudes (see Figure 1 for an example of artifacts). Movement artifacts should *not* be retained as these high and low VPA readings have the potential to skew subsequent data analyses. Photoplethysmograph data requires processing for detection and removal of these artifacts.

The gold-standard procedure for identifying movement artifacts in photoplethysmograph data consists of visual inspection of the data and manual removal of the artifacts by a trained experimenter (Laan et al., 1995; Meston & Heiman, 1998). The experimenter scrolls through the raw data on a computer screen and identifies artifacts by visually comparing each peak to its neighboring peaks and removing peaks that do not fit in size with the others. The description of this manual data-processing procedure in academic papers is usually brief and general, failing to specify the exact definition of a movement artifact (e.g., see Chivers, Roy, Grimbos, Cantor, & Seto, 2014; Heard-Davison, Heiman, & Kuffel, 2007; Laan et al., 1995; Meston & Heiman, 1998; Peterson, Janssen, & Laan, 2010; Rellini & Meston, 2011; Suschinsky & Lalumière, 2011). Some researchers, however, provide more specificity, such as noting that data points must be greater than a 100% increase from nearby data points in order to be deemed an artifact (Bloemers et al., 2010), and others set the criterion at “3 standard deviations above the mean of the 15-second period [immediately before]” (van der Made et al., 2009, p. 432). Still others allow for “peak-to-trough measures  $\pm 3$  SD from the mean” and identify any points outside that range as artifacts (Lorenz, Harte, Hamilton, & Meston, 2012, p. 115). Although these three artifact descriptions are more specific, they vary significantly between research laboratories. To our knowledge, none of these more specific artifact criteria have been examined empirically, therefore little is known about their validity and potential impact on study outcomes. The procedures for artifact detection are inconsistent between research laboratories, and these discrepancies may introduce unnecessary variability to study findings reported by these laboratories. Manual data processing is also subject to human error such as fatigue and distraction, which might introduce variability into the data retained within a single lab. Other types of physiological data such as ECG and EEG are processed fully automatically (Nolan, Whelan, & Reilly, 2010) or partially automatically (Delorme & Makeig, 2004) with software toolboxes, suggesting automatic solutions for processing vaginal photoplethysmograph data may be possible.

A few other research groups have previously addressed the weakness of the current data-processing procedure for VPA data; however, none of these methods have become widely adopted. One such method performs a graphical inspection of the data via a two-pass algorithm that automatically identifies artifacts to be deleted (Both et al., 2008). This method was developed by the Technical Support Department of the

University of Amsterdam and to our knowledge is used exclusively by research teams in the Netherlands (Both Laan, et al., 2008; Both, Spiering, et al., 2008). This method does not appear to have been formally published, which limits other researchers from utilizing the method. Another method involves the use of wavelet transforms to denoise the VPA data. Rather than identifying the artifacts for removal (as in most other methods), the wavelet method reduces the size of the artifacts to bring them into line with the rest of the neighboring peaks and maintains these revised peak values in the data set (Prause, Williams, & Bosworth, 2010; Rafiee, Rafiee, & Michaelsen, 2009; Rafiee, Rafiee, Prause, & Schoen, 2011). Some benefits of this method are that it retains all of the data, it can be adapted for use with data recorded at different sampling rates, and it reduces data-processing time in comparison to the manual method. One drawback of this method is that it requires purchasing specialized software (Matlab software). Despite these many benefits, the wavelet method does not appear to be in widespread use at this time; a literature review revealed a total of three citations (Clifton, Seehuus, & Rellini, 2015; Prause, Barela, Roberts, & Graham, 2013; Prause, Staley, & Finn, 2011) since the method was published in 2010. The wavelet method is available for free via download on the MatLab Central File Exchange.

We propose an automated VPA data-processing procedure to identify movement artifacts for deletion that is free, rapid, and easily adapted for use with data recorded at different sampling rates. Our procedure was programmed in the open-source R software environment (R Foundation, 2014), and although R itself is often regarded as somewhat syntactically complex, we have organized the procedure in a program that is straightforward to use. For data recorded at 200 Hz, using this procedure requires nothing more than loading the raw data into the R software and running the program. If the data are recorded at a frequency other than 200 Hz, the user must change a single value indicating the recording frequency within the program, and then the rest of the process is fully automated. The proposed method uses a series of smoothing regression splines to model the data and identify outliers. The process involves an iterative approach in which the data are modeled, artifacts are identified and removed, and then the data are modeled again and the process repeated until all of the significant artifacts have been identified and removed. The automatic process produces an output file of all data points with the artifacts flagged. We compared the accuracy of our new method to the gold-standard manual-processing method on two sets of data: neutral data with no stimuli present in which participants were instructed to make movements to create artifacts intentionally, and sexual-arousal response data in which participants viewed a short erotic film. Potential advantages of the proposed automated method include reducing data-processing time and reducing the human error associated with the manual-processing approach.

## General method

### Overview

In both experiments women came to the laboratory and electrocardiogram and vaginal photoplethysmograph equipment was used to record their heart rate and genital response. In the first experiment, women were instructed to physically move at certain time points to intentionally create movement artifacts in their data, and in the second experiment women viewed a short neutral film clip followed by a short erotic film clip. The second experiment featured the standard erotic film presentation paradigm commonly used in female sexual psychophysiology research (Hoon, Wincze, & Hoon, 1976; Laan et al., 1995).

### Participants

For both experiments, women over the age of 18 were recruited to participate in the experiment with online and print advertisements. Interested women contacted the lab and completed an eligibility screening via phone or email. Eligible women were invited to schedule a study session.

## Measures

### Demographics

Participants completed a short demographics questionnaire assessing age, sexual orientation, race, ethnicity, marital status, and level of education.

### Vaginal response

Vaginal photoplethysmography was used to assess genital response. Electrocardiography (ECG) was used to assess heart rate simultaneously with the photoplethysmograph assessment. Heart rate was assessed with a three-lead ECG with disposable electrodes positioned by the experimenter using a three-limb site paradigm based on Einthoven's triangle (upper right chest, lower left chest, and inner right ankle). The vaginal photoplethysmograph (Behavioral Technology, Inc., Salt Lake City, UT) is a tampon-shaped device with an infrared light-emitting diode embedded inside, inserted into the vagina by the participant, and secured with a placement shield around the cord to ensure the device remains at a consistent depth within the vagina during the experiment. ECG and VPA data were collected at a sampling rate of 200 Hz and recorded on a Dell personal computer using AcqKnowledge software version 3.9.1 and a Model MP100 data acquisition unit (BIOPAC Systems, Inc., Santa Barbara, CA). The VPA signal was amplified by 20 and was bandpass filtered with a high pass filter of .5 Hz during data collection. Heart rate was recorded in volts (V) and the amplitude of the VPA peak-to-trough waveforms was recorded in millivolts (mV).

### Procedure

All study procedures were performed in accordance with the ethical standards of the institutional and national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. All study procedures were approved by the Institutional Review Board of The University of Texas at Austin. A research administrator greeted participants and informed consent was obtained from all individual participants included in the study. The administrator oriented participants to the study procedures, including how to insert the vaginal photoplethysmograph and attach the ECG monitoring equipment. The research administrator cleaned the participant's skin with a small alcohol pad and attached the disposable electrodes to participant's upper right chest, lower left chest, and inner right ankle. The participant was left alone in the experiment room to attach the equipment. The participant and administrator communicated through an intercom system for the rest of the session. At this point the procedure for the two studies diverged; specific details for each experiment are provided in their respective sections.

### Overview of data processing and comparison

The current article presents two experiments: Experiment 1 on intentional artifact neutral data, and Experiment 2 on sexual-arousal response data. Data from both experiments were processed and evaluated in the same manner, therefore we present the common data processing and comparison approach here. We note that the evaluation of the automatic program to replicate the results of the manual program occurred within each experiment respectively.

In order to compare the manual and automated data-processing methods, each method was applied to the raw data to identify artifacts. Each processing method produced a Microsoft Excel file that included time, VPA magnitude, and a column denoting each VPA peak as either no artifact (0) or artifact (1). For the manual method, the first author (a researcher with five years of experience in manual data processing), who was blind to the times when participants were instructed to make the movements to create artifacts, visually reviewed the raw data files in AcqKnowledge and noted artifacts manually in the Excel file for the manual method. The automated method was conducted in the R open-source software environment and includes the following three steps: (1) identification of heartbeats, (2) identification of VPA peaks, and (3) identification of artifacts, and produces an Excel output file with the artifacts flagged (0 for no artifact, 1 for artifact). Next the results from both methods were compiled in a single Excel file

and the classification of VPA peaks was compared, with the manual process treated as the gold-standard method. The following sections explain the steps of the manual and automated processes in detail.

### ***Manual data reduction and processing***

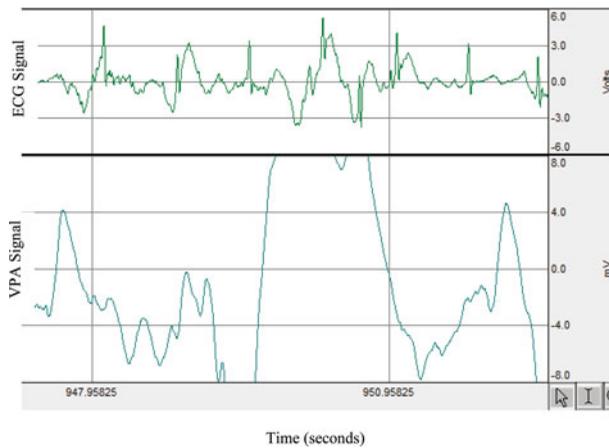
For the manual data-processing method, VPA data were manually extracted from the AcqKnowledge files into Excel, and the first author viewed the raw data file to determine artifacts, which she noted in the Excel file. VPA peaks were identified using the Find Peak tool in AcqKnowledge software version 3.9.1, in which the researcher sets a threshold in the ECG channel to capture the R peak of each QRS complex, and these heartbeat internals are used to define VPA peaks in the VPA channel (see BIOPAC Systems, Inc., 2014, p. 264, for more detailed information on the Find Peak tool). The Peak Finder tool identifies R peaks to highlight the VPA peak associated with each heartbeat interval. The researcher sets the threshold for the ECG channel by viewing the ECG waveforms and selecting a value low enough to capture the R peaks but high enough not to capture the T waves that occur after the R peaks. The threshold value used in manual data processing is generally between 1.0 and 2.0 volts. If the selected threshold misses some of the R peaks, it can be reset to a lower value during data extraction to ensure that all R peaks are identified and therefore all VPA peaks are defined. The Find Peak tool prints the time stamp from the second ECG peak surrounding each VPA peak and the VPA peak magnitude into the AcqKnowledge Journal, which can then be exported to Excel.

The Find Peak tool was used to extract the time stamps and VPA magnitudes from each raw data file, and data were then exported to Excel. Raw VPA data were visually inspected by the first author by scrolling through the data files and flagging VPA movement artifacts in the Excel file, using the same procedure that was used in previous studies (e.g., Chivers et al., 2013; Laan et al., 1995; Rellini & Meston, 2011). Movement artifacts were defined as VPA peaks that extended beyond the standard display window (8 mV to +8 mV), or that appeared noticeably different from their neighboring peaks in size. If a participant had a high baseline VPA such that the majority of her VPA peaks extended beyond the standard display window, then the window was enlarged to include the entire waveform so that artifacts could be accurately evaluated. Sections in which movement artifacts occurred in the ECG channel and R peaks could not be clearly identified were also marked as artifact; these sections will be referred to as out-of-phase peaks. Physical movement interrupts both ECG and VPA data collection, therefore artifacts in the ECG channel are generally accompanied by artifacts in the VPA channel.

### ***Automatic data reduction and processing***

For the automatic VPA data-processing method, raw ECG and VPA data were exported from AcqKnowledge software version 3.9.1 to Excel. Data were recorded at 200 Hz; therefore each Excel file included 200 readings per second for the full duration of data collection. These Excel files contained columns for “time,” “heart rate,” and “vpa” records. Data for each subject were individually submitted to the automatic processing tool, programmed in the open-source R software environment that produces an output file with the artifacts flagged. The processing tool locates the ECG peaks in order to identify the intervals for the VPA peaks (akin to the Find Peak tool in AcqKnowledge used in the manual-processing method), flags out-of-phase heartbeat peaks (explained in detail later), calculates the VPA peaks, and then flags the VPA movement artifacts. For the current study, the automatic program was set up to process ECG and VPA data recorded at 200 Hz, but can be easily modified to process data recorded at other sampling rates. The steps of the automatic data-processing program will be explained in detail in the following section, yet all that is required of the user is to import the raw data from Excel to R, set the processing speed if the speed used was different than 200 Hz, and run the automatic program, which will produce an output file with artifacts and out-of-phase heartbeat peaks flagged.

The first step of the automatic procedure is to identify the peaks of the QRS complex (R peaks) from each heartbeat waveform. For each subject, the mean and standard deviation of all the ECG records are computed and then the records that are greater than two standard deviations from the mean are identified; this provides ranges of records within which the R peaks lie. Data records within each range were examined in order to identify the specific R-peak record within each heartbeat waveform. The longest possible gap in a QRS complex is approximately .12 seconds (Cacioppo, Tassinari, & Berntson,



**Figure 2.** Example of VPA and out-of-phase heart rate peak artifacts in raw electrocardiogram and vaginal photoplethysmograph data. ECG = electrocardiography; VPA = vaginal pulse amplitude.

2007), indicating that any records within .12 seconds of one another should belong to the same heartbeat. In our data files, recorded at 200 Hz, .12 seconds corresponds to 24 records. Therefore any two records separated by less than 25 records were considered part of the same heartbeat. Once the heartbeat ranges were identified, the highest value within each heartbeat was denoted as the R peak, and each peak was associated with a time and row number in the data frame.

As with all physiological recordings, ECG assessment is also sensitive to physical movement by participants (Nolan et al., 2010). When participants move during the experiment, artifacts in the heart-rate waveform may occur in addition to the artifacts in the VPA waveform. Artifacts in the ECG channel are usually accompanied by artifacts in the VPA channel, but the reverse is not true; artifacts in the VPA channel are *not* necessarily accompanied by artifacts in the ECG channel. The time between the two R peaks of these out-of-phase heartbeats will have different durations than normal heartbeats and will not accurately define the window of a VPA peak. Therefore these out-of-phase heartbeats must be removed (see Figure 2 for examples of out-of-phase heartbeats). The automatic process calculates the mean and standard deviation of the time intervals between heartbeats and performs an outlier test using a Bonferroni-adjusted alpha of .05 (.05/number of intervals) on the time intervals between R peaks to identify out-of-phase heartbeat peaks. These out-of-phase peaks are flagged in the output file of the processing program where they may later be examined and retained or deleted according to the standards of the experimenter. In the current study, out-of-phase heartbeats were considered artifact and combined with the VPA artifacts to calculate total number of artifacts per file.

With the boundaries of each heartbeat identified, the next step is identifying the VPA peak (i.e., the maximum VPA record) within each heartbeat interval. We modeled this part of the process after the Find Peak tool in AcqKnowledge, which uses the interval between two ECG peaks to identify VPA peaks (BIOPAC Systems, Inc., 2014). In the Find Peak tool, the time stamp for the second ECG peak is the time associated with the VPA peak for that interval, therefore we used the time stamp from the second ECG peak in our automated system as well to facilitate comparison with the manual-processing method. The automatic program creates a data frame including these time stamps and VPA peaks (i.e., maximum VPA records).

With time and VPA values having been determined and accumulated in the data frame, the automatic method uses a standard generalized cross-validated smoothing spline to create a smooth trajectory through all of the VPA data records. Critical ancillary packages used within the R open-source software environment were the mgcv package (Wood, 2006, 2011) and the car package (Fox & Weisberg, 2011). In the mgcv package, the gam function was used, which selects the degree of smoothing via generalized cross-validation, which (among other things) approximates the well-known Akaike Information Criterion (AIC; Akaike, 1974) for optimal model selection. Default settings in the gam function were used:

family was set to gaussian, weights were set to null, and method was set to “GCV.Cp.” The car package was used for the Bonferroni-adjusted outlier tests. To account for the fact that VPA data are time-oriented, we also included an auto-correlated error term during the gam call (i.e., using the correlation = corAR1() option). The smoothing spline solution represents the VPA trajectory through the experimental time period for one individual subject. The residuals are assumed to be zero-mean and normally distributed, and the standard deviation of the residuals is used to compute  $z$  scores for each VPA record (Please see the Appendix for the statistical code for the automated method). The  $z$  scores are converted to probability values for each record. If the probability value for a record is less (i.e., more extreme) than  $1/N$ , with  $N$  being the total number of VPA records in the data file, then that record is identified as an artifact. Identified artifacts are removed from the data frame to facilitate the next step of the process (note that all artifacts are retained in the output Excel file in which artifacts are flagged).

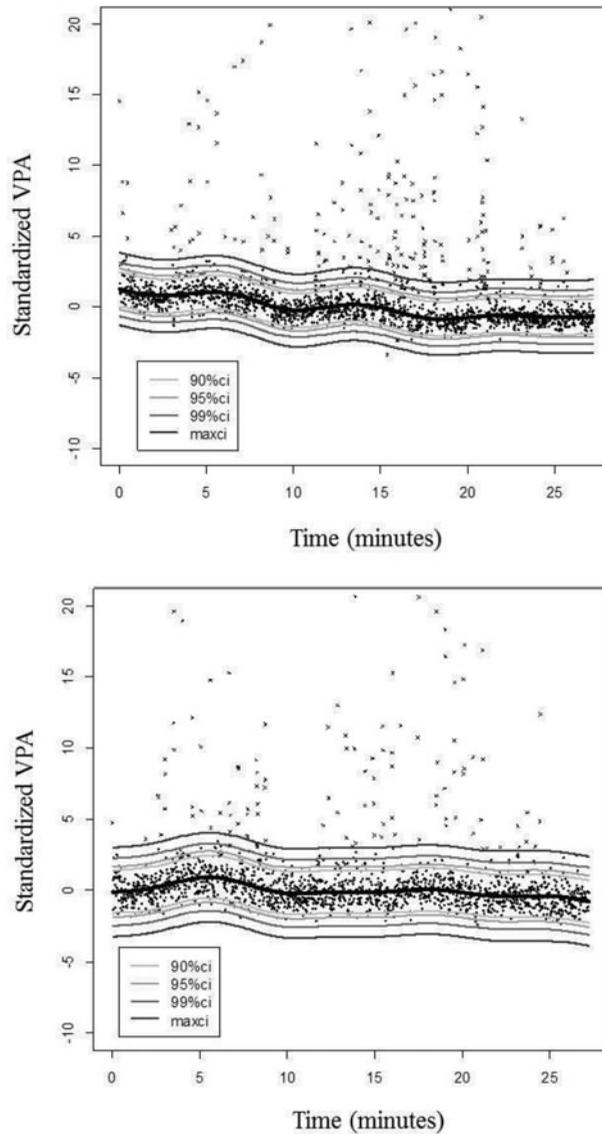
The process then repeats, starting with fitting a new smoothing spline through the remaining VPA records; the shape of this smoothing spline will necessarily be slightly different than it was in the first pass as the most extreme outliers will have been removed. The removal of those artifacts affects the new smoothing spline as well as the standard deviation used to calculate the  $z$  scores and  $p$  values, and alters the  $N$  used to evaluate the probability values to determine which records are artifacts. This iterative process continues through as many cycles as necessary to identify and eliminate all of the artifacts. The iterations terminate and all artifacts are deemed to have been removed when there are no longer any  $p$  values produced that are less than  $1/N$  (see Figure 3 for examples of processed data). Finally, any VPA record situated between two artifacts with less than three heartbeats between them was also denoted as artifact. Physical movement by the participant typically produces artifacts that span more than a single VPA peak, therefore any VPA record of normal magnitude between two VPA records identified as artifacts was also identified as an artifact. In Figure 3, these normal magnitude peaks within an artifact are represented by the Xs within the confidence bounds. The processing tool produces an Excel output file with the VPA movement artifacts and out-of-phase heartbeats flagged. The artifacts and out-of-phase heartbeats may then be reviewed by the experimenter.

### Statistical analyses

The automatic-processing method was compared to the gold-standard manual method with measures of sensitivity and specificity. For the automatic method, both VPA artifacts and out-of-phase heart-rate peaks were considered artifacts: the column noting VPA artifacts (0/1) and the column noting out-of-phase heart-rate peaks (0/1) were combined into a “total artifacts” column, and any data point that received a 1 for either type of artifact (or both types) was considered artifact in future steps of the comparison process. As previously noted, data from both methods were compiled in a single Excel file in which each VPA peak was noted as no artifact (0) or artifact (1) by each method (time stamps for each VPA peak were identical for the two methods). VPA peaks identified by both methods as artifacts (1) were labeled “hits.” VPA peaks identified through the manual method as artifact (1) and the automatic method as no artifact (0) were labeled “misses.” VPA peaks identified through the manual method as no artifact (0) and the automatic method as artifact (1) were labeled “false alarms.” VPA peaks identified through both methods as no artifact (0) were labeled “correct rejections.” Then, the values for hits, misses, false alarms, and correct rejections were used to calculate sensitivity and specificity indices.

Accuracy of artifact detection was evaluated with statistics of sensitivity and specificity;  $d'$  prime is a sensitivity index based on signal-detection theory that compares accuracy in correctly identifying artifacts when they occur (i.e., hit rate versus false alarm rate; Macmillan & Creelman, 1991). The  $d'$ -prime statistic was calculated with the following formula:  $d' = z(H) - z(FA)$ ,  $H$  = hit rate,  $FA$  = false alarm rate. The hit rate and false alarm rate were standardized by converting them to  $z$  scores, and then  $d'$  prime was computed for each participant individually. The  $d'$ -prime value for each participant, as well as the mean and standard deviation  $d'$ -prime values for the sample as a whole, are reported. Higher values of  $d'$  prime indicate more sensitive signal detection. Acceptable  $d'$ -prime values are approximately 2.0 or above (Macmillan & Creelman, 1991).

Sensitivity was also evaluated by calculating the percentage of artifacts identified by the gold-standard manual method that the automated method correctly identified as artifacts (i.e., sensitivity = detected



**Figure 3.** Graphical representation of the automated artifact-detection program for two individual subjects. The black line in the center of the band is the smoothing regression spline, and the gray lines on each side represent confidence intervals around the smoothing spline. Data are represented by dots, and artifacts are represented by Xs. Data that fall outside of the widest confidence band are identified as artifacts. Data points that did not appear to be artifacts by their magnitude but occurred immediately between two other artifacts were also identified as artifacts, and they account for the Xs within the confidence intervals. VPA = vaginal pulse amplitude.

artifacts/total artifacts; Nolan et al., 2010). Specificity refers to correctly reporting when artifacts do *not* occur (i.e.,  $\text{specificity} = 1 - [\text{false alarms}/\text{total nonartifacts}]$ ). Sensitivity and specificity percentages for each participant, as well as the mean and standard deviation values for the sample as a whole, are reported. The ideal signal-detection method is 100% sensitive (identifies every artifact correctly) and 100% specific (identifies every nonartifact correctly), yet that ideal is rarely achieved. Researchers aim to strike a balance between the detection process being too sensitive and leading to a high false-alarm rate, and being too specific and leading to a high miss rate. Comparison of the misses and false alarms of a signal-detection method can help to evaluate the relative balance of errors.

We also examined the impact of the two data-processing methods on study summary statistics. Artifacts identified by each method were removed from the data files respectively, and means, standard

deviations, and change scores (for Experiment 2 data only) were computed and compared with paired samples *t* tests. The results of these comparisons are reported and discussed.

## Experiment 1

In Experiment 1, participants were asked to make specific physical movements that have previously been associated with movement artifacts to create intentional movement artifacts in their data. No stimuli were presented during the experiment.

## Method

### Participants

Participants were women over the age of 18 recruited from the psychology department of a large public university. Exclusion criteria for the study included history of sexually transmitted diseases, history of pelvic surgery, current pelvic or urinary tract infections, and experiencing menopause or perimenopause. The sample included 17 women ages 21 to 32, with a mean age of 24.88 ( $SD = 2.96$ ). All participants had completed some years of college, with the majority having completed college (82.35%). In terms of sexual orientation, 88% of the sample identified as heterosexual, one woman as bisexual, and one woman as lesbian. Eighty-two percent of the sample was Caucasian and 18% was Asian. The sample was approximately evenly split between single women (47.06%) and women in a committed relationship (52.94%).

### Procedure

A research administrator greeted participants and oriented them to the study procedures. The administrator explained the methodology of vaginal photoplethysmography and that the purpose of the study was for participants to make intentional movements to create movement artifacts in their data on purpose in order to compare data-processing tools for artifact detection. Participants were seated in a large comfortable armchair during the experiment. The administrator then explained the seven specific movements participants would be asked to make on cue during the experiment. These movements included (a) a Kegel exercise (contracting the muscles of the pelvic floor); (b) lifting her left leg; (c) lifting her left arm; (d) moving from left to right in the chair (described by the research administrator as “wiggling back and forth in the chair”); (e) pushing her body up in the chair using the armrests; (f) slouching down in the chair; and (g) reaching for the intercom with her right hand. These movements were selected as they are the ones most commonly reported by participants in previous studies in our laboratory and movements we have observed participants make in the chair during the orientation section of previous studies. Contraction of the pelvic floor (i.e., a Kegel exercise) has been associated with movement artifacts in previous laboratory research on women’s sexual-arousal response (Laan et al., 1995). Participants were asked to practice each movement for the research administrator to ensure they understood and could adequately perform each one. Participants were told they would be asked to make each movement at the administrator’s verbal signal for one full second. Each movement was made six times with a pause of 30 seconds between each movement.

The research administrator left participants alone to insert the probe and attach the ECG leads, and subsequently communicated with the participants via the intercom system. To enable participants to focus their full attention on making the movements at the indicated times, no audio or audiovisual stimuli were presented. After participants attached the equipment, they were instructed to sit as still as possible in the chair and baseline data were recorded for three minutes. Then the research administrator followed a written time line in which she used the intercom to instruct participants to make each movement for one full second, with a 30-second pause between each movement. This procedure resulted in a total of 42 movements. The research administrator tagged the time she gave the instruction to make each movement with the typed-in labels in AcqKnowledge. At the end of the series of 42 intentional movements, the

**Table 1.** Experiment 1 intentional artifact neutral descriptive data.

Subject	Manual Artifacts	Auto Artifacts	Hits	Misses	False Alarms	Correct Rejections	<i>d</i> prime	Sensitivity (%)	Specificity (%)
1	184	339	299	15.90	14.02	2,132	2.48	87.50	90.86
2	140	138	754	6.11	33.38	2,259	3.17	81.43	98.87
3	271	305	1,673	13.98	76.67	2,182	4.52	99.26	98.12
4	249	297	641	13.96	30.12	2,128	3.85	96.39	96.96
5	218	216	907	8.69	36.50	2,485	4.14	94.04	99.51
6	209	231	636	12.34	33.97	1,872	3.15	89.00	97.29
7	216	234	1,389	11.24	66.75	2,081	4.35	98.15	98.82
8	266	184	623	10.67	36.14	1,724	3.02	66.54	99.52
9	490	457	1,149	21.85	54.92	2,092	3.04	85.51	97.63
10	257	295	936	13.08	41.51	2,255	3.49	97.67	97.30
11	259	269	1,217	13.03	58.96	2,064	3.40	90.73	98.12
12	286	345	454	16.70	21.97	2,066	3.39	95.10	95.90
13	319	305	1,047	11.85	40.68	2,574	3.52	89.03	99.07
14	314	291	533	14.82	27.15	1,963	3.28	85.67	98.67
15	181	170	802	9.52	44.90	1,786	4.32	92.27	99.81
16	197	196	391	10.85	21.64	1,807	3.81	92.39	99.13
17	254	239	521	12.11	26.39	1,974	3.09	83.07	98.37
<i>M</i>	253.53	265.35	266.71	27.47	39.24	1,792	3.53	89.63	97.88
<i>SD</i>	77.57	77.55	68.18	23.63	40.37	231.24	.56	7.99	2.09

Note. This table presents the number of artifacts identified by the manual process and the automatic process for the intentional artifact neutral data and, the number of hits, misses, false alarms, and correct rejections when treating the manual process as the gold-standard method of data processing. This table also presents the *d*-prime sensitivity index (acceptable *d*-prime values are 2.0 or above; Macmillan & Creelman, 1991), and the sensitivity and specificity percentages (as described by Nolan et al., 2010).

participant was again asked to remain as still as possible and an additional three minutes of data were recorded. The data collection period lasted approximately 27 minutes and 12 seconds (1,632 seconds) for each participant. After participants removed the probe and the ECG wires, they completed the self-report questionnaire on demographics and were debriefed by the research administrator.

## Results and discussion

For the neutral data, the number of artifacts identified by the manual method ranged from 140 to 490, with a mean of 253.53 ( $SD = 77.57$ ). The automated method identified between 138 and 457 artifacts per file, with a mean of 265.35 ( $SD = 77.55$ ; note that the automatic artifact totals include both VPA artifacts and out-of-phase heart-rate peaks). Using the manual method as the gold standard, artifact detection was compared between the two methods to determine hits, misses, false alarms, and correct rejections. The *d*-prime sensitivity index ranged from 2.48 to 4.52, with a mean of 3.53 ( $SD = .56$ ). See Table 1 for number of artifacts, hits, misses, false alarms, correct rejections, and *d*-prime values for each method, per participant. As *d*-prime values of 2.0 and above represent acceptable sensitivity (Macmillan & Creelman, 1991), it appears that the automatic method performed comparably to the manual method in detection of artifacts. The other metric of sensitivity (i.e., correctly identifying artifacts when they occur; Nolan et al., 2010) ranged from 66.54% to 99.26%, with a mean of 89.63% ( $SD = 7.99\%$ ). The metric for specificity (i.e., correctly identifying when artifacts have *not* occurred; Nolan et al., 2010) ranged from 90.86% to 99.81%, with a mean of 97.88% ( $SD = 2.09\%$ ). The additional sensitivity and specificity metrics support the conclusion that the automated method acceptably approximates the manual method in detection of artifacts. As previously noted, signal-detection programs aim to balance between misses and false alarms. There was no significant difference between the number of misses and false alarms,  $t(16) = .97$ ,  $p = .35$ , with a paired samples *t* test, suggesting that when the automated program makes errors, those errors are balanced. The automatic program does not appear to be overly sensitive or overly specific, rather the threshold for artifact detection appears to be appropriate.

The impact of each data-processing method on summary statistics for the neutral data set was also examined. As the current study did not include any stimuli, the mean VPA and VPA standard deviation for the entire study period were calculated after artifacts identified by the manual method and the automated method were removed from the data respectively (see Table 2 for summary statistic means and standard deviations per participant). Mean VPA with the manual method was normally distributed

**Table 2.** Experiment 1 intentional artifact neutral data summary statistics.

Subject	Manual		Auto	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
1	8.69	2.95	8.93	2.70
2	5.50	1.34	5.46	1.25
3	2.42	1.14	2.35	1.00
4	6.87	1.84	6.78	1.70
5	5.83	2.45	5.83	2.43
6	16.35	4.16	16.33	4.07
7	4.04	1.18	4.00	1.11
8	13.69	5.80	14.18	6.17
9	9.51	3.58	9.59	3.56
10	4.66	2.11	4.53	1.87
11	6.25	2.34	6.15	2.12
12	8.69	3.96	8.23	3.04
13	6.06	1.74	6.02	1.70
14	11.43	3.28	11.40	3.23
15	7.25	2.27	7.28	2.28
16	9.64	3.24	9.60	3.11
17	6.53	3.30	6.46	3.26
<i>M</i>	7.85	2.75	7.83	2.62
<i>SD</i>	3.53	1.23	3.60	1.28

Note. This table presents the means and standard deviations for the data retained (i.e., nonartifact data) by both the manual process and the automatic process for the intentional artifact neutral data. There were no statistically significant differences between the mean VPA and VPA standard deviation in files processed with the two methods respectively. VPA = vaginal pulse amplitude.

according to the Shapiro-Wilk test,  $W(17) = .94, p = .31$ , and mean VPA with the automatic method was also normally distributed,  $W(17) = .93, p = .25$ . A paired samples *t* test revealed no significant difference in mean VPA between the two methods of data processing,  $t(16) = .41, p = .68$ . VPA standard deviation for the manual method was normally distributed,  $W(17) = .94, p = .36$ , and VPA standard deviation was also normally distributed,  $W(17) = .91, p = .10$ . A paired samples *t* test indicated no significant difference in VPA standard deviation between the two methods of data processing,  $t(16) = 2.02, p = .06$ . The discrepancies between data identified as artifact by the manual and automated processing methods did *not* significantly affect summary statistics in this experiment, suggesting that the automated method is a promising tool for VPA data processing.

The manual-processing method took approximately two hours to process each 27-minute and 12-second (1,632 seconds) file. The automated procedure took approximately 20 minutes to process each file on an 8GB Intel® Core™ Due CPU Dell personal computer. The approximate one hour and 40-minute reduction in time to process a single file suggests that the automatic process could reduce the time burden of processing VPA data.

Given that the automatic process identified artifacts at a comparable rate to the manual process on neutral data, the efficacy of the automatic program was next tested on sexual-arousal response data. These data differed in two ways from the neutral data: (1) VPA magnitude increased in response to the erotic stimulus, and (2) data included naturally occurring artifacts rather than intentional artifacts.

## Experiment 2

In Experiment 2, participants viewed a short neutral film clip followed by a short erotic film clip while their heart rate and genital sexual-arousal response were measured. These data were collected as part of a larger study on women's sexual arousal response.

## Method

### Participants

Participants were women over the age of 18 recruited from the local community. Exclusion criteria for the study included history of sexually transmitted diseases, history of pelvic surgery, current pelvic or

urinary tract infections, experiencing menopause or perimenopause, being pregnant or breastfeeding, and current diagnosis of a psychotic disorder. Inclusion criteria for the study included self-report of a heterosexual or bisexual sexual orientation, as the stimulus film featured a heterosexual couple. The sample included 20 women ages 18 to 39, with a mean age of 25.65 ( $SD = 5.87$ ). In terms of education history, 40% of the sample completed some years of college, 35% completed college, 20% held an advanced degree, and 5% completed some years of high school. In terms of sexual orientation, 85% of the sample identified as heterosexual, and 15% of the sample identified as bisexual. The racial composition of the sample included 75% Caucasian women, 15% Asian women, one African American/Black woman, and one multiracial woman. Eighty-five percent of participants were married or in a committed relationship, and 15% of participants were single.

## Materials

The stimulus was a 10-minute film composed of a neutral (four-minute) and an erotic (six-minute) film clip. The neutral portion featured nature scenes accompanied by classical music, and the erotic portion featured a heterosexual couple engaged in foreplay (two minutes), cunnilingus (two minutes), and vaginal intercourse (two minutes). This erotic film has previously been found to elicit both genital and psychological sexual arousal in heterosexual and bisexual women in our laboratory.

## Procedure

A research administrator greeted participants and oriented them to the study procedures. Participants were left alone in a private room to attach the equipment and communicated with the research administrator through an intercom system. Prior to the film presentation, participants completed two short questionnaires on subjective sexual arousal and state affect that were not analyzed in the current study. Women viewed the erotic film while their heart rate and genital sexual arousal were assessed. Participants also reported continuous subjective sexual arousal during the film presentation. After the film, participants removed the equipment, got dressed, and were escorted into a separate room to complete a battery of self-report questionnaires on a computer, starting with the demographics questionnaire described in the General Method section. The other questionnaires were not analyzed in the current study. The research experimenter debriefed participants at the end of the study.

## Results and discussion

For the sexual response data, the number of artifacts identified by the manual method ranged from 11 to 151, with a mean of 58.45 ( $SD = 38.13$ ). The automated method identified between 12 and 138 artifacts per file, with a mean of 51.70 ( $SD = 34.70$ ; note that the automatic artifact totals include both VPA artifacts and out-of-phase heart-rate peaks). The  $d$ -prime sensitivity index ranged from 1.16 to 4.27, with a mean of 2.61 ( $SD = .76$ ). See Table 3 for number of artifacts, hits, misses, false alarms, correct rejections, and  $d$ -prime values for each method, per participant. Given that acceptable  $d$ -prime values are 2.0 and greater (Macmillan & Creelman, 1991), the automatic method acceptably identified artifacts in the sexual response data. Only three participants' data showed  $d$ -prime values under 2.0, and these will be discussed in the General Discussion. The sensitivity metric (Nolan et al., 2010) ranged from 18.91% to 100%, with a mean of 60.05% ( $SD = 26.08\%$ ). The specificity metric (Nolan et al., 2010) ranged from 96.94% to 99.81%, with a mean of 98.62% ( $SD = .87\%$ ). There was no significant difference between the number of hits and false alarms per file,  $t(19) = 1.30$ ,  $p = .21$ , with a paired samples  $t$  test, suggesting that the automatic program is balanced in terms of signal detection (i.e., the program is not overly sensitive or overly specific).

The impact of each data-processing method on summary statistics for the sexual response data set was also examined. Artifacts identified by each method were removed from the data files respectively, and then means and standard deviations were computed for the neutral and erotic film sections separately. Studies using vaginal photoplethysmography typically focus on the change in VPA from the neutral

**Table 3.** Experiment 2 sexual-arousal response descriptive data.

Subject	Manual Artifacts	Auto Artifacts	Hits	Misses	False Alarms	Correct Rejections	<i>d</i> prime	Sensitivity (%)	Specificity (%)
1	13	12	9	4	3	1,137	3.29	69.23	99.74
2	151	138	126	25	12	1,251	3.32	83.44	99.05
3	83	28	24	59	4	973	2.09	28.92	99.59
4	59	18	16	43	2	893	2.23	27.12	99.78
5	41	55	36	5	19	1,394	3.38	87.80	98.66
6	71	69	49	22	20	1,140	2.61	69.01	98.28
7	26	58	15	11	43	1,362	2.07	57.69	96.94
8	80	85	68	12	17	1,210	3.24	85.00	98.61
9	77	36	27	50	9	1,382	2.10	35.06	99.35
10	37	27	7	30	18	855	1.16	18.92	97.94
11	33	14	10	23	4	1,174	2.19	30.30	99.66
12	107	70	55	52	15	1,077	2.24	51.40	98.63
13	29	43	12	17	31	1,391	1.80	41.38	97.82
14	64	55	26	38	29	1,702	1.89	40.63	98.32
15	88	67	49	39	18	1,084	2.28	55.68	98.37
16	51	36	24	27	12	1,053	2.21	47.06	98.87
17	17	44	17	1	27	1,268	3.63	100.00	97.92
18	11	14	11	1	3	1,561	4.27	100.00	99.81
19	17	37	15	2	22	1,108	3.25	88.24	98.05
20	114	128	96	18	32	1,030	2.88	84.21	96.99
<i>M</i>	58.45	51.70	34.6	23.95	17	1,202.25	2.61	60.05	98.62
<i>SD</i>	38.13	34.70	31.58	18.09	11.36	215.66	.76	26.08	.87

Note. This table presents the number of artifacts identified by the manual process and the automatic process for the sexual-arousal response data, the number of hits, misses, false alarms, and correct rejections when treating the manual process as the gold-standard method of data processing. This table also presents the *d*-prime sensitivity index (acceptable *d*-prime values are 2.0 or above; Macmillan & Creelman, 1991), and the sensitivity and specificity percentages (as described by Nolan et al., 2010).

to the erotic film presentations, which is computed by subtracting the mean VPA during the neutral film from the mean VPA during the erotic film (i.e., change in VPA = mean VPA erotic – mean VPA neutral; Chivers et al., 2013; Clifton et al., 2015). Change in mean VPA was calculated in this manner, and difference scores were also compared between the data files with artifacts removed by the manual- and automatic-processing methods respectively (see Table 4 for summary statistic means and standard deviations per participant).

Mean VPA during the neutral film that was processed with the manual method was normally distributed according to the Shapiro-Wilk test,  $W(20) = .93$ ,  $p = .14$ , and mean VPA during the neutral film that was processed with the automatic method was normally distributed,  $W(20) = .93$ ,  $p = .13$ . A paired samples *t* test indicated no significant difference in mean VPA between the two methods of data processing,  $t(19) = .08$ ,  $p = .94$ . Using the manual method, VPA standard deviation during the neutral film was not normally distributed,  $W(20) = .84$ ,  $p = .004$ , which appeared to be due to one outlier (Participant 2 with a value of 4.01 see Table 4) that was greater than three standard deviations above the mean. When this outlier was removed from the data set, the data were normally distributed,  $W(19) = .96$ ,  $p = .53$ . Using the automatic method, VPA standard deviation during the neutral film was not normally distributed,  $W(20) = .87$ ,  $p = .01$ , which appeared to be due to the same outlier (Participant 2 with a value of 4.05 see Table 4). With the outlier removed from the data set, the data were normally distributed,  $W(19) = .95$ ,  $p = .46$ . A paired samples *t* test revealed no significant difference in VPA standard deviation between the two methods of data processing,  $t(18) = 1.24$ ,  $p = .23$ , for the neutral film. The discrepancies in data identified as artifact between the manual- and automated-processing methods did not significantly affect summary statistics for the neutral film, suggesting that the automatic approach is acceptable.

Mean VPA during the erotic film that was processed with the manual method was normally distributed,  $W(20) = .95$ ,  $p = .44$ , and mean VPA during the erotic film that was processed with the automatic method was also normally distributed,  $W(20) = .95$ ,  $p = .44$ . A paired samples *t* test indicated a statistically significant difference in mean VPA between the two methods of data processing,  $t(19) = 2.62$ ,  $p = .02$ . Mean VPA for the manual method was slightly higher with a group mean of 8.96, while mean VPA for the automatic method was 8.88—a difference of .08. Discrepancies in the data identified as artifact between the two processing methods will be examined in the General Discussion in an

**Table 4.** Experiment 2 sexual-arousal response data summary statistics.

Subject	Neutral Film				Erotic Film				M Change	
	Manual		Auto		Manual		Auto		Manual	Auto
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
1	5.11	1.79	5.11	1.79	5.37	1.48	5.36	1.45	.26	.24
2	12.95	4.01	13.13	4.05	9.79	3.72	9.84	3.75	-3.16	-3.30
3	8.08	1.97	8.19	2.15	16.80	5.12	16.79	5.06	8.73	8.60
4	7.01	1.48	7.01	1.48	10.90	2.24	10.89	2.21	3.89	3.87
5	5.06	2.71	4.96	2.64	9.05	2.13	9.06	2.11	3.99	4.11
6	7.51	1.19	7.53	1.22	12.81	2.82	12.74	2.60	5.30	5.21
7	3.85	1.58	3.73	1.25	5.10	2.16	4.80	1.69	1.25	1.06
8	7.80	1.47	7.81	1.41	9.43	3.32	9.29	3.14	1.63	1.48
9	6.59	2.15	6.61	2.17	6.96	2.36	6.94	2.30	.38	.32
10	6.49	1.63	6.46	1.66	9.01	2.79	8.83	2.54	2.52	2.37
11	5.09	1.14	5.09	1.14	10.15	2.33	10.15	2.33	5.06	5.06
12	5.90	1.39	5.87	1.32	8.11	1.39	8.09	1.40	2.21	2.22
13	2.83	0.93	2.83	.92	3.91	.81	3.85	.64	1.08	1.02
14	3.65	2.21	3.89	2.55	7.15	4.39	6.63	3.75	3.50	2.74
15	9.54	2.18	9.41	1.86	9.86	1.24	9.86	1.24	.33	.45
16	4.83	1.13	4.73	.86	5.25	.96	5.25	.96	.42	.52
17	3.67	1.07	3.60	.83	8.56	.95	8.50	.81	4.90	4.90
18	6.35	1.41	6.33	1.38	8.94	1.32	8.94	1.32	2.60	2.61
19	5.90	1.61	5.87	1.53	7.84	1.26	7.78	1.15	1.94	1.91
20	7.75	2.02	7.75	2.02	14.14	3.73	14.00	3.47	6.39	6.25
<i>M</i>	6.30	1.75	6.30	1.71	8.96	2.33	8.88	2.20	2.66	2.58
<i>SD</i>	2.33	.70	2.36	.76	3.13	1.22	3.16	1.17	.80	.80

Note. This table presents the mean and standard deviations for the data retained (i.e., nonartifact data) by both the manual process and the automatic process for the sexual-arousal response data. There were small but statistically significant differences between the mean VPA and standard deviation VPA during the erotic film. *M* Change refers to the difference between the mean VPA during the erotic film and the mean VPA during the neutral film; there was no statistically significant difference between the mean change VPA from data processed by the two methods. VPA = vaginal pulse amplitude.

effort to shed light on this significant difference. VPA standard deviation for the erotic film that was processed with the manual method was normally distributed,  $W(20) = .93$ ,  $p = .13$ , and VPA standard deviation for the erotic film that was processed with the automatic method was also normally distributed,  $W(20) = .94$ ,  $p = .21$ . A paired samples  $t$  test indicated a significant difference in VPA standard deviation between the two methods of data processing,  $t(19) = 3.33$ ,  $p = .004$ . Manual VPA standard deviation to the erotic film was slightly higher with a group mean of 2.33, while VPA standard deviation for the automatic method was 2.19—a difference of .14. To further examine the source of these differences, discrepancies between data identified as artifact between the two processing methods will be explored in the General Discussion.

Change in mean VPA from the neutral to the erotic films was calculated for each method. Change in mean VPA for the manual method was normally distributed,  $W(20) = .97$ ,  $p = .83$ , and change in mean VPA for the automatic method was also normally distributed,  $W(20) = .97$ ,  $p = .72$ . A paired samples  $t$  test showed no significant difference for change in mean VPA between the two methods of data processing,  $t(19) = 1.92$ ,  $p = .07$ . Although there were small but statistically significant differences in mean VPA and VPA standard deviation to the erotic film, when the mean VPA scores were combined into change scores, there was no significant difference in outcomes between the two methods of data processing.

In terms of processing time, the manual method took approximately one hour to process each sexual-arousal response file, which averaged 15 minutes (900 seconds) in length. The automated procedure took approximately 10 minutes to process each file on an 8GB Intel® Core™ Due CPU Dell personal computer. The automatic method processes data more rapidly than the manual method.

## General discussion

Vaginal photoplethysmography is the most frequently used measure of women's sexual-arousal response, and data are typically processed by visual inspection of the raw data and manual removal of artifacts. This

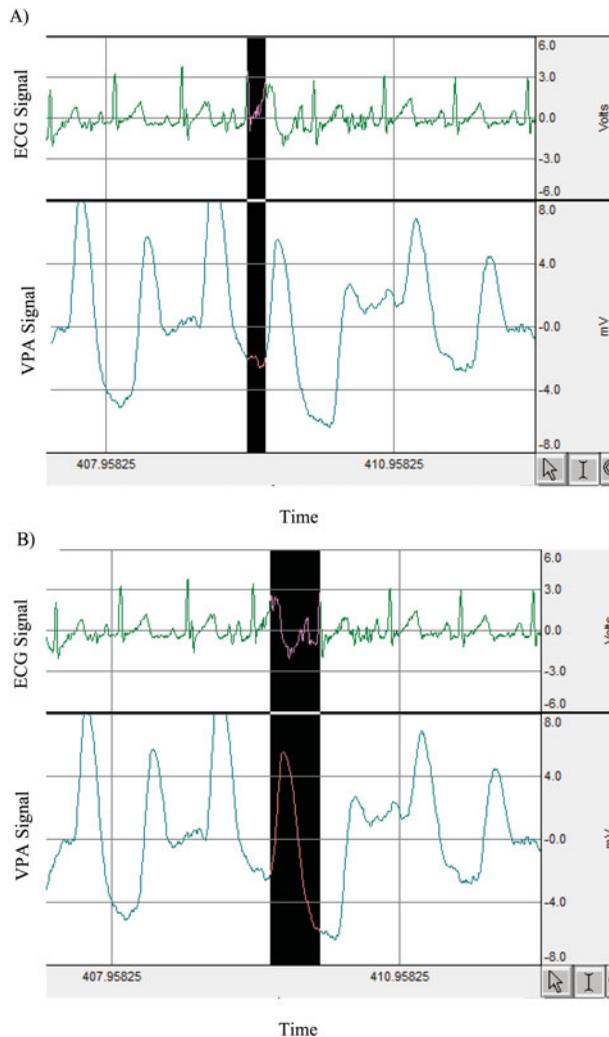
process is time-consuming and subject to human error from fatigue and distraction. We proposed a tool that attempts to automate this process. We tested the efficacy of the automated-processing procedure in two experiments, one on neutral data and one on sexual-arousal response data. Across both experiments the automated process identified the same artifacts as the manual process at an acceptable level of sensitivity and specificity, suggesting the automated tool is a promising method of data processing worthy of further testing. The errors committed by the automatic process in terms of missing artifacts that had occurred and falsely identifying artifacts when none were present appeared to be balanced. Balancing between misses and false alarms is one important goal of signal-detection programs.

When comparing the sensitivity and specificity indices between the two experiments, we saw that the automatic program performed better on the neutral data than on the sexual response data ( $d$  prime of 3.53 vs. 2.61; sensitivity metric of 89.63% vs. 60.05%). Additionally, when comparing the study outcomes for the sexual response data, there were small but statistically significant differences between the mean VPA and standard deviation VPA (although not *change* in mean VPA). In an attempt to understand the source of these discrepancies, we examined the data points for which the manual and automatic processes disagreed on artifact classification (i.e., the misses and false alarms).

Two types of discrepancy patterns were identified, which were both associated with ECG out-of-phase peaks. Although out-of-phase peaks are typically accompanied by VPA artifacts, this is not always the case. One of the discrepancies we found was the occurrence of a single out-of-phase heartbeat peak accompanied by nonartifact VPA data. The manual process identified both of the VPA records associated with this out-of-phase heartbeat peak as artifact (i.e., the VPA record associated with this peak as well as the next VPA record, following the logic that in the presence of an ECG artifact the VPA interval would not be correctly defined), but the automatic method only identified the *first* VPA record as artifact (see Figure 4). As seen in Figure 4, the second ECG interval does indeed capture the VPA peak for this interval and visually appears to be valid data.

Another notable issue that emerged was that there were entire sections with artifacts in the ECG channel but no artifacts in the VPA channel (see Figure 5). In this case the manual rater identified the entire section as artifact, but the automatic method did not identify any artifacts in this section because it was able to identify the R peaks to define the VPA intervals. When we examined the magnitudes of the VPA peaks from these sections, they appeared to be within the acceptable range for each participant, suggesting it is appropriate to retain them in the data set. This latter issue was apparent multiple times but only in a small number of data files, suggesting that some participants may be predisposed to produce ECG data of this pattern. The reader should keep in mind that in our automatic method the out-of-phase heart-rate peak identification occurs before the smoothing spline process, so that any VPA records from the second section of an out-of-phase heartbeat peak are also later subject to the smoothing spline artifact identification process (i.e., if the VPA magnitude of these records is artifact, it will be identified in this later step). In the Experiment 1 neutral data files, this discrepancy was present in approximately 2.34% of the data. In the Experiment 2 sexual data files, this discrepancy was present in approximately 1.23% of the data. In Experiment 2 these two discrepancy issues occurred most frequently in the files for Participants 10, 13, and 14, which were the three files with  $d$ -prime statistics below the acceptable threshold of 2.0. It is possible that these issues are somewhat responsible for the low  $d$ -prime values for these three files. Combining the new automatic method with the manual method could help researchers to justify retaining more of their original data than when using the manual method alone.

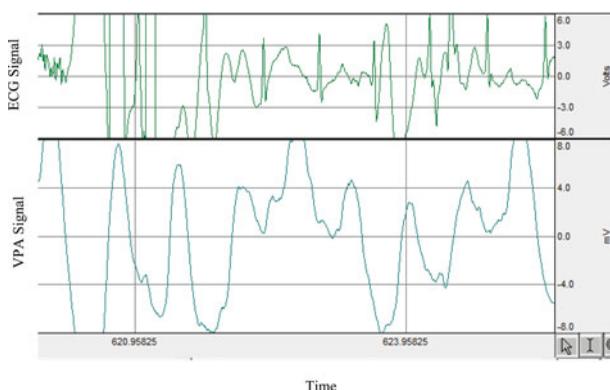
Another benefit of the automatic method is that it processed files much faster than the manual method. The manual method required approximately one to two hours to process each file, depending on the length of the study session, yet the automatic method took just 10 to 20 minutes to process each file. Studies using vaginal photoplethysmography have traditionally required a significant amount of time for data processing, as the manual process is lengthy and requires the research administrator to be alert and conscientious during processing. Studies with long protocols or multiple conditions require even more time for data processing. The automated tool is capable of processing files much more quickly than the manual method. Moreover, the automated tool can be set up to process a set of files while the researcher is working on other tasks or is not even in the laboratory.



**Figure 4.** This figure illustrates the single out-of-phase heartbeat peak discrepancy issue between the manual- and automatic-processing programs. The out-of-phase ECG heartbeat peak is shown in the center of the image. The manual method identified both the VPA record ending with this ECG R peak as well as the next ECG R peak as artifact. The automatic method, however, only identifies the VPA record associated with the out-of-phase heartbeat peak (highlighted in Panel A) as artifact. The automatic program retained the second VPA record (highlighted in Panel B) as valid data. Panel B shows that the second heartbeat interval appears to capture an artifact-free VPA peak, which should likely be retained in the data set. ECG = electrocardiography; VPA = vaginal pulse amplitude.

This study had a number of limitations that warrant mention. In Experiment 1, the research administrator followed a written schedule and used a stopwatch to verbally instruct participants to make the physical movements, which could introduce human error into the timing of these instructions. In retrospect we could have exercised more experimental control by using a prerecorded audio of the instructions, ensuring that all participants received the same instructions at the exact same times. Another limitation is that the threshold used for outlier detection becomes more stringent as the experimental sessions lengthen, and it is possible that this could change the efficacy of artifact detection for longer data files. Accuracy of the automatic-processing procedure on data files longer than 27 minutes is an important direction for future study.

There are few fully automatic methods for detecting artifacts in physiological data (Nolan et al., 2010). Although fully automated processes are alluring, best practices suggest reviewing the raw data for major issues, and using a combination of automatic processes with manual review to evaluate the efficacy of automatic tools. The automatic method we propose here appears to be a promising tool for helping



**Figure 5.** This figure illustrates the multiple out-of-phase heartbeat peak discrepancy issue between the manual- and automatic-processing programs. The manual method identified this entire section as artifact due to the ECG artifacts, but the automatic method did not identify any artifacts in this section as its method of identifying R peaks was not impaired by the ECG artifacts. The magnitude of these VPA peaks were examined and appeared to be within the range for the rest of this file, suggesting that if these data points can be extracted with the automatic method, it is appropriate to retain them in the data set. ECG = electrocardiography; VPA = vaginal pulse amplitude.

researchers to process their VPA data more quickly and in a standardized manner. More research is needed to further test the properties of this automatic process. We encourage other researchers to experiment with the automated artifact-detection tool presented here. We will make the current version of the program available on request. Ultimately, we plan to make it available as an official R package on the Comprehensive R Archive Network (CRAN) repository. R is an open-source statistical environment, and all R packages are available for free via download. The automated tool presented here detects artifacts at an acceptable level as compared to the gold standard of manual processing, is free, processes data more quickly than the manual method, and can help to standardize artifact detection between research laboratories.

## References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. doi:10.1109/TAC.1974.1100705
- Berman, J. R., Berman, L. A., Werbin, T. J., Flaherty, E. E., Leahy, N. M., & Goldstein, I. (1999). Clinical evaluation of female sexual function: Effects of age and estrogen status on subjective and physiologic sexual responses. *International Journal of Impotence Research*, 11(s1), S31–S38. doi:10.1038/sj.ijir.3900468
- BIOPAC Systems, Inc. (2014). *AcqKnowledge 4 software guide*. Goleta, CA: Author.
- Bloemers, J., Gerritsen, J., Bults, R., Koppeschaar, H., Everaerd, W., Olivier, B., & Tuiten, A. (2010). Induction of sexual arousal in women under conditions of institutional and ambulatory laboratory circumstances: A comparative study. *The Journal of Sexual Medicine*, 7(3), 1160–1176. doi:10.1111/j.1743-6109.2009.01660.x
- Both, S., Laan, E., Spiering, M., Nilsson, T., Oomens, S., & Everaerd, W. (2008). Appetitive and aversive classical conditioning of female sexual response. *The Journal of Sexual Medicine*, 5(6), 1386–1401. doi:10.1111/j.1743-6109.2008.00815.x
- Both, S., Spiering, M., Laan, E., Belcome, S., van den Heuvel, B., & Everaerd, W. (2008). Unconscious classical conditioning of sexual arousal: Evidence for the conditioning of female genital arousal to subliminally presented sexual stimuli. *The Journal of Sexual Medicine*, 5(1), 100–109. doi:10.1111/j.1743-6109.2007.00643.x
- Cacioppo, J. T., Tassinary, L. G., & Berntson, G. G. (2007). *Handbook of psychophysiology* (3rd ed.). New York, NY: Cambridge University Press.
- Chivers, M. L., Roy, C., Grimbs, T., Cantor, J. M., & Seto, M. C. (2013). Specificity of sexual arousal for sexual activities in men and women with conventional and masochistic sexual interests. *Archives of Sexual Behavior*, 43(5), 931–940. doi:10.1007/s10508-013-0174-1
- Clifton, J., Seehuus, M., & Rellini, A. H. (2015). Testing cognitive predictors of individual differences in the sexual psychophysiological responses of sexually functional women. *Psychophysiology*, 52(7), 957–968. doi:10.1111/psyp.12423
- Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21. doi:10.1016/j.jneumeth.2003.10.009
- Fox, J., & Weisberg, S. (2011). *Companion to applied regression* (2nd ed.). Thousand Oaks, CA: Sage. Retrieved from <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>

- Heard-Davison, A., Heiman, J. R., & Kuffel, S. (2007). Genital and subjective measurement of the time course effects of an acute dose of testosterone vs. placebo in postmenopausal women. *The Journal of Sexual Medicine*, 4(1), 209–217. doi:10.1111/j.1743-6109.2006.00406.x
- Hoon, P. W., Wincze, J. P., & Hoon, E. F. (1976). Physiological assessment of sexual arousal in women. *Psychophysiology*, 13(3), 196–204. doi:10.1111/j.1469-8986.1976.tb00097.x
- Kukkonen, T. M., Binik, Y. M., Amsel, R., & Carrier, S. (2010). An evaluation of the validity of thermography as a physiological measure of sexual arousal in a non-university adult sample. *Archives of Sexual Behavior*, 39(4), 861–873. doi:10.1007/s10508-009-9496-4
- Kukkonen, T. M., Paterson, L., Binik, Y. M., Amsel, R., Bouvier, F., & Khalifé, S. (2006). Convergent and discriminant validity of clitoral color Doppler ultrasonography as a measure of female sexual arousal. *Journal of Sex & Marital Therapy*, 32(4), 281–287. doi:10.1080/00926230600666220
- Laan, E., & Everaerd, W. (1995). Determinants of female sexual arousal: Psychophysiological theory and data. *Annual Review of Sex Research*, 6(1), 32–76.
- Laan, E., Everaerd, W., & Evers, A. (1995). Assessment of female sexual arousal: Response specificity and construct validity. *Psychophysiology*, 32(5), 476–485. doi:10.1111/j.1469-8986.1995.tb02099.x
- Lorenz, T. A., Harte, C. B., Hamilton, L. D., & Meston, C. M. (2012). Evidence for a curvilinear relationship between sympathetic nervous system activation and women's physiological sexual arousal. *Psychophysiology*, 49(1), 111–117. doi:10.1111/j.1469-8986.2011.01285.x
- Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York, NY: Cambridge University Press.
- Meston, C. M., & Heiman, J. R. (1998). Ephedrine-activated physiological sexual arousal in women. *Archives of General Psychiatry*, 55(7), 652–656. doi:10.1001/archpsyc.55.7.652
- Nolan, H., Whelan, R., & Reilly, R. B. (2010). FASTER: Fully automated statistical thresholding for EEG artifact rejection. *Journal of Neuroscience Methods*, 192(1), 152–162. doi:10.1016/j.jneumeth.2010.07.015
- Peterson, Z. D., Janssen, E., & Laan, E. (2010). Women's sexual responses to heterosexual and lesbian erotica: The role of stimulus intensity, affective reaction, and sexual history. *Archives of Sexual Behavior*, 39(4), 880–897. doi:10.1007/s10508-009-9546-y
- Prause, N., Barela, J., Roberts, V., & Graham, C. (2013). Instructions to rate genital vasocongestion increases genital and self-reported sexual arousal but not coherence between genital and self-reported sexual arousal. *The Journal of Sexual Medicine*, 10(9), 2219–2231. doi:10.1111/jsm.12228
- Prause, N., Staley, C., & Finn, P. (2011). The effects of acute ethanol consumption on sexual response and sexual risk-taking intent. *Archives of Sexual Behavior*, 40(2), 373–384. doi:10.1007/s10508-010-9718-9
- Prause, N., Williams, K., & Bosworth, K. (2010). Wavelet denoising of vaginal pulse amplitude. *Psychophysiology*, 47(2), 393–401. doi:10.1111/j.1469-8986.2009.00941.x
- Rafiee, J., Rafiee, M. A., & Michaelsen, D. (2009). Female sexual responses using signal processing techniques. *The Journal of Sexual Medicine*, 6(11), 3086–3096. doi:10.1111/j.1743-6109.2009.001452.x
- Rafiee, J., Rafiee, M. A., Prause, N., & Schoen, M. P. (2011). Wavelet basis functions in biomedical signal processing. *Expert Systems with Applications*, 38(5), 6190–6201. doi:10.1016/j.eswa.2010.11.050
- Rellini, A. H., & Meston, C. M. (2011). Sexual self-schemas, sexual dysfunction, and the sexual responses of women with a history of childhood sexual abuse. *Archives of Sexual Behavior*, 40, 351–362. doi:10.1007/s10508-010-9694-0
- R Foundation. (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.r-project.org/>
- Sintchak, G., & Geer, J. H. (1975). A vaginal plethysmograph system. *Psychophysiology*, 12(1), 113–115. doi:10.1111/j.1469-8986.1975.tb03074.x
- Suschinsky, K. D., & Lalumière, M. L. (2011). Category-specificity and sexual concordance: The stability of sex differences in sexual arousal patterns. *Canadian Journal of Human Sexuality*, 20(3), 93–108.
- Suschinsky, K. D., Lalumière, M. L., & Chivers, M. L. (2009). Sex differences in patterns of genital sexual arousal: Measurement artifacts or true phenomena? *Archives of Sexual Behavior*, 38(4), 559–573. doi:10.1007/s10508-008-9339-8
- van der Made, F., Bloemers, J., van Ham, D., El Yassem, W., Kleiverda, G., Everaerd, W., ... Tuiten, A. (2009). Childhood sexual abuse, selective attention for sexual cues and the effects of testosterone with or without vardenafil on physiological sexual arousal in women with sexual dysfunction: A pilot study. *The Journal of Sexual Medicine*, 6(2), 429–439. doi:10.1111/j.1743-6109.2008.01103.x
- Wood, S. N. (2006). *Generalized additive models: An introduction with R*. Boca Raton, FL: Taylor & Francis Group.
- Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 73(1), 3–36. doi:10.1111/j.1467-9868.2010.00749.x

## Appendix

# Each individual participant will have one input data file which should contain columns for time, heartrate, and vpa readings (using these *exact* column names).

# This is where the input file is, and where output files will be written.

```
setwd("/name/of/directory/path/here")
```

# Change the names of the input and output files below by inserting the desired filenames.

# If the output file already exists, delete it before running this.

# This output file contains all the original data plus flags for whether the VPA measurement is an artifact and a flag for an out-of-phase heartbeat.

```
rawinput_filename <- "filename.csv"
```

```
prepruned_heartbeat_VPA_output_filename <- "filename_prepruned_v2.csv"
```

```
recording_frequency <- 200 # Input the recording frequency in Hz, or measurements per second.
```

```
#####
```

```
#####
```

```
#####
```

```
data <- read.csv(rawinput_filename)
```

```
data$row_id <- c(1:nrow(data))
```

```
library(mgcv)
```

```
heartrate_gam <- gam(heartrate ~ s(time), data=data)
```

```
data$heartrate_res <- heartrate_gam$residuals
```

# This addresses stationarity.

```
mean_res <- mean(data$heartrate_res)
```

```
sd_res <- sd(data$heartrate_res)
```

```
data$heartrate_res_z <- (data$heartrate_res - mean_res)/sd_res
```

```
prelim_indices <- as.vector(which(data$heartrate_res_z > 2))
```

# 95% CI corresponds to  $\sim z > 2$ . This appears to work well, but can be changed by the user if desired.

```
#####
```

```
#####
```

```
#####
```

# The loop below iterates through the flagged indices.

# If successive indices are different by less than the threshold, ( $\text{recording\_frequency}/8$ ), they are assumed to be in the same R-peak.

# 0.12s is considered to be the upper bound of the QRS width and is a sensible value.

# At 200Hz, if the next index is 25 units or more away, it should be a member of a new peak in a normal heart rhythm.

# If the data are recorded at a frequency different than 200Hz, input the frequency at the top of the file and the cutoff will be adjusted here.

```

cluster <- 1
clustervector <- c(1, rep(0,length(prelim_indices)-1))

# The above initializes everything.

for (i in 2:length(prelim_indices)) {

  if (prelim_indices[i]-prelim_indices[i-1] < (recording_frequency/8)) {cluster <- cluster} else {cluster
<- cluster+1}
  clustervector[i] <- cluster

}

#####

#####

#####

# With high heart value records identified and assigned a cluster membership, we flag those records in
the main dataset.

data$prelim_indices <- rep(0,nrow(data))

# This initializes a 0/1 vector in data to flag the rows in prelim_indices.

for (i in 1:length(prelim_indices)) {

  targetrow <- prelim_indices[i]
  data$prelim_indices[targetrow] <- 1

}

#####

#####

#####

# Here we extract all variables for the high heart value records and place them in a separate dataset.
# We also add the cluster membership information and declare it to be a factor.

data_prelim_indices_subset <- data[data$prelim_indices == 1,]

data_prelim_indices_subset$clustervector <- clustervector

data_prelim_indices_subset$clustervector <- as.factor(data_prelim_indices_subset$clustervector)

#####

#####

```

```
#####

# Here we pull each cluster, identify the record in that cluster with the highest value, and store the asso-
ciated row_id.

final_indices <- rep(0,nlevels(data_prelim_indices_subset$clustervector))

for (i in 1:nlevels(data_prelim_indices_subset$clustervector)) {

  tempdata <- data_prelim_indices_subset[data_prelim_indices_subset$clustervector == i,]
  final_indices[i] <- tempdata$row_id[which.max(tempdata$heartrate)]

}

#####

#####

#####

# And in this section we flag the records corresponding to the peak heartrate in each cluster.

data$final_indices <- rep(0,nrow(data))

# This initializes a 0/1 vector in data to flag the rows in final_indices.

for (i in 1:length(final_indices)) {

  targetrow <- final_indices[i]
  data$final_indices[targetrow] <- 1

}

#####

#####

#####

# Right now, final_indices is still a vector with the actual index values in it – which is helpful.
# (In contrast to what we just created, which is a vector in the dataframe called data$final_indices that
has values 0/1.)
# This module will identify the VPA (max minus min) within the heartbeat intervals and write that VPA
and the time (end of interval) to an output file.

time <- rep(0, length(final_indices)-1) # Initializing the vector.
vpa <- rep(0, length(final_indices)-1) # Initializing the vector.

for (i in 1:length(final_indices)-1) {

  beginrow <- final_indices[i]+1 # This is the row after the ith entry in final_indices. (The ith entry is
the end of the prior interval, so the next row is the start of the current one.)
```

```

endrow <- final_indices[i+1] # This is the i+1th entry in final_indices. (The i+1th entry is the end
of the current interval.)
currentsubset <- data[data$row_id >= beginrow & data$row_id <= endrow,]
time[i] <- data$time[endrow]
vpa[i] <- max(currentsubset$vpa)-min(currentsubset$vpa)

}

phase1output <- data.frame(time,vpa)

#####

#####

#####

# This module is designed to identify out-of-phase rows.
# These will eventually be written to the pre-pruned copy of phase1output as a 0/1 vector.
# This is based on a Bonferroni-corrected outlier test on the heartbeat intervals.
# This appears to perform better (fewer false alarms) than a non-parametric boxplot on a lagged regres-
sion of the differences.
# It does a reasonable job of (usually) capturing the first aphasic short interval while not flagging the
second.

diff1 <- numeric()
diff1[1] <- NA
diff1 <- append(diff1,diff(phase1output$time,differences=1))
phase1output$time_diff1 <- diff1

library(car)

phasetest <- outlierTest(lm(phase1output$time_diff1~1, data=phase1output))

outofphasevector <- as.vector(as.numeric(row.names(as.matrix(phasetest$rstudent))))

# NOTE: At present, out-of-phase rows are NOT being deleted automatically – just flagged in the output
file.
# If we want to automatically delete them, take the programming just after the end of the repeat loop that
writes the 0/1 flag and move it up here.
# Then filter out the out-of-phase rows from data_working.

phase1output$time_diff1 <- NULL # Removing this vector as its utility is over.

#####

#####

#####

# This is the major analysis module.

# Everything that follows below is expecting an input of a time and a vpa.

```

```

phase1outputcopy <- phase1output # This copy is going to be used to write out the phase1 extract plus
the flag for whether the row is an artifact.
data_working <- phase1output # This will be used for most of the processing, graph production, and
output of post-pruned data.
data_working$row_id <- c(1:nrow(data_working))

data_working$artifact <- rep(0,nrow(data_working))
artifact_vector <- NULL

# The above initializes an artifact variable.

repeat {

artifact_vector_length_prior <- length(artifact_vector)

data_working <- data_working[data_working$artifact==0,]

vpa_mean <- mean(data_working$vpa)
vpa_sd <- sd(data_working$vpa)

data_working$vpa_z <- (data_working$vpa-vpa_mean)/vpa_sd

time_ac <- gamm(vpa_z ~ s(time), method = "ML", correlation = corAR1(), data=data_working)

sd_res <- sd(time_ac$gam$residuals)

data_working$pred_z <- as.vector(predict(time_ac$gam))

max_z <- qnorm(1-(1/(2*nrow(data_working))))

data_working$sd_pred_z <- rep(0,nrow(data_working))
data_working$p_artifact <- rep(0,nrow(data_working))
data_working$lclmax <- rep(0,nrow(data_working))
data_working$uclmax <- rep(0,nrow(data_working))
data_working$lcl99 <- rep(0,nrow(data_working))
data_working$ucl99 <- rep(0,nrow(data_working))
data_working$lcl95 <- rep(0,nrow(data_working))
data_working$ucl95 <- rep(0,nrow(data_working))
data_working$lcl90 <- rep(0,nrow(data_working))
data_working$ucl90 <- rep(0,nrow(data_working))

for (i in 1:nrow(data_working)) {

z <- (data_working$vpa_z[i]-data_working$pred_z[i])/sd_res
data_working$sd_pred_z[i] <- sd_res
data_working$p_artifact[i] <- 1-(2*(1-pnorm(abs(z))))
data_working$lclmax[i] <- data_working$pred_z[i]-(max_z*sd_res)
data_working$lcl99[i] <- data_working$pred_z[i]-(2.58*sd_res)
data_working$lcl95[i] <- data_working$pred_z[i]-(1.96*sd_res)
data_working$lcl90[i] <- data_working$pred_z[i]-(1.64*sd_res)
data_working$uclmax[i] <- data_working$pred_z[i]+(max_z*sd_res)

```

```

data_working$ucl99[i] <- data_working$pred_z[i]+(2.58*sd_res)
data_working$ucl95[i] <- data_working$pred_z[i]+(1.96*sd_res)
data_working$ucl90[i] <- data_working$pred_z[i]+(1.64*sd_res)

}

max_p <- 1-(2*(1-pnorm(max_z)))

data_working_artifact_rows <- data_working[data_working$p_artifact > max_p,]

artifact_vector <- append(artifact_vector, data_working_artifact_rows$row_id, after =
length(artifact_vector))

artifact_vector_length_current <- length(artifact_vector)

if (artifact_vector_length_current == artifact_vector_length_prior) { break }

data_working$artifact <- (data_working$p_artifact > max_p)*1

}

##### Just visually setting off the end of the major repeat loop #####

#####

#####

#####

# The following does a bit of post-processing on the artifact vector created above.

artifact_vector # This shows the rows in the original data that are artifacts.

artifact_vector <- sort(artifact_vector) # Sorts the artifact vector.

newrows_vector <- NULL # Initiates a new vector into which additional artifact rows will be written.

for (i in 1:(length(artifact_vector)-1)) {

  if (artifact_vector[i+1] - artifact_vector[i] == 2) {newrows_vector <- append(newrows_vector,
c(artifact_vector[i]+1))} else {}
  if (artifact_vector[i+1] - artifact_vector[i] == 3) {newrows_vector <- append(newrows_vector,
c(artifact_vector[i]+1, artifact_vector[i]+2), after=length(newrows_vector))} else {}

}

# The loop above identifies additional rows as artifacts – i.e., any rows in between two artifact rows
separated by three heartbeats or less.

artifact_vector <- append(artifact_vector, newrows_vector, after=length(artifact_vector)) # Adds the
additional rows to the artifact vector.

```

```

artifact_vector <- sort(artifact_vector) # Sorts the final artifact vector.

#####

#####

#####

# The following writes out all of this into files.

phase1outputcopy$artifact <- rep(0,nrow(phase1outputcopy)) # Initializes the variable.

for (i in 1:length(artifact_vector)) {
  targetrow <- artifact_vector[i]
  phase1outputcopy$artifact[targetrow] <- 1
}

phase1outputcopy$outofphase <- rep(0,nrow(phase1outputcopy)) # Initializes the variable.

for (i in 1:length(outofphasevector)) {
  targetrow <- outofphasevector[i]
  phase1outputcopy$outofphase[targetrow] <- 1
}

write.csv(phase1outputcopy, prepruned_heartbeat_VPA_output_filename, row.names=FALSE)

# The above writes the pre-pruned data with the flag for whether each row wound up being an artifact.

plot(density(data_working$p_artifact))

# This shows the density of the p-values for the pruned data.
# If the above does not look close to uniform, there are some outliers.

plot(data_working$time,data_working$vpa_z, type="p", pch=20, cex=0.6, xlab="time", ylab="vpa_z",
main="Pruned Data and Confidence Limits")
lines(data_working$time,data_working$pred_z, type="l", col="black", lwd=4)
lines(data_working$time,data_working$lclmax, type="l", col="red", lwd=2)
lines(data_working$time,data_working$uclmax, type="l", col="red", lwd=2)
lines(data_working$time,data_working$lcl99, type="l", col="orange", lwd=2)
lines(data_working$time,data_working$ucl99, type="l", col="orange", lwd=2)
lines(data_working$time,data_working$lcl95, type="l", col="yellow", lwd=2)
lines(data_working$time,data_working$ucl95, type="l", col="yellow", lwd=2)
lines(data_working$time,data_working$lcl90, type="l", col="green", lwd=2)
lines(data_working$time,data_working$ucl90, type="l", col="green", lwd=2)
legend(locator(1), c("90%ci", "95%ci","99%ci","maxci"), col=c("green", "yellow", "orange", "red"),
lty=c(1,1,1,1), lwd=c(2,2,2,2))

# This graph shows the smooth function and confidence limits for the pruned data.

```