IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received December 27, 2018, accepted February 8, 2019, date of publication February 14, 2019, date of current version March 4, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2899485

Analysis of Reproducibility of Noninvasive Measures of Sympathetic Autonomic Control Based on Electrodermal Activity and Heart Rate Variability

HUGO F. POSADA-QUINTERO[1], (Member, IEEE), TANYA DIMITROV[1], AURELIE MOUTRAN[1], SEOHYOUNG PARK[2], AND KI H. CHON[1], (Member, IEEE)

[1]Department of Biomedical Engineering, University of Connecticut, Storrs, CT 06269, USA
[2]Northern Highlands Regional High School, Allendale, NJ 07401, USA

Corresponding author: Hugo F. Posada-Quintero (h.posada@engr.uconn.edu)

This work was supported by the Office of Naval Research Work Unit under Grant N00014-15-1-2236.

ABSTRACT Heart rate variability (HRV) and electrodermal activity (EDA) are useful tools for assessing the central and peripheral dynamics of the sympathetic nervous system and detecting the effects of numerous systemic diseases and life-challenging situations. However, the indices of HRV and EDA are highly influenced by mental stress, environmental conditions, body position, and other physiological conditions that introduce variability. In this paper, we assessed the five-day reproducibility of HRV and EDA measures of sympathetic control, for N = 20 subjects undergoing 70° head-up tilt test (HUT) and Stroop task tests. We made the assessment in highly controlled conditions without environmental causes of variability, to have a good baseline understanding of the consistency of the various indices of HRV and EDA. Therefore, we assessed intra-subject variation (using the coefficient of variation, CV) and consistency (using the intra-class correlation coefficient, ICC) of the test-to-baseline differences produced by both tests on the studied measures. The low-frequency component of HRV (HRVLF), and its normalized variant was computed as HRV measures of sympathetic control. For EDA, the skin conductance level, frequency of non-specific skin conductance responses, spectral index (EDASympn), and time-varying index (TVSymp) were computed. TVSymp (ICC = 0.85) and HRV indices exhibited higher consistency during the HUT (ICC $\geq$ 0.8), compared to other EDA measures, and HRVLF was the least variable measure (CV = 85.4%). EDA indices exhibited higher consistency (except for the EDASympn) during the Stroop task (ICC $\geq$ 0.79) when compared to HRV, and TVSymp was the least variable measure (CV = 97.2%). Remarkably, TVSymp proved to be a reproducible measurement (low variation and high consistency) in both scenarios. These results are the necessary groundwork for studying the use of EDA and HRV in real-world conditions, as reproducibility of the indices has remarkable importance in clinical practice.

INDEX TERMS Autonomic nervous system, electrodermal activity, heart rate variability, reproducibility, Stroop task, tilt table test.

I. INTRODUCTION

Wearable sensors and diagnostic systems are a promising advancement for the assessment of systemic diseases (diabetes, hypertension, heart failure, and so on) and other life-challenging conditions (stress, drowsiness, and so on). Those conditions can be potentially assessed using data from the central and peripheral modulations of the sympathetic branch of the autonomic nervous system (ANS) [1]–[7]. There are two main tools available for noninvasive assessment of the dynamics of the ANS: the heart rate variability (HRV) [8] and the electrodermal activity (EDA) [9], [10]. However, variables such as environmental conditions, body position, and other physiological functions of the subjects may lessen

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaodong Yang.

the reliability of such techniques for the assessment of the ANS [11], [12]. Repeatability is especially important for wearable technologies, because their readings are provided continuously to the subject, and the consistency of the measurements will greatly affect adoption rates [13]–[15]. Hence, assessing the reproducibility of HRV and EDA is a necessary research task.

Traditional techniques for assessing sympathetic function involve hemodynamic measurements [16], pharmacological blockade [17], noradrenaline measurement [18], microneurography [19], plasma noradrenaline kinetics [20], imaging techniques [21], HRV analysis [8], [22], or measurement of sweat production (e.g. EDA)[23]. Among those techniques, only HRV and EDA can provide continuous, non-invasive and affordable means for sympathetic function assessment. Because of the simplicity of the technique, HRV analysis is the most widely employed technique to evaluate sympathetic function. HRV is a non-invasive indicator that reflects the homeostatic interplay between perturbations in central (cardiovascular) functions and the dynamic responses of the cardiovascular regulatory systems [8]. The low-frequency components of HRV (0.04-0.15 Hz) are commonly used as a marker of sympathetic control, even though this frequency band is also known to be influenced by the parasympathetic nervous system. Hence, the LF band is not an accurate measure of the sympathetic dynamics.

In clinical practice, reliable techniques for the assessment of sympathetic tone of the ANS are needed because of the prevalence of autonomic balance impairment in certain cardiovascular diseases and pathophysiological conditions [24], [25]. A widely used noninvasive means to assess the dynamics of the ANS is to compute the power spectral density of HRV [8]. The high-frequency components of HRV are known to be solely influenced by the parasympathetic system. In contrast, the low-frequency (0.045-0.15 Hz) components of HRV (HRVLF or HRVLFn when normalized to total power of HRV) are influenced by both the sympathetic and parasympathetic nervous systems. Results concerning the reproducibility of HRV both in the time and frequency domain are controversial. Some found poor reproducibility [26], [27], but others found the opposite result [28]–[30]. Overall, higher reproducibility has been found for spectral parameters, when compared to temporal parameters [26], [31]. General reproducibility can be achieved using controlled breathing [1]. Nevertheless, the reproducibility of the spectral indices of HRV is still a matter of debate [32], [33].

EDA is being increasingly used as a measure of the sympathetic function [10]. One of the advantages of EDA is that sudomotor activity is known to be solely controlled by the sympathetic nervous system [34]–[36]. Traditionally, analysis of EDA has been in the time-domain [9], using skin conductance level (SCL) and nonspecific skin conductance responses (NS.SCRs). However, several studies have reported low reproducibility of these time-domain indices [11], [37]. Time-invariant and time-variant spectral analysis of electrodermal activity have recently been reported as tools for sympathetic tone assessment [37], [38]. The resulting indices, EDASymp and TVSymp, demonstrated lower intra-subject variability compared to time-domain measures of EDA, and higher consistency and sensitivity to orthostatic and cognitive stress compared to SCL and NS.SCRs.

There have been many reports questioning the consistency of EDA data even during short duration recordings with minimal motion artifacts [9], [11], [37], [39], [40]. Despite this prevailing skepticism, no study has ever analyzed the consistency and reliability of short-term EDA signals in an extremely controlled environment, which is the logical first step. Hence, our study specifically deals with this problem, as we rely on highly-controlled signals to compute the indices on which we evaluate the consistency and repeatability of EDA data. For this study, we explored the intra-subject repeatability of HRV and EDA measures, for healthy subjects performing the 70° head-up tilt (HUT) test and the Stroop task. Data were collected from the same group of subjects over five different days, without breathing control.

II. MATERIALS AND METHODS

A. PROTOCOL

Twenty healthy volunteers (11 males, 9 females) of ages 22 ± 5.6 years old (mean ± standard deviation), weight 70 ± 9.7 kg, and height 173.5 ± 7 cm, were enrolled in this study. Participants were asked to avoid caffeine and alcohol during the 24 hours preceding the first test and at least five hours before each subsequent test. The study was conducted in a quiet, dimly lighted room (ambient temperature, 26-27 °C). Before each sub-test, the subjects were asked to stay still in the supine position for 5 minutes to procure hemodynamic stabilization. ECG and EDA data were recorded simultaneously for each subject. An HP ECG monitor (HP 78354A) and an EDA ADInstruments module were used, respectively. Hydrogel Ag-AgCl electrodes were employed for ECG signal collection. For the EDA, a pair of stainless-steel electrodes were placed on index and middle fingers. Subjects' skin was prepared with alcohol before placing the ECG and EDA electrodes. All signals were recorded at a sampling frequency of 400 Hz.

Participants were asked to put the three ECG electrodes on themselves. The first electrode was placed on the inside of the right wrist, the second electrode was placed on the inside of the left wrist, and the third electrode was placed on the left lower rib. The EDA electrodes were placed on the index and middle fingers of each subject's right hand. Every day, subjects underwent two sub-tests: 70° head-up tilt (HUT) and the Stroop task. For all subjects, the order of tests was always HUT first, then the Stroop task. These tests are described below.

1) 70° HEAD-UP TILT (HUT) TEST

The HUT is a simple standardized test. A procedure similar to the one described in [41, p.] was used. For baseline measurements, the subjects were asked to lie down on the

table and were then strapped to it (horizontal is defined as 0°). We allowed 5 minutes for hemodynamic stabilization, prior to starting the ECG and EDA signal recording that was continuous throughout the test. The subjects stayed in the supine position for 5 minutes, to collect baseline data. Afterwards, the table was tilted to an angle of 70°, with the subject's head up. The subject remained in that position for 5 minutes.

2) STROOP TASK

The Stroop task is a simple test widely used for measuring cognitive stress. A procedure similar to the one documented in [42] was used. After baseline readings, a five-minute Stroop test was conducted. For baseline measurements, subjects were asked to lie down on the tilt table described above. Subjects were in the supine position (0°) throughout the entire experiment. The subjects were asked to remain in the supine position, with their eyes closed for five minutes, while we collected baseline data. After that, subjects were asked to speak the color of the ink of a word presented to them which named a color. They were shown congruent (the word was written in the color it expressed) and incongruent (the word and the color it was printed in were different) combinations to induce cognitive stress (Stroop effect) [42]. The words and colors were "blue," "yellow," "green," "red," "purple" and "black". The background also changed to be randomly congruently or incongruently colored with the word. A computerized version of the original Stroop task was designed. The Stroop task was 5 minutes total. ECG and EDA signals were recorded throughout the test. On day one, subjects underwent a one-minute training version of the Stroop task.

B. PHYSIOLOGICAL INDICES OF THE AUTONOMIC NERVOUS SYSTEM

Measures to assess the sympathetic function of the ANS based on HRV and EDA were computed using data collected during the five days of testing for each subject undergoing the HUT and Stroop task. We chose the indices of HRV and EDA based on the recommendations of the standards of measurement, physiological interpretation, and clinical use of HRV [8] and the published recommendations for EDA measurements [9]. Furthermore, we included indices of EDA recently developed in our lab [37], [38]. To ensure the quality of physiological data, subjects were asked to stay still while performing the tests, with no movement other than the tilting of the table.

The duration of the data segments was chosen based on the minimum duration of the data considered to be usable [8], [9]. In other words, if data are collected continuously using a wearable device, a data segment of at least 4 minutes for ECG or 2 minutes for EDA needs to be collected (for example, based on an implemented signal-quality index), so the index of sympathetic control can be computed. The assumption is that although data collection on wearable devices can be corrupted (e.g. by motion) most of the time, there will be segments of at least such duration available for computing the indices. Tools for the assessment of the quality of ambulatory EDA data have recently been published [43].

1) INDICES OF ELECTRODERMAL ACTIVITY

The EDA data were inspected for spikes introduced by motion artifacts. All identified events were removed by connecting the end points of the motion artifact spike using cubic spline. A median filter (1-second width) was also applied to remove noise. For each trial, two minutes of clean EDA signal were extracted during baseline rest and for each of the two tests (HUT and Stroop task), to compute the measures of EDA. From baseline recordings, two complete minutes of data ending 30 seconds before the tests started were used to assure the most stabilized signal. From the HUT and Stroop task, the two-minute EDA data segment was extracted starting 30 seconds after the subject started the test, to avoid any distortion caused by the transition from baseline to test. Measures of EDA were computed by analyzing the data in the time and frequency domains.

In the time domain, the EDA signal was decomposed into tonic and phasic components, using the convex optimization approach [44]. The skin conductance level (SCL, expressed in microsiemens, μS), an index related to the slow shifts of EDA, was computed as the mean value of the tonic component of EDA taken during a two-minute period [9]. The frequency of non-specific skin conductance responses (SCRs), termed NS.SCRs, an index related to fast changes of EDA, was measured as the number of SCRs whose amplitudes were higher than a given threshold (0.05 μS, in this study), per minute [9].

The time-invariant spectra of EDA were calculated using Welch's periodogram method with 50% data overlap. A Blackman window (length of 128 points) was applied to each segment, the fast Fourier transform was calculated for each windowed segment, and the power spectra of the segments were averaged. The power spectral index of EDA, EDASymp [μS2], was computed by integrating the power in the range from 0.045 to 0.25 Hz, as such range was previously found to be sensitive to cognitive stress [37].

To compute the time-varying index of EDA, TVSymp, the time-frequency spectra of EDA data were computed using variable frequency complex demodulation (VFCDM), a time-frequency spectral analysis technique that provides accurate amplitude estimates and one of the highest time-frequency resolutions [45]. The components comprising the frequency power in the range from 0.08 to 0.24 Hz were used to compute TVSymp, as defined in a previous study [38]. Amplitudes of the time-varying components in this band are summed together to obtain an estimated reconstructed EDA signal ($X'(t)$), which is then normalized to unit variance, and its instantaneous amplitude is computed using the Hilbert transform [46], as follows

$$Y'(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} X'(\tau) \big/ (t - \tau) d\tau \tag{1}$$

where P indicates the Cauchy principal value. $X'(t)$ and $Y'(t)$ form the complex conjugate pair, so we can define an analytic signal, $Z(t)$, as

$$Z(t) = X'(t) + iY'(t) = a(t)e^{j\theta(t)} \tag{2}$$

in which

$$a(t) = \left[X'^2(t) + Y'^2(t)\right]^{1/2} \tag{3}$$

$$\theta(t) = arctan\left(Y'(t)/X'(t)\right). \tag{4}$$

The resulting $a(t)$ is considered the instantaneous amplitude of $Z(t)$. TVSymp is computed as the mean amplitude of $a(t)$. Notice that TVSymp is a dimensionless quantity as it was normalized to the standard deviation.

2) INDICES OF HEART RATE VARIABILITY

For each trial of the two tests (HUT and Stroop task), four minutes of clean ECG signals during baseline and test were used to compute HRV indices. ECG signals were band-pass filtered (0.05-40 Hz) to reduce noise and motion artifacts. The R-waveform peaks were detected using the detection algorithm that defines a delineation function based on the envelope of the ECG signal [47], [48]. In addition, all segments were visually inspected to ensure that no R peak (beat) was missed. The R-R interval series were converted to an evenly time-sampled signal (4 Hz) by cubic spline interpolation. Using the heart rate (HR) time series for each trial, the power spectra of HRV were then calculated using Welch's periodogram with 50% data overlap. A Blackman window (length of 256 points) was applied to each segment and the fast Fourier transform was calculated for each windowed segment. Finally, the power spectra of the segments were averaged. The low-frequency index (HRVLF [ms2], 0.045 to 0.15 Hz), and normalized version of it (HRVLFn = HRVLF/Total power of HRV) were computed [8]. Indices from the low frequency range of HRV (HRVLF and HRVLFn) are widely used as indices of sympathetic tone [8]. Indices from the high frequency power (0.15 to 0.4 Hz), termed HRVHF an d HRVHFn, are used as indices of parasympathetic function, and were not considered in this study, as we focused on measurements of the sympathetic control in the ANS.

C. STATISTICS

The set of measures of sympathetic control based on HRV and EDA computed in this study is: HRVLF, HRVLFn, SCL, NS.SCRs, EDASymp and TVSymp. For the HUT and Stroop task, ten measurements of each index were collected (i.e. baseline and test measurement for each of the five days). Two-way repeated measures analysis was performed to test the consistency of the test-to-baseline differences during the five days (defined as the first factor), and the individual consistency of the measures (defined as the second factor). Normality of the measurements through the five days was tested using the Kolmogorov-Smirnov test [49]–[51]. For repeated measures analysis in normally-distributed data, the two-way analysis of variance (ANOVA) was performed to test for significant differences between measures. If non-normality was found in a specific index, the Friedman test was used [52]. The Bonferroni method was used for correction of multiple comparisons.

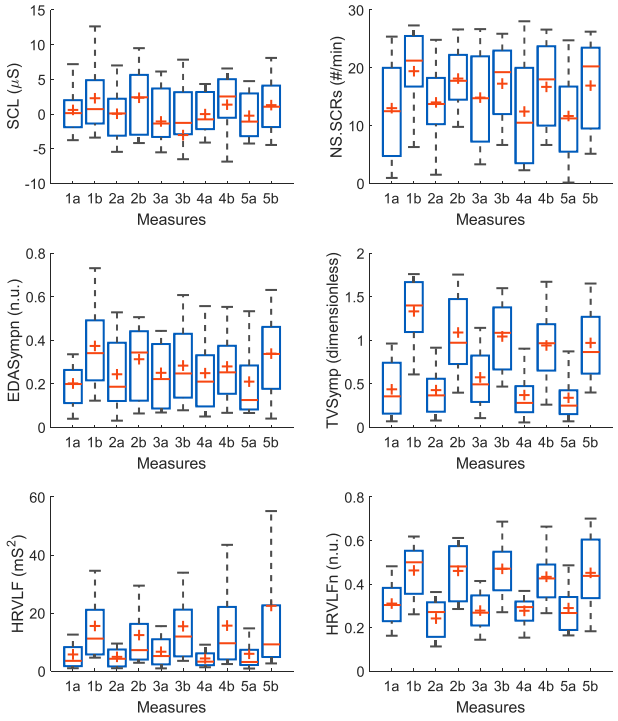

FIGURE 1. Baseline and HUT measures of HRV and EDA for the five days, for all subjects. The letter "a" represents baseline measurements, and "b" represents test measurements.

The difference between test and baseline measures was computed every day for the HUT and Stroop task. The analysis of reproducibility of such differences was carried out using the computed test-baseline differences. The intra-subject five-day coefficient of variation (CV) (i.e. the standard deviation divided by the mean) of each measure for every subject was computed. The mean and standard deviation of the intra-subject CV values (for the N = 20 subjects) was calculated to assess the overall level of variation of each of the eight measures. The intra-class correlation (ICC) was computed for each measure to estimate its degree of consistency [53], for the N = 20 independent subjects, using the five day measures. For the interpretation of ICC, an ICC lower than 0.4 was considered poor consistency, 0.4 to 0.75 represented good consistency, and >0.75 represented excellent consistency beyond chance [54].

III. RESULTS

Figs. 1 and 2 are the box plots for baseline and test measurements for all the HRV and EDA measures computed, for the H UT and Stroop task, respectively, during the five days of testing, for all subjects. Table 1 includes the significant differences found between HUT and baseline measures. For the

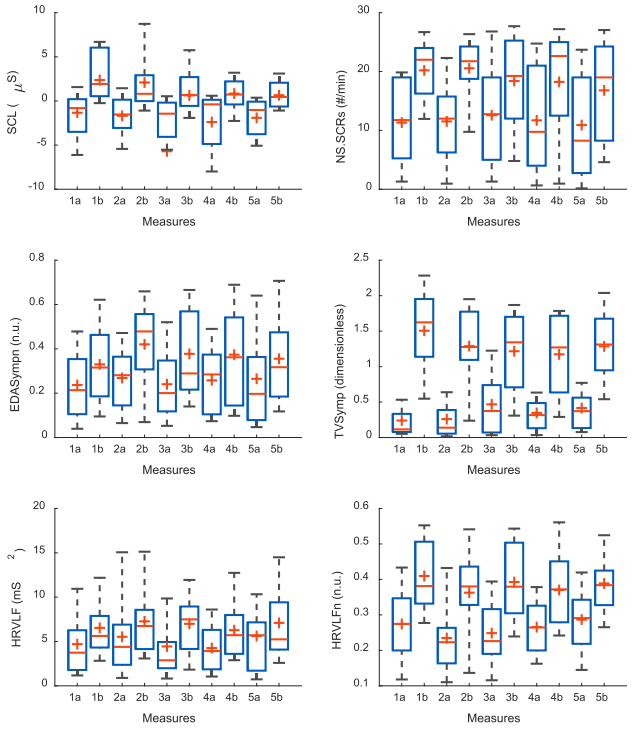

FIGURE 2. Baseline and Stroop task measures of HRV and EDA for the five days. The letter "a" represents baseline measurements, and "b" represents test measurements.

TABLE 1. Significance of test-to-baseline differences in EDA and HRV indices for the HUT and stroop task during the five day experiment, for all subjects.

Tilt Table

TVSymp

Baseline \ HUT	Day 1	Day 2	Day 3	Day 4	Day 5
Day 1	*	*	*	*	*
Day 2	*	*	*	*	*
Day 3	*	*	*		
Day 4	*	*	*	*	*
Day 5	*	*	*	*	*

HRVLF

Baseline \ HUT	Day 1	Day 2	Day 3	Day 4	Day 5
Day 1	*		*	*	*
Day 2	*	*	*	*	*
Day 3	*		*	*	*
Day 4	*	*	*	*	*
Day 5	*		*	*	*

HRVLFn

Baseline \ HUT	Day 1	Day 2	Day 3	Day 4	Day 5
Day 1	*	*	*		*
Day 2	*	*	*	*	*
Day 3	*	*	*	*	*
Day 4	*	*	*	*	*
Day 5	*	*	*	*	*

Stroop task

SCL

Baseline \ HUT	Day 1	Day 2	Day 3	Day 4	Day 5
Day 1					
Day 2	*	*			*
Day 3	*	*			
Day 4	*	*			*
Day 5	*	*			

NS.SCRs

Baseline \ HUT	Day 1	Day 2	Day 3	Day 4	Day 5
Day 1	*	*			
Day 2	*	*			
Day 3					
Day 4	*	*			
Day 5	*	*			

TVSymp

Baseline \ HUT	Day 1	Day 2	Day 3	Day 4	Day 5
Day 1	*	*	*	*	*
Day 2	*	*	*	*	*
Day 3	*	*	*	*	*
Day 4	*	*	*	*	*
Day 5	*	*	*	*	*

HRVLFn

Baseline \ HUT	Day 1	Day 2	Day 3	Day 4	Day 5
Day 1	*		*		
Day 2	*	*	*	*	*
Day 3	*		*	*	*
Day 4	*		*		*
Day 5	*				

* respresents statistically significant difference found in the repeated measures analysis

HUT test, the repeated measures analysis found that test-to-baseline differences in SCL, NS.SCRs, and EDASympn were not significantly different. TVSymp, HRVLFn and HRVLF were significantly different in most HUT measures, compared to baseline measures. Day-2 HUT measures of HRVLF were only different to the baseline of days 2 and 4. As for the second factor (individual consistency), all indices were significantly affected by the inter-subject variability of the HUT ($p < 0.05$).

For the Stroop task, only TVSymp exhibited significant differences between all test measures and all baseline measures (Table 1). SCL and NS.SCRs showed many differences between test and baseline measures, although some test measures were too low (e.g. SCL on day 3, NS.SCRs on days 4 and 5). EDASympn and HRVLF were not found to be significantly different between test and baseline measures in the repeated measures analysis. Significant differences were found between test and baseline measures in HRVLFn, for most of the days. The second factor (individual consistency) was statistically significant for all indices, excluding TVSymp ($p = 0.64$). No baseline-to-baseline or test-to-test differences were found in any index.

Results from the analysis of reproducibility of test-to-baseline difference indices of HRV and EDA, for both the HUT and Stroop task, are shown in Tables 2 and 3, respectively. For the HUT, the highest variation (assessed by CV) was exhibited by NS.SCRs (634%), followed by EDASympn (498%), SCL (413%), and HRVLFn (182%). TVSymp exhibited a variation of less than 200%, and HRVLF was the only measure with a CV lower than 100%. As for the consistency (assessed by ICC), HRVLF exhibited the highest consistency (0.83), followed by TVSymp (0.75), and HRVLFn (0.73); only these measures exhibited excellent consistency beyond chance. SCL and NS.SCRs exhibited good consistency (0.46 and 0.63, respectively), and EDASympn exhibited poor consistency (<0.4).

For the Stroop task, the highest variation was observed in HRVLF (424%), HRVLFn (371%), SCL (280%), and EDASympn (264%). Only TVSymp and NS.SCRs exhibited a CV lower than 200%, and TVSymp was the least variable index overall. Based on the ICC, HRVLFn exhibited the highest consistency (0.7), followed by TVSymp (0.66), NS.SCRs (0.64), and SCL (0.42). Those values are all considered good consistency. HRVLF (0.38) and EDASympn (0.39) exhibited poor consistency.

TABLE 2. Results for test-to-baseline differences of EDA and HRV indices for the HUT, for all subjects.

	Day 1	Day 2	Day 3	Day 4	Day 5	CV (%) (LB UB)	ICC (LB UB)
SCL (μS)	1.7±5.5	2.3±3.8	-2±21	1.4±4.3	1.5±2.8	413 (379 447)	0.46 (-0.025 0.76)
NS.SCRs (#/min)	6.4±5.5	4.1±6.8	2.4±6	4.3±6.5	5.3±7.6	634 (581 686)	0.63 (0.29 0.83)
EDASympn (n.u.)	0.17±0.27	0.07±0.28	0.03±0.25	0.03±0.21	0.13±0.24	498 (457 539)	0.38 (-0.19 0.72)
TVSymp	0.89±0.54	0.66±0.7	0.47±0.74	0.57±0.6	0.63±0.6	146 (134 158)	0.75 (0.52 0.89)
HRVLF (ms^2)	9.8±12	7.5±10	8.8±10	11±13	17±37	81.8 (75.1 88.6)	0.83 (0.68 0.92)
HRVLFn (n.u.)	0.15±0.15	0.22±0.17	0.19±0.15	0.16±0.15	0.16±0.19	182 (167 197)	0.73 (0.49 0.88)

Values are mean ± standard deviation

SCL, skin conductance level; NS.SCRs, non-specific skin conductance responses; EDASympn, normalized sympathetic component of the EDA; TVSymp, Time-varying index of sympathetic tone; HRVLF, low frequency components of HRV; HRVLFn, normalized low frequency components of HRV; CV, Coefficient of variation; LB and UB are upper and lower bounds of the confidence interval of CV and ICC with a level of significance of 0.05.

TABLE 3. Results for test-to-baseline differences of EDA and HRV indices for the Stroop task test, for all subjects.

	Day 1	Day 2	Day 3	Day 4	Day 5	CV (%) (LB UB)	ICC (LB UB)
SCL (μS)	3.7±3.7	3.8±4.1	6.3±13	3.2±3.9	2.6±3	280 (257 303)	0.42 (-0.11 0.74)
NS.SCRs (#/min)	8.9±8.3	9±6.7	5.8±10	6.5±8.6	5.9±6.8	133 (122 144)	0.64 (0.32 0.84)
EDASympn (n.u.)	0.09±0.24	0.15±0.2	0.14±0.27	0.12±0.25	0.091±0.27	264 (243 286)	0.39 (-0.16 0.73)
TVSymp	1.3±0.72	1±0.67	0.75±0.89	0.83±0.68	0.87±0.66	107 (97.9 115)	0.66 (0.34 0.85)
HRVLF (ms^2)	1.8±3.6	1.7±6	2.6±4.8	2±3.2	1.5±5.3	424 (389 459)	0.38 (-0.18 0.72)
HRVLFn (n.u.)	0.14±0.11	0.13±0.15	0.14±0.13	0.1±0.11	0.1±0.14	371 (341 401)	0.7 (0.42 0.86)

Values are mean ± standard deviation

SCL, skin conductance level; NS.SCRs, non-specific skin conductance responses; EDASympn, normalized sympathetic component of the EDA; TVSymp, Time-varying index of sympathetic tone; HRVLF, low frequency components of HRV; HRVLFn, normalized low frequency components of HRV; CV, Coefficient of variation; LB and UB are upper and lower bounds of the ICC with a level of significance of 0.05.

IV. DISCUSSION

In this study, we evaluated the five-day reproducibility of HRV and EDA measures of sympathetic control, for subjects undergoing HUT and Stroop task tests. We aimed to elucidate the question of which measure of sympathetic control available from HRV and EDA analyses is more reproducible in the presence of postural and cognitive stress. For that, we assessed intra-subject variation and consistency of the test-to-baseline differences produced by both tests on the studied measures. The HRV indices of sympathetic control computed in this study (HRVLF and HRVLFn) exhibited higher consistency in the HUT, compared to the Stroop task. This suggests that postural stimulation produces more reproducible changes in the central level (HRV is caused by the sympathetic modulation on the heart). Three of the computed EDA indices (SCL, NS.SCRs, and EDASympn) exhibited overall low consistency for the HUT and the Stroop task. Nevertheless, TVSymp, a measure resulting from the time-varying spectral analysis of EDA, exhibited low variability compared to all other indices, and good-to-high consistency for both tests, which makes it the most reproducible measure overall. Interestingly, the effect of the Stroop task on TVSymp is consistent among subjects, as no individual effect was found in the two-way analysis.

Results of the repeated measures analysis were consistent with these findings. Test-to-baseline differences were more consistent in HRV measures for the HUT test, and in EDA measures for the Stroop task. Beyond the consistency analysis, there is another observation possible in Figs. 1 and 2. The TVSymp captures a ''learning effect'' for both the HUT and Stroop task tests. Apparently, test measures on day 1 were particularly high compared to other days (although not statistically significant). Day 2 through day 5 measures seem more stable, like a plateau. It suggests an extra stress on day 1 caused by the expectation and/or ignorance of the subjects about the effects and difficulty of the tasks. This variation is also captured by the SCL for the Stroop task.

Sympathetic tone normally increases with postural stimulation [55]. The HUT has been previously used to elicit sympathetic activation, and HRVLF components have been shown to be sensitive to such a test [56], [57]. In this study, we found measures computed using such components to be highly reproducible in response to the HUT. This suggests that the central autonomic mechanisms elicited by HUT produce highly consistent effects on the low-frequency innervations of the heart. Likewise, some previous studies have found a significant increase in HRVLF components in response to the Stroop task [58], [59]. The opposite results have also been reported [38]. In this study, we found those indices were

highly variable intra- and inter-subject, and their consistency was only fair.

The EDA dynamics exhibit both tonic and phasic changes, regulated by sympathetic innervation of the sweat glands. Variations in EDA are a product of the innervation of sweat glands that result in changing levels of sweat in the ducts [60]. Functionally, EDA is associated with central mechanisms that play different roles, including gross movements, thermoregulatory sweating, affective processes, orientation and attention, and fine control [61], [62]. Although sweat glands make up part of the sympathetic-cholinergic system and were thought to exclusively respond to peripheral stimulus (i.e. thermoregulatory sweating), the electrodermal response is inhibited in response to pharmacological central depressants in a manner analogous to its action on other sympathetic systems [63], [64]. This has led some researchers to conclude that a central adrenergic inhibitory mechanism is involved in the regulation of the electrodermal activity [64], [65]. In general, the time-domain measures (SCL and NS.SCRs) are known to be consistent with sympathetic arousal, as they are elevated by administration of dextroamphetamine, caffeine, and threatening situations [66], [67]. They also exhibited relatively low within-subject variability in a study looking at the test-retest (one repetition) correlation, but high vari ability between subjects [11]. This long-running concern about the variability of EDA has impeded the widespread use of these indices for assessing the state of activation of the sympathetic system. Our results show that a robust index of EDA based on the time-frequency spectrum (TVSymp) is more consistent and less variable than traditional measures of EDA.

Very few studies have looked at the response of EDA to the HUT task. A study reported an increase in the SCL in the tilt-negative group [68]; another study found significant differences in time-domain and spectral indices of EDA when subjects were tilted, compared to the supine position [38]. In this study, we found poor consistency of SCL in response to HUT. NS.SCRs exhibited good consistency, but were highly variable. As for the spectral indices, TVSymp was the most consistent measure in response to HUT, with moderate variations intra- and inter-subject. The sensitivity of measures of EDA to the Stroop task have been shown before [37], [38]. In this study, we found that the changes in SCL, NS.SCRs and TVSymp are highly consistent in response to such a test. EDASympn exhibited high variability and poor consistency for both the HUT and Stroop task.

V. CONCLUSION

This study evaluated the reliability of non-invasive sympathetic indices derived from EDA data. In this work, we examined the reliability and consistency of short-duration EDA during controlled conditions with minimal artifacts. This was motivated by the lack of studies even addressing the consistency and reliability of short-term EDA signals during controlled conditions. We believe that lack of faith in the consistency of EDA data is one of the primary reasons why the use of EDA has not yet gained wide acceptance by the research community. Analysis on long-duration EDA data with motion artifacts is the logical next step which requires advanced algorithms for detecting and removing motion artifacts.

We found that the autonomic sympathetic response to postural stimulation is more reproducible in the HRV, compared to the EDA. Despite their sensitivity, most EDA indices exhibited low reproducibility in response to postural and cognitive stress. However, the measure resulting from the time-varying analysis of EDA, TVSymp, was found to be a reproducible measure in both scenarios. For practical applications using wearable technologies, where reproducibility of the indices has remarkable importance, TVSymp along with measures of HRV can be obtained to produce more reliable tests for the evaluation of sympathetic function, and possibly enable the assessment of the progression of systemic diseases that affect the sympathetic autonomic response in humans.

REFERENCES

[1] A. J. Burger, M. Charlamb, L. A. Weinrauch, and J. A. D'Elia, "Short- and long-term reproducibility of heart rate variability in patients with long-standing type I diabetes mellitus," *Amer. J. Cardiol.*, vol. 80, no. 9, pp. 1198–1202, Nov. 1997.

[2] M. Glos, I. Fietze, A. Blau, G. Baumann, and T. Penzel, "Cardiac autonomic modulation and sleepiness: Physiological consequences of sleep deprivation due to 40 h of prolonged wakefulness," *Physiol. Behav.*, vol. 125, pp. 45–53, Feb. 2014.

[3] A. O. Konrady, O. G. Rudomanov, O. I. Yacovleva, and E. V. Shlyakhto, "Power spectral components of heart rate variability in different types of cardiac remodelling in hypertensive patients," *Med. Sci. Monitor*, vol. 7, no. 1, pp. 58–63, Feb. 2001.

[4] H. F. Posada-Quintero, J. B. Bolkhovsky, N. Reljin, and K. H. Chon, "Sleep deprivation in young and healthy subjects is more sensitively identified by higher frequencies of electrodermal activity than by skin conductance level evaluated in the time domain," *Frontiers Physiol.*, vol. 8, p. 409, Jun. 2017.

[5] S. Scalvini *et al.*, "Is heart rate variability a reliable method to assess autonomic modulation in left ventricular dysfunction and heart failure?: Assessment of autonomic modulation with heart rate variability," *Int. J. Cardiol.*, vol. 67, no. 1, pp. 9–17, Nov. 1998.

[6] J. Vicente, P. Laguna, A. Bartra, and R. Bailon, "Drowsiness detection using heart rate variability," *Med. Biol. Eng. Comput.*, vol. 54, no. 6, pp. 927–937, 2016.

[7] F. Weber, H. Schneider, T. von Arnim, and W. Urbaszek, "Heart rate variability and ischaemia in patients with coronary heart disease and stable angina pectoris: Influence of drug therapy and prognostic value," *Eur. Heart J.*, vol. 20, no. 1, pp. 38–50, Jan. 1999.

[8] Task force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology, "Heart rate variability. standards of measurement, physiological interpretation, and clinical use," *Eur. Heart J.*, vol. 17, no. 3, pp. 354–381, 1996.

[9] W. T. Roth, M. E. Dawson, and D. L. Filion, "Publication recommendations for electrodermal measurements," *Psychophysiology*, vol. 49, no. 8, pp. 1017–1034, Aug. 2012.

[10] R. Freeman and M. W. Chapleau, "Testing the autonomic nervous system," *Handbook Clin. Neurol.*, vol. 115, pp. 115–136, Jan. 2013.

[11] A. Crider and R. Lunn, "Electrodermal lability as a personality dimension," *J. Exp. Res. Personality*, vol. 5, no. 2, pp. 145–150, 1971.

[12] D. J. Ewing, J. M. Neilson, C. M. Shapiro, J. A. Stewart, and W. Reid, "Twenty four hour heart rate variability: Effects of posture, sleep, and time of day in healthy controls and comparison with bedside tests of autonomic function in diabetic patients," *Heart*, vol. 65, no. 5, pp. 239–244, May 1991.

[13] C. R. Reid *et al.*, "Wearable technologies: How will we overcome barriers to enhance worker performance, health, and safety?" *Proc. Hum. Factors Ergon. Soc. Annu. Meeting*, vol. 61, no. 1, pp. 1026–1030, Sep. 2017.

[14] M. C. Schall, Jr., R. F. Sesek, and L. A. Cavuoto, "Barriers to the adoption of wearable sensors in the workplace: A survey of occupational safety and health professionals," *Hum. Factors, J. Hum. Factors Ergonom. Soc.*, vol. 60, no. 3, pp. 351–362, Jan. 2018.
[15] M. A. Case, H. A. Burwick, K. G. Volpp, and M. S. Patel, "Accuracy of smartphone applications and wearable devices for tracking physical activity data," *JAMA*, vol. 313, no. 6, pp. 625–626, Feb. 2015.
[16] C. B. Thomas, J. A. Stanley, and M. A. Kendrick, "Observations on some possible precursors of essential hypertension and coronary artery disease: VII. The subjective reaction to the cold pressor test as expressed in the verbal response," *J. Chronic Diseases*, vol. 14, pp. 355–365, Sep. 1961.
[17] G. Grassi *et al.*, "Heart rate as marker of sympathetic activity," *J. Hypertension*, vol. 16, no. 11, pp. 1635–1639, Nov. 1998.
[18] M. Esler *et al.*, "Assessment of human sympathetic nervous system activity from measurements of norepinephrine turnover," *Hypertension*, vol. 11, no. 1, pp. 3–20, Jan. 1988.
[19] A. B. Vallbo, K. E. Hagbarth, H. E. Torebjork, and B. G. Wallin, "Somatosensory, proprioceptive, and sympathetic activity in human peripheral nerves," *Physiol. Rev.*, vol. 59, no. 4, pp. 919–957, Oct. 1979.
[20] T. Bradley and P. Hjemdahl, "Renal extraction of endogenous and radiolabelled catecholamines in the dog," *Acta Physiol. Scandinavica*, vol. 126, no. 4, pp. 505–510, Apr. 1986.
[21] D. S. Goldstein, "Clinical assessment of catecholaminergic function," in *Stress, Catecholamines, and Cardiovascular Disease*. Oxford, U.K.: Oxford Univ. Press, pp. 234–286, 1995.
[22] S. Akselrod, D. Gordon, J. B. Madwed, N. C. Snidman, D. C. Shannon, and R. J. Cohen, "Hemodynamic regulation: Investigation by spectral analysis," *Amer. J. Physiol.*, vol. 249, no. 4, p. H867–H875, Oct. 1985.
[23] P. H. Ellaway, A. Kuppuswamy, A. Nicotra, and C. J. Mathias, "Sweat production and the sympathetic skin response: Improving the clinical assessment of autonomic function," *Auton. Neurosci.*, vol. 155, nos. 1–2, pp. 109–114, Jun. 2010.
[24] K. H. Chon *et al.*, "A novel quantitative method for diabetic cardiac autonomic neuropathy assessment in type 1 diabetic mice," *J. Diabetes Sci. Technol.*, vol. 8, no. 6, pp. 1157–1167, Nov. 2014.
[25] G. Grassi and M. Esler, "How to assess sympathetic activity in humans," *J. Hypertension*, vol. 17, no. 6, pp. 719–734, Jun. 1999.
[26] S. W. Lord, R. R. Senior, M. Das, A. M. Whittam, A. Murray, and J. M. McComb, "Low-frequency heart rate variability: Reproducibility in cardiac transplant recipients and normal subjects," *Clin. Sci.*, vol. 100, no. 1, pp. 43–46, Jan. 2001.
[27] M. Piepoli *et al.*, "Reproducibility of heart rate variability indices during exercise stress testing and inotrope infusion in chronic heart failure patients," *Clin. Sci.*, vol. 91, pp. 87–88, Jan. 1996.
[28] B. L. Marks and J. T. Lightfoot, "Reproducibility of resting heart rate variability with short sampling periods," *Can. J. Appl. Physiol.*, vol. 24, no. 4, pp. 337–348, Aug. 1999.
[29] G. Parati *et al.*, "Reproducibility of beat-by-beat blood pressure and heart rate variability," *Blood Pressure Monit.*, vol. 6, no. 4, pp. 217–220, Aug. 2001.
[30] G. D. Pinna *et al.*, "Heart rate variability measures: A fresh look at reliability," *Clin. Sci.*, vol. 113, no. 3, pp. 131–140, Aug. 2007.
[31] M. V. Højgaard, N.-H. Holstein-Rathlou, E. Agner, and J. K. Kanters, "Reproducibility of heart rate variability, blood pressure variability and baroreceptor sensitivity during rest and head-up tilt," *Blood Pressure Monit.*, vol. 10, no. 1, pp. 19–24, Feb. 2005.
[32] G. Parati, G. Mancia, M. Di Rienzo, and P. Castiglioni, "Point: Cardiovascular variability is/is not an index of autonomic control of circulation," *J. Appl. Physiol.*, vol. 101, no. 2, pp. 676–682, 2006.
[33] G. R. H. Sandercock, P. D. Bromley, and D. A. Brodie, "The reliability of short-term measurements of heart rate variability," *Int. J. Cardiol.*, vol. 103, no. 3, pp. 238–247, Sep. 2005.
[34] B. M. W. Illigens and C. H. Gibbons, "Sweat testing to evaluate autonomic function," *Clin. Auton. Res.*, vol. 19, no. 2, pp. 79–87, Apr. 2009.
[35] C. Setz, B. Arnrich, J. Schumm, R. L. Marca, G. Tröster, and U. Ehlert, "Discriminating stress from cognitive load using a wearable EDA device," *IEEE Trans. Inf. Technol. Biomed.*, vol. 14, no. 2, pp. 410–417, Mar. 2010.
[36] J. A. Healey and R. W. Picard, "Detecting stress during real-world driving tasks using physiological sensors," *IEEE Trans. Intell. Transp. Syst.*, vol. 6, no. 2, pp. 156–166, Jun. 2005.
[37] H. F. Posada-Quintero *et al.*, "Power spectral density analysis of electrodermal activity for sympathetic function assessment," *Ann. Biomed. Eng.*, vol. 44, no. 10, pp. 3124–3135, Oct. 2016.
[38] H. F. Posada-Quintero, J. P. Florian, D. Orjuela-Cañón, and K. H. Chon, "Highly sensitive index of sympathetic activity based on time-frequency spectral analysis of electrodermal activity," *Amer. J. Physiol.-Regulatory, Integr. Comparative Physiol.*, vol. 311, no. 3, pp. R582–R591, Sep. 2016.
[39] J. A. Horne, "A review of the biological effects of total sleep deprivation in man," *Biol. Psychol.*, vol. 7, nos. 1–2, pp. 55–102, Sep. 1978.
[40] E. Miró, M. C. Cano-Lozano, and G. Buela-Casal, "Electrodermal activity during total sleep deprivation and its relationship with other activation and performance measures," *J. Sleep Res.*, vol. 11, no. 2, pp. 105–112, Jun. 2002.
[41] E. Gil, M. Orini, R. Bailón, J. M. Vergara, L. Mainardi, and P. Laguna, "Time-varying spectral analysis for comparison of HRV and PPG variability during tilt table test," in *Proc. IEEE Conf. Annu. Int. Conf. Eng. Med. Biol.*, Aug./Sep. 2010, pp. 3579–3582.
[42] J. R. Stroop, "Studies of interference in serial verbal reactions," *J. Exp. Psychol.*, vol. 18, no. 6, pp. 643–662, 1935.
[43] I. R. Kleckner *et al.*, "Simple, transparent, and flexible automated quality assessment procedures for ambulatory electrodermal activity data," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 7, pp. 1460–1467, Jul. 2018.
[44] A. Greco, G. Valenza, A. Lanata, E. Scilingo, and L. Citi, "cvxEDA: A convex optimization approach to electrodermal activity processing," *IEEE Trans. Biomed. Eng.*, vol. 63, no. 4, pp. 797–804, Aug. 2015.
[45] K. H. Chon, S. Dash, and K. Ju, "Estimation of respiratory rate from photoplethysmogram data using time–frequency spectral estimation," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 8, pp. 2054–2063, Aug. 2009.
[46] N. E. Huang *et al.*, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.
[47] M.-E. Nygårds and L. Sörnmo, "Delineation of the QRS complex using the envelope of the e.c.g," *Med. Biol. Eng. Comput.*, vol. 21, no. 5, pp. 538–547, Sep. 1983.
[48] C. Vidaurre, T. H. Sander, and A. Schlögl, "BioSig: The free and open source software library for biomedical signal processing," *Comput. Intell. Neurosci.*, vol. 2011, Dec. 2011, Art. no. 935364.
[49] F. J. Massey, Jr., "The Kolmogorov-Smirnov test for goodness of fit," *J. Amer. Statist. Assoc.*, vol. 46, no. 253, pp. 68–78, 1951.
[50] L. H. Miller, "Table of percentage points of Kolmogorov statistics," *J. Amer. Stat. Assoc.*, vol. 51, no. 273, pp. 111–121, 1956.
[51] J. Wang, W. W. Tsang, and G. Marsaglia, "Evaluating Kolmogorov's distribution," *J. Stat. Softw.*, vol. 8, no. 18, pp. 1–14, 2003.
[52] M. Friedman, "The use of ranks to avoid the assumption of normality implicit in the analysis of variance," *J. Amer. Statist. Assoc.*, vol. 32, no. 200, pp. 675–701, Dec. 1937.
[53] K. O. McGraw and S. P. Wong, "Forming inferences about some intraclass correlation coefficients," *Psychol. Methods*, vol. 1, no. 1, pp. 30–46, 1996.
[54] J. R. Landis and G. G. Koch, "The measurement of observer agreement for categorical data," *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977.
[55] D. J. Ewing and B. F. Clarke, "Diagnosis and management of diabetic autonomic neuropathy," *Brit. Med. J., Clin. Res. Ed.*, vol. 285, no. 6346, pp. 916–918, Oct. 1982.
[56] Y. Takimoto, K. Yoshiuchi, T. Ishizawa, Y. Yamamoto, and A. Akabayashi, "Autonomic dysfunction responses to head-up tilt in anorexia nervosa," *Clin. Auton. Res.*, vol. 24, no. 4, pp. 175–181, Aug. 2014.
[57] J. P. Florian, E. E. Simmons, K. H. Chon, L. Faes, and B. E. Shykoff, "Cardiovascular and autonomic responses to physiological stressors before and after six hours of water immersion," *J. Appl. Physiol.*, vol. 115, no. 9, pp. 1275–1289, Nov. 2013.
[58] A. Garafova, A. Penesova, E. Cizmarova, A. Marko, M. Vlcek, and D. Jezova, "Cardiovascular and sympathetic responses to a mental stress task in young patients with hypertension and/or obesity," *Physiol. Res.*, vol. 63, p. S459–S467, Jan. 2014.
[59] Z. Visnovcova *et al.*, "Complexity and time asymmetry of heart rate variability are altered in acute mental stress," *Physiol. Meas.*, vol. 35, no. 7, pp. 1319–1334, Jul. 2014.
[60] R. Edelberg, "Electrodermal mechanisms: A critique of the two-effector hypothesis and a proposed replacement," in *Progress in Electrodermal Research*, J.-C. Roy, W. Boucsein, D. C. Fowles, and J. H. Gruzelier, Eds. Boston, MA, USA: Springer, 1993, pp. 7–29.
[61] W. Boucsein, *Electrodermal Activity*. Boston, MA, USA: Springer, 2012.
[62] R. Edelberg, "Mechanisms of electrodermal adaptations for locomotion, manipulation, or defense," *Prog. Physiol. Psychol.*, vol. 5, pp. 155–209, 1973.

[63] M.-N. Girardot and M. C. Koss, "A physiological and pharmacological analysis of the electrodermal response in the rat," *Eur. J. Pharmacol.*, vol. 98, no. 2, pp. 185–191, Feb. 1984.
[64] M. C. Koss and M. A. Davison, "The electrodermal response as a model for central sympathetic reactivity: The action of clonidine," *Eur. J. Pharmacol.*, vol. 37, no. 1, pp. 71–78, May 1976.
[65] S. A. Shields, K. A. MacDowell, S. B. Fairchild, and M. L. Campbell, "Is mediation of sweating cholinergic, adrenergic, or both? A comment on the literature," *Psychophysiology*, vol. 24, no. 3, pp. 312–319, May 1987.
[66] T. P. Zahn, J. L. Rapoport, and C. L. Thompson, "Autonomic effects of dextroamphetamine in normal men: Implications for hyperactivity and schizophrenia," *Psychiatry Res.*, vol. 4, no. 1, pp. 39–47, Feb. 1981.
[67] Y. Zhong *et al.*, "Autonomic nervous nonlinear interactions lead to frequency modulation between low- and high-frequency bands of the heart rate variability spectrum," *Amer. J. Physiol.-Regulatory, Integr. Comparative Physiol.*, vol. 293, no. 5, p. R1961–R1968, Nov. 2007.
[68] A. McGrady, C. Kern-Buell, E. Bush, S. Khuder, and B. P. Grubb, "Psychological and physiological factors associated with tilt table testing for neurally mediated syncopal syndromes," *Pacing Clin. Electrophysiol.*, vol. 24, no. 3, pp. 296–301, Mar. 2001.

HUGO F. POSADA-QUINTERO received the B.S. degree in electronic engineering from the Universidad Distrital Francisco José de Caldas, Bogotá D.C., Colombia, the M.S. degree in electronics and computers engineering from the Universidad de los Andes, Bogotá D.C., and the Ph.D. degree in biomedical engineering from the University of Connecticut, Storrs. He is currently an Assistant Research Professor with the Department of Biomedical Engineering, University of Connecticut. His current research interests include biomedical signal processing and biomedical instrumentation.

TANYA DIMITROV is currently a senior student in biomedical engineering with a minor in computer science with the University of Connecticut. With a passion for physiological signals and signal processing, she plans to pursue a career in personal healthcare devices and to attend graduate school.

AURELIE MOUTRAN is currently a sophomore student with the Department of Biomedical Engineering, University of Connecticut.

SEOHYOUNG PARK is currently a senior high school student with the Northern Highlands Regional High School.

KI H. CHON received the B.S. degree in electrical engineering from the University of Connecticut, Storrs, the M.S. degree in biomedical engineering from The University of Iowa, Iowa, and the M.S. degree in electrical engineering and the Ph.D. degree in biomedical engineering from the University of Southern California, Los Angeles. He is currently the John and Donna Krenicki Chair Professor and the Head of biomedical engineering with the University of Connecticut, Storrs, CT, USA.

He is a Fellow of the American Institute of Medical and Biological Engineering and of the International Academy of Medical and Biological Engineering. He has chaired many international conferences including his role as the Program Co-Chair for the IEEE EMBS conference in NYC, in 2006, and as the Conference Chair for the 6th International Workshop on Biosignal Interpretation in New Haven, CT, USA, in 2009. He was an Associate Editor of the IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING from 2007 to 2013. He is a Co-Founder of Mobile Sense Technologies which is located in the TIP Center, Farmington, CT, USA. The company has recently been granted both NIH and NSF SBIR grants.