A Normalization Algorithm to compare Scores from different Sample Size Data derived from Discrete Stochastic Models

Jeffery Jonathan Joshua (ישוע) Davis, Florian Schübeler
The Embassy of Peace, Whitianga, New Zealand
joshua_888@yahoo.com, florian@theembassyofpeace.com

In recent years several types of biofeedback systems have become available for both research and public use. Particularly interesting are the types of biofeedback systems that measure Heart Rate Variability (HRV) based on Interbeat Intervals (IBI or R-R intervals). Such biofeedback systems have been available for public use for a while allowing the user to keep records of progress via coherence scores based on frequency measures derived from HRV signals. Here we explore new metrics for the comparison between different types of scores available as measures of psychophysiological coherence and we propose a normalization algorithm that allows us to compare some of these metrics when derived from different sample sizes. This methodology, we suggest, will be useful when comparing coherence scores from different participants in different modalities in research settings. Ideally, this could be incorporated as part of the new generation of biofeedback systems to support users having a quantitative measure of their progress under different modalities or activities in daily life.

Keywords - *Heart Rate Variability (HRV), Psychophysiological Coherence, Coherence Scores, Euclidean Distance, Semi-Markov Processes, Markov Chain Analysis, Inner Peace.*

INTRODUCTION

Psychophysiological coherence has been broadly studied in recent years in association with HRV (McCraty et al., 2006), (McCraty, 2002), (Davis et al., 2019a). Several systems are available for research and public use varying in purpose, functionality, quality, interface and price. Here we briefly describe the emWave system (Quantum Intech, Inc., 2010) since we have derived our data from recordings provided by users of this device-system. The emWave system allows the user to measure HRV via an earlobe sensor into a data acquisition, analysis and score system. The scores are derived from a coherence ratio that is computed based on the different frequency components or bands from the frequency spectrum of the IBI (Tarvainen & Niskanen, 2012), (Medicore, n.d.). The coherence scores are then computed based on a set of thresholds and criteria that relate to different levels or degrees of desired mastery known as challenge level (HeartMath, Inc., 2018).

When measuring HRV under different activities or situations, a user may derive very important information from his or her scores that will support him or her in making choices to correct or manage certain psychological or physiological responses (Davis & Schübeler, 2019). For example, listening to an inspirational reading may have very different effects on the user's psychophysiological makeup than meditating or reading complex, energy consuming material (Davis et al., 2019). The emWave system provides a colour coding to discriminate between significantly stress related (red), relaxed (blue) or coherent (green) states in terms of the coherence ratio measure (McCraty et al., 2001), (Childre & Cryer, 2008). These states are monitored numerically by a coherence score as well as a cumulative coherence score around every five (5) seconds. Here we present the reader with two (2) types of score measures: (1) normal score (S) and (2) cumulative score (CS).

The normal scores are derived from three (3) different kinds of simulated Markov processes (Ross, 1983), (Law & Kelton, 1991), (Howard, 1971) associated to three (3) different types of transition probability matrices (TPM), which we have called red, blue and green matrix all derived from experimental data (Davis et al., 2019). These TPM will be used to simulate and model three (3) distinct types of psychophysiological scenarios related predominantly to stressed, relaxed or coherent states. Note that CS is derived from S. In the first section of this work we present the reader with four (4) types of metrics that

allow us to compare different stochastic processes, in our case derived from red, blue and green matrices as stated before. These metrics are: (1) a Euclidean Distance (Fraleigh & Beauregard, 1995) based on normal scores (ED_S), (2) a Euclidean Distance based on cumulative scores (ED_CS), (3) a distance based on the final cumulative score (CS_final) and (4) the average of the ideal minus the normal scores (Mean_S).

In the second section we present a novel (to our knowledge) computation for the comparison of the ED_CS from two (2) data sets with different sample sizes. This requires the solution of some integrals in order to derive a normalization factor.

Ideally, when collecting experimental data for HRV we should design it in a way that guarantees a same sample size for different runs or trials. However, this is sometimes difficult to achieve: (1) due to time and resource constraints and participant availability, (2) some data loss and (3) when comparing data across studies.

Finally, we present a comparative analysis between different scenarios for different metrics, for different and equal sample sizes. We also compare metrics derived from different sample sizes and derive some considerations and conclusions for further research.

EXPERIMENTAL METHODS AND MODELS

This section briefly describes the methods and models that were utilized in order to derive the TPM used in the simulation models. First of all, we analyzed coherence scores derived from an experimental setting with three (3) modalities: meditation, listening to a reading from a book and visioning, as described in (Davis et al., 2019). Each data set was exported and prepared for a comparative analysis between participants, modalities and sessions that involved a counting process based on normal scores in order to estimate the experimental TPM per participant, modality and session. This gave us the means to estimate an average TPM across sessions for each participant per modality.

We selected three (3) distinct matrices that best matched our requirements where one matrix represents a predominantly 'green' state, another one represents a predominantly 'blue' state and the third one a predominantly 'red' state.

In Figure 1 we display the three (3) average TPM that were chosen for this study.

Based on these three (3) TPM (associated to green, blue and red states) we carried out

$$G_{ij} = \begin{matrix} & 0 & 1 & 2 \\ 0 & \\ 1 & \\ 2 & \end{matrix} \begin{bmatrix} 0.41 & 0.43 & 0.16 \\ 0.11 & 0.57 & 0.32 \\ 0.01 & 0.05 & 0.94 \end{bmatrix}$$ Dominant Coherent State (green)

$$B_{ij} = \begin{matrix} & 0 & 1 & 2 \\ 0 & \\ 1 & \\ 2 & \end{matrix} \begin{bmatrix} 0.8 & 0.16 & 0.04 \\ 0.23 & 0.49 & 0.28 \\ 0.03 & 0.1 & 0.87 \end{bmatrix}$$ Dominant Relaxed State (blue)

$$R_{ij} = \begin{matrix} & 0 & 1 & 2 \\ 0 & \\ 1 & \\ 2 & \end{matrix} \begin{bmatrix} 0.74 & 0.26 & 0 \\ 0.25 & 0.5 & 0.25 \\ 0.06 & 0.22 & 0.72 \end{bmatrix}$$ Dominant Stressed State (red)

$$\forall\, i = 1, 3\ \&\ j = 1, 3$$

Figure 1
displays the three (3) chosen matrices associated to different tendencies for coherent, relaxed or stressed states; green (G), blue (B) and red (R) respectively.

simulations for a variety of experimental settings, where we varied the number of simulated points from 200 to 1000 and the number of sessions from 10 to 100.

From these simulations we obtained the simulated scores (S) and we then applied our methodology to carry out the main analysis.

METHODOLOGY

The methodology we used is explained in great detail in (Davis & Schübeler, 2019), (Davis et al., 2019). Here we present the reader with a summary that allows for a basic understanding of the concepts and computations that follow.

Let us call $X_t = \{0, 1, 2\}$ the possible states of the system at time t where $X_t = 0$ is associated to green (coherent state), $X_t = 1$ blue (relaxed state) and $X_t = 2$ red (stressed state).

We define the discrete time TPM as:

$$P_{ij} = \begin{pmatrix} P_{00} & P_{01} & P_{02} \\ P_{10} & P_{11} & P_{12} \\ P_{20} & P_{21} & P_{22} \end{pmatrix} \quad i, j = 0, 1, 2 \quad (1)$$

that describes the transitions from state i corresponding to X_t to state j, corresponding to X_{t+1} for t (t=1, 2, 3, ...n) where each unit t corresponds to a five (5) second period that moves as follows: 5, 10,.... 5*t, ... 5*n. In this study, states i and j can only take the values of 0, 1 or 2.

Also for this study, we computed the simulated scores (S_t) and the cumulative scores (CS_t) where S_t is a function of X_t and computed as follows:

$$X_t = \begin{cases} 0, \text{ for a coherent state, green} \\ 1, \text{ for a relaxed state, blue} \\ 2, \text{ for a stressed state, red} \end{cases}$$

$$\forall\, t = 1, 2, 3, \dots\dots\dots$$

$$S_t = \begin{cases} 2 \text{ when } X_t = 0 \\ 1 \text{ when } X_t = 1 \\ -1 \text{ when } X_t = 2 \end{cases}$$

$$\forall\, t = 1, 2, 3, \dots\dots\dots$$

Then the cumulative score CS_t is computed

as follows:

$$CS_t = S_0 + \sum_{(t=1)}^{N} S_t \qquad (2)$$

In order to assess the performance of a participant, we need what we could call an ideal ongoing state which in our case is described by the highest possible score, the 'Ideal Score', IS_t where: $IS_t = \max\{S_t\} = 2$ and N is the sample size for a session. Similarly, the 'Ideal Cumulative Score' is ICS_t, where:

$$ICS_t = \sum_{(t=1)}^{N} IS_t \text{ where } IS_t = 2 \quad \forall \ t \qquad (3)$$

These ideal scores allow us to compute the gap between the actual and the ideal state of the system, which can be thought of as a gap or an error and which can be estimated via a metric.

With this in mind we are equipped to compute four (4) metrics which may be useful for different situations, purposes and data sets. These metrics are ED_S, ED_CS, CS_final and Mean_S, where:

$$ED\_S = \sqrt{\left(\sum_{(t=1)}^{N} (IS_t - S_t)^2 \right)} \qquad (4)$$

$$ED\_CS = \sqrt{\left(\sum_{(t=1)}^{N} (ICS_t - CS_t)^2 \right)} \qquad (5)$$

$$CS\_final = (ICS_N - CS_N) \qquad (6)$$

$$Mean\_S = \frac{\left(\sum_{(t=1)}^{N} ISt - St \right)}{N} \qquad (7)$$

Where N is the sample size and ICS_N is the last ideal cumulative score computed, *ICS_final*.

NORMALIZATION ALGORITHM

When dealing with having to compare processes coming from different sample sizes we may face several issues. One of them, perhaps the most common one is the calculation of confidence intervals which is directly affected by sample size within a certain range. However, the real challenge arises here for us when we deal with particularly two (2) of our metrics, ED_S and ED_CS. Following we will explore why.

Let us say we gather ten (10) data points each with a value of one (1). That means that the average of these values would be equal to one (1) and let's assume that they come from the same population. In such a situation we would be able to compare the two sample means with the appropriate two-sample t-test for equal means and then reject or accept. However, if we used the metrics ED_S or ED_CS applying these tests would be inappropriate since the first metric is a Euclidean Distance involving the square root of a sum of squares, as shown in equation (4). For the second metric, we compute an

even more complex Euclidean Distance that involves the square root of a cumulative sum of squares, as shown in equation (5).

If we invoke the same example for let's say purely red and blue processes, then we would expect an ED_S = 6.32 for the red process and an ED_S = 3.16 for the blue one, each with a sample size of n=10. For a sample size of n=5 the values would be 4.47 and 2.23 for the red and blue process respectively. The reader must note that it would be meaningless to compare blue or red processes of different sample sizes unless we normalize the results while also taking care of the confidence intervals.

Here we will only deal with the issue of normalizing same processes either red, blue or green without addressing the associated complexities in computing the appropriate confidence intervals for the normalized data set, since this is outside the scope of this work. We will also leave for another study an in depth analysis in applying these metrics for a mixed red-blue-green process.

Theoretical values for the purely red, blue and green process are derived via the following mathematical models for ED_S and ED_CS for the three (3) limiting cases or processes. Also, these formulas will be used when in need for normalization due to different sample sizes, as will be explained and illustrated in the next section.

ED_S$_2$ (pure red process) (8)

$$ED\_S_2 = \sqrt{\left(\sum_{(t=1)}^{N} (2-(-1))^2 \right)} = \sqrt{\left(\sum_{(t=1)}^{N} (3)^2 \right)}$$

$$ED\_S_2 = \sqrt{\left(\sum_{(t=1)}^{N} 9 \right)} = \sqrt{(N * 9)} = 3 * \sqrt{N}$$

ED_S$_1$ (pure blue process) (9)

$$ED\_S_1 = \sqrt{\left(\sum_{(t=1)}^{N} (2-1)^2 \right)} = \sqrt{\left(\sum_{(t=1)}^{N} (1)^2 \right)}$$

$$ED\_S_1 = \sqrt{\left(\sum_{(t=1)}^{N} 1 \right)} = \sqrt{(N * 1)} = \sqrt{N}$$

ED_S$_0$ (pure green process) (10)

$$ED\_S_0 = \sqrt{\left(\sum_{(t=1)}^{N} (2-2)^2 \right)} = \sqrt{\left(\sum_{(t=1)}^{N} (0)^2 \right)}$$

$$ED\_S_0 = \sqrt{\left(\sum_{(t=1)}^{N} 0 \right)} = \sqrt{(N * 0)} = \sqrt{0} = 0$$

ED$_2$ (ED_CS, pure red process) (11)

$$D_2 = \int_0^{N_2} (2t - 0)^2 * dt = \int_0^{N_2} (2t)^2 * dt$$

$$D_2 = \frac{4t^3}{3}\Big|_0^{N_2} = \frac{4(N_2)^3}{3}$$

$$ED_2 = \sqrt{D_2} = \frac{2(N_2)^{3/2}}{\sqrt{3}}$$

ED_1 (ED_CS, pure blue process)   (12)

$$D_1 = \int_0^{N_1} (2t - t)^2 * dt = \int_0^{N_1} t^2 * dt$$

$$D_1 = \frac{t^3}{3}\Big|_0^{N_1} = \frac{(N_1)^3}{3}$$

$$ED_1 = \sqrt{D_1} = \frac{N_1^{3/2}}{\sqrt{3}}$$

ED_0 (ED_CS, pure green process)   (13)

$$D_0 = \int_0^{N_0} (2t - 2t)^2 * dt$$

$$D_0 = 0$$

$$ED_0 = \sqrt{D_0} = 0$$

ANALYSIS OF DATA

In this section we will evaluate the four (4) metrics we obtained regarding their feasibility to best analyse HRV data in order to better interpret the results obtained from it. We also aim to make some initial assessment of what number of data points per session (sample size) give the best results, while at the same time meeting feasibility requirements in obtaining such quality of results. For example, to have sessions of 30 minutes may be more feasible in certain experimental settings than obtaining data from several hours of recording. Finally, we provide some recommendations for the number of sessions that should be recorded in order to derive meaningful HRV analysis.

First, in the following Figure 2, we use the three (3) matrices G_{ij}, B_{ij} and R_{ij} that we introduced in Figure 1 in order to compute the limiting probabilities (LP). The LP values for R_{ij} are: $P(X_t=2) = 0.39$ (red state), $P(X_t=1) = 0.32$ (blue state) and $P(X_t=0)= 0.29$ (green state). The LP values for B_{ij} are: $P(X_t=2) = 0.29$ (red state), $P(X_t=1) = 0.19$ (blue state) and $P(X_t=0)= 0.52$ (green state). Finally, the LP values for G_{ij} are: $P(X_t=2) = 0.04$ (red state), $P(X_t=1) = 0.14$ (blue state) and $P(X_t=0)= 0.82$ (green state).

The reader must note that the state variable in the real system (level of coherence or state, X_t) has been observed to remain only for a relatively short time in the blue state, which can be understood as a transition state and the system is more likely to gravitate towards either green or red. When the system spends a large percentage of time in blue, we interpret this as a system with very frequent transitions between green and red, as we can observe in real life experimental data associated with R_{ij}.
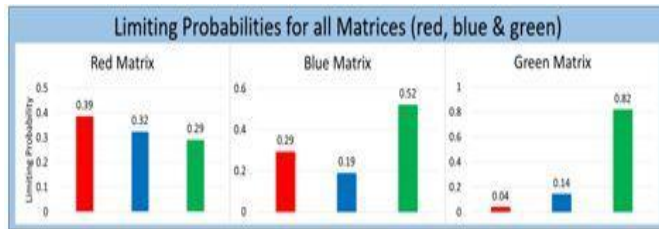
Figure 2

displays the limiting probabilities for the red (R_{ij}), blue (B_{ij}) and green (G_{ij}) matrix (left to right) respectively.

We used each matrix to simulate the stochastic processes for a set of sessions (10, 20, 30, 40, 50 and 100 runs) with different sample sizes (200, 400, 600, 800 and 1000 points). Note that 200 points ≈ 30 minutes of real life HRV recordings with the emWave system. In Figure 3 we display the results for the mean values for different sample sizes and runs associated with the Euclidian Distance (ED, Norm2) (Fraleigh & Beauregard, 1995) of the difference between the ideal cumulative score vector (ICS_t) and the cumulative simulated score vector (CS_t). For this metric (ED_CS), we also computed the corresponding confidence intervals (CI) (Allen, 1978), (Law & Kelton, 1991). We conducted the same type of analyses for the other three (3) metrics:

(1)   The mean values across sessions for the ED of the difference between the ideal score vector (IS_t) and the simulated score vector (S_t), which we call ED_S.

(2)   The mean values across sessions of the final cumulative score (CS_final).

(3)   The grand mean (Mean_S) of the computed means of S_t across sessions.

All four (4) metrics display similar tendencies. Whether calculated for ten (10) sessions or more, the mean differs very little for different sample sizes and the CI are also relatively small. However, for a sample size of 200 points we observe slightly larger CI around the mean and we interpret that to be a minimum good or acceptable sample size below which the statistical significance of the estimated means would be questionable.

In Figure 4 we present the results obtained from the simulation for the ED_CS metric. While we can identify some variations in the mean from 10 to 100 sessions, the differences are very small, however the CI gets noticeably smaller as the number of sessions increases, something expected. When we computed the same type of comparison for the other three (3) metrics, we found very similar results. For all metrics we observed large CI and relatively poor statistical significance in the results for ten (10) sessions which we regard as a very minimum number of sessions in real life experiments. It is certainly recommended that twenty (20) or more sessions be recorded for very good quality results. This applies to all sample sizes which means that for very good results we could record 30-60 minutes of data, which is a comfortable recording time

for participants, according to our experience. This would mean 200 to 400 points per session.
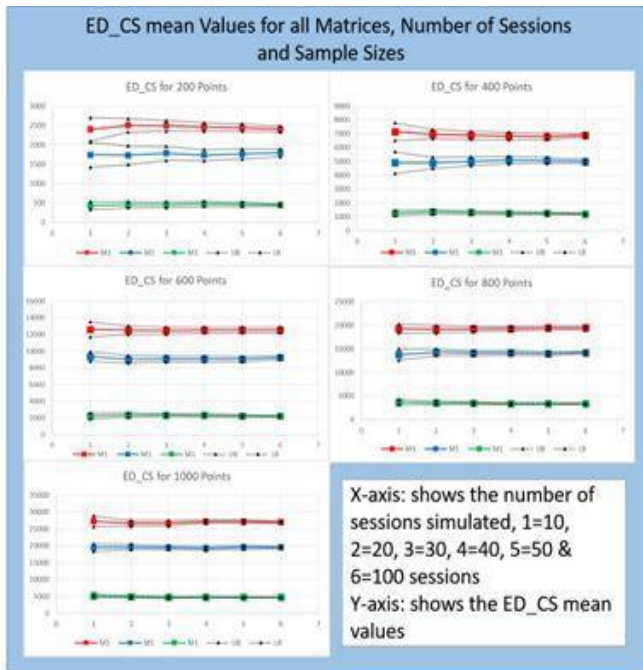

Figure 3
displays the simulation means (M1) of the ED_CS metric for all three (3) matrices (red, blue & green) and for all the combinations of sample sizes and sessions. Sample size of 200 points (top left), 400 points (top right), 600 points (middle left), 800 points (middle right) and 1000 points (bottom left), each value in a graph represents the mean for a particular number of sessions (10, 20, 30, 40, 50 & 100 sessions, 1 to 6 on the x-axis respectively).

We then computed the coefficient of variation for each of the four (4) metrics and again for all simulations, sample sizes and number of
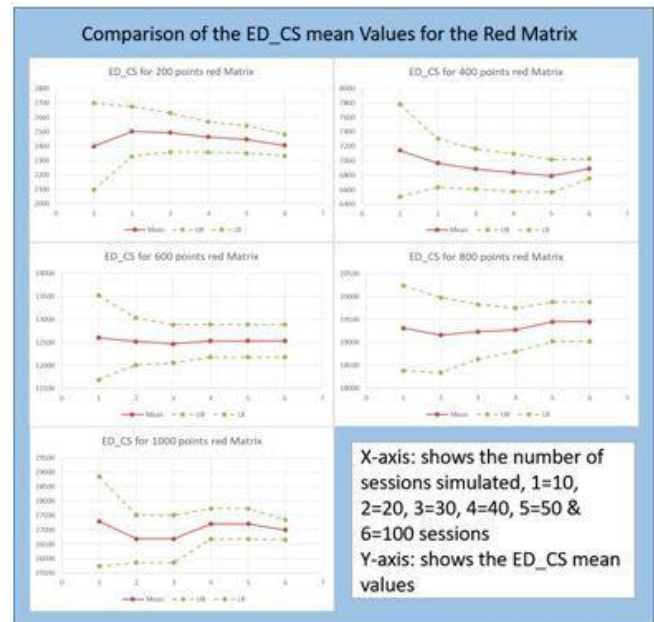

Figure 4
displays the ED_CS for the red matrix, for all number of points and all number of sessions.

sessions for all three (3) matrices. Following, we present the general formula we used to compute the sampled coefficient of variation:

$$\hat{c}_v = \frac{s}{\bar{x}} \qquad (14)$$

The coefficients of variation are displayed in Figure 5 and we can appreciate very similar results for these computations per metric and per type of matrix. It is important to note that the graph for the CS_final (bottom left) shows the results associated to the red matrix as the one with the highest values, while the results associated with the green matrix are the smallest. This is expected since the green matrix will generate a CS much closer to the

ideal CS than the blue and red matrix, therefore displaying a larger CS_final value. This in turn results in a larger CS_final mean value over all simulations, which finally will result in a smaller coefficient of variation. For all the other cases, the analysis is trivial since the green matrix is expected to show the highest scores while the blue shows slightly better scores than the red matrix.

In Figure 6 we compare Mean_S with the expected or theoretical mean, as well as the simulated and theoretical standard deviations
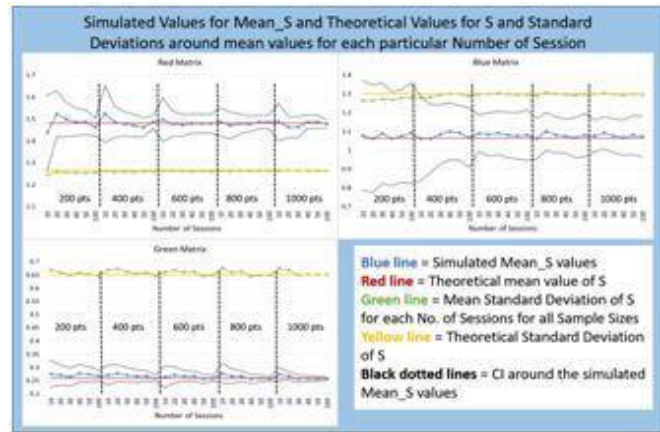

Figure 6
displays the theoretical mean values (red line) and the simulated means (blue dots) together with the standard deviations of the theoretical mean (yellow line) and the standard deviation of the simulated means (green dots).

Figure 5
displays the coefficient of variation (c_v) for each of the four (4) metrics for the red matrix (red line), blue matrix (blue line) and green matrix (green line).

over each particular number of sessions. The reader can easily observe that for all three (3) matrices, the simulated mean is very close to the theoretical mean. The same applies to the standard deviations. Again, the standard

deviation around the simulated means per session follows very closely the tendency of the standard deviation of the theoretical mean values. Note that for all sample sizes the results associated with ten (10) sessions is noticeably larger in CI further supporting our previous observations.

Finally, we present in Figure 7 the Mean_S together with associated CI, for all sample sizes for all number of sessions.

We can observe that as sample size increases, the CI gets closer to the mean, and again for a number of ten (10) sessions the CI increases noticeably for all sample sizes. We can also observe, as expected, that for any particular sample size, the CI gets closer to the mean as the number of sessions
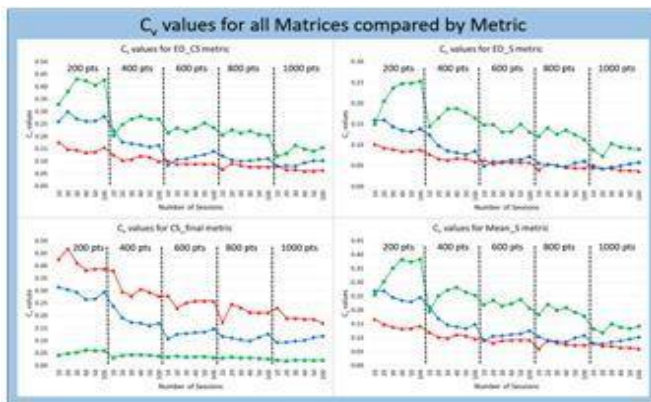
increases. While the CI for the green matrix are already very small for simulations with 200 points sample size, the red and blue matrix produce means with relatively larger CI, when compared to the green. Overall, we are able to properly classify the different processes (red, blue and green) for all sample sizes and all number of sessions for all metrics. This means that even though the CI are larger for ten (10) sessions and a sample size of 200, still we can achieve a good classification.

Figure 7
displays the Mean_S values for all simulated scenarios for the green matrix (green line), the blue matrix (blue line) and the red matrix (red line) and their corresponding CI (black dotted lines). Note that the six (6) values within each of the different sample size regions for 200, 400, 600, 800 & 1000 points correspond to the 10, 20, 30, 40, 50 & 100 runs respectively.

In order to complement our CI analysis we calculated the t-values and p-values for the null hypotheses H_0: $\mu = \mu_0$ and the alternative H_1: $\mu \neq \mu_0$ where μ_0: {E(ED_CS), E(ED_S),

E(CS_final), E(IS-S)} and each theoretical expected value for each metric where N is the sample size for a particular session, LP are the limiting probabilities and IS is the ideal score with a value of two (2) and are computed as follows:

$$E(ED\_CS) = (IS - S) * \sqrt{(N^3/3)} \tag{15}$$

$$E(ED\_S) = \sqrt{(E(IS - S)^2)} * \sqrt{N} \tag{16}$$

$$E(CS\_final) = \overline{S} * N \tag{17}$$

$$E(IS-S) = \sum_{(i=1)}^{3} (IS - S_i) * LP (IS - S_i) \tag{18}$$

The t-statistic is computed as:

$$t = \frac{\overline{x} - \mu_0}{s/\sqrt{n}} \text{ where n is the no. of sessions} \tag{19}$$

In our case $\overline{x}$ can be any of the four (4) metrics: Mean ED_CS, Mean ED_S, Mean CS_final, Mean_S. Following we present in Table I the results for the case of ED_CS for all three (3) matrices (red, blue and green).

It is important to note that p-values greater than or equal to α = 0.05 would lead us to accept the null hypothesis H_0 which is the case for all p-values for all sample sizes and number of sessions. Furthermore, these results give us the confidence that we simulated the processes for a large enough sample size and number of sessions, even for
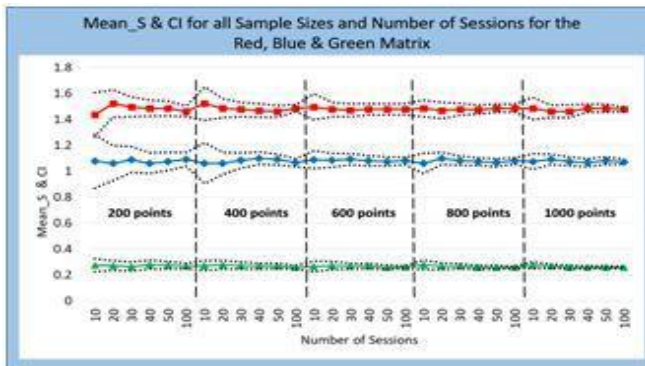
the very minimum of 200 points and ten (10) sessions. All of this means to us that we could expect good enough results in real life measurements for half an hour to an hour per session, for a total of ten (10) to twenty (20) day sessions. All of this is fairly easy to achieve with a relatively small budget. Finally, it is important to mention that we performed the same analysis for the other two (2) metrics with similar results, where all the p-values were greater than or equal to α= 0.05.

# Points	# Sessions	ED_CS	ED_CS	ED_CS	Mean_S	Mean_S	Mean_S
200	10	0.78	0.97	0.98	0.57	0.89	0.50
200	20	0.44	0.87	0.45	0.41	0.97	0.54
200	30	0.39	0.71	0.41	0.68	0.60	0.80
200	40	0.62	0.82	0.13	0.86	0.95	0.28
200	50	0.83	0.79	0.14	0.87	0.73	0.34
200	100	0.41	0.37	0.26	0.37	0.30	0.43
400	10	0.35	0.94	0.42	0.48	0.98	0.51
400	20	0.53	0.82	0.05	0.87	0.95	0.25
400	30	0.88	0.62	0.11	0.89	0.50	0.37
400	40	0.82	0.22	0.24	0.66	0.14	0.54
400	50	0.51	0.20	0.27	0.41	0.18	0.42
400	100	0.70	0.44	0.99	0.71	0.86	0.75
600	10	0.99	0.23	0.75	0.73	0.42	0.82
600	20	0.76	0.89	0.26	0.89	0.45	0.40
600	30	0.54	0.88	0.13	0.70	0.21	0.46
600	40	0.73	0.86	0.15	0.90	0.37	0.47
600	50	0.69	0.98	0.57	0.88	0.49	0.77
600	100	0.58	0.16	0.57	0.83	0.24	0.91
800	10	0.86	0.81	0.22	0.85	0.94	0.28
800	20	0.57	0.28	0.34	0.70	0.12	0.58
800	30	0.61	0.54	0.46	0.90	0.33	0.48
800	40	0.64	0.37	0.90	0.79	0.42	0.73
800	50	0.76	0.66	0.90	0.73	0.85	0.77
800	100	0.66	0.07	0.71	0.63	0.17	0.59
1000	10	0.77	0.97	0.09	0.90	0.66	0.25
1000	20	0.33	0.54	0.28	0.43	0.16	0.27
1000	30	0.23	0.96	0.76	0.33	0.47	0.72
1000	40	0.63	0.54	0.87	0.73	0.85	0.71
1000	50	0.59	0.67	0.93	0.70	0.26	0.63
1000	100	0.64	0.68	0.75	0.69	0.44	0.70

Table I

p-values for ED_CS and Mean_S

NORMALIZATION FOR DIFFERENT SAMPLE SIZES

In the following part of this analysis section, we introduce the reader to the results obtained from our normalization algorithms according to the formulas introduced in section 'Normalization algorithm'.

When the need arises to compare the ED_CS results (a metric derived from a cumulative process) coming from different simulations, which are different in sample size, we apply our normalization formulas as depicted in equations (8) to (13). We always normalize from the larger sample size to the smaller one in order to be conservative with CI. For example, for the red matrix with 200 points and 10 sessions we got an ED_CS value of 2398.048 while for 400 points and 10 sessions we computed a value of 7143.003. We then normalized the results from a sample size of 400 points by applying equation (11) and obtained a normalized ED_CS value of 2525.43307 comparable to 2398.048. We normalized all possible combinations based on the simulation sample size, which means for example, that for a sample size of 200 points we normalized ED_CS from 400 to 200, from 600 to 200, from 800 to 200 and from 1000 to 200 points.

As an illustration, we present in Figure 8 the normalized results for all ED_CS obtained for 10 and 100 sessions only. The blue bars present the ED_CS value computed from simulations that are the normalization target (smaller sample size) and the red bars

present the normalized ED_CS values from simulations with larger sample size (normalization source).

Here the reader can visually identify that for all normalizations, the source values are very close to the values they were normalized to

Figure 8

displays ED_CS target value (blue bars) vs. the ED_CS source value after normalizing the ED_CS to the target value (red bars). These results are shown for 10 (left) and 100 (right) sessions for the red (top), blue (middle) and green (bottom) matrix.

(target values). This we would expect since we used the same matrices to simulate different sample sizes in order to test the normalization equations. Even though we are confident that these formulas are reliable to allow us to compare data from different

sample sizes, we still recommend a further in depth statistical analysis, which is outside the scope of this paper.

CONCLUSIONS

The methodology presented here is applied to discrete HRV simulated signals based on Markovian transition probability matrices. However, this methodology can be applied to any other real-life or simulated signal that fits the requirements of a discrete stochastic process, like for example, simulated discrete semi-Markovian models, ARIMA models or biological signals, like the ones derived from brain dynamics and stress hormone levels when sampled discretely.

From our analysis we have derived some valuable insights in terms of what metrics are more convenient and appropriate for what type of analysis, as well as what combination of sample size and number of sessions are most ideal when weighing statistical requirements for data analysis with efforts required in order to obtain such data. From our analysis we conclude that a recording of 200 points over a period of ten (10) sessions just suffices for most purposes of statistical data comparison. However, if a larger confidence is required for smaller CI, then a larger sample size and number of sessions may be more appropriate. The choice of sample size and number of sessions will remain dependent on the study's objectives as well as the available equipment and
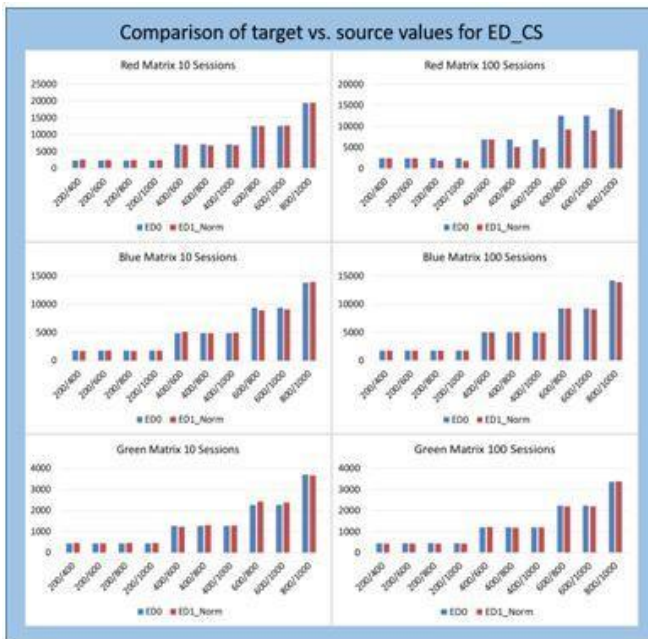
resources. For example, while large sample sizes for many days are relatively easy to obtain when using a more permanent device such as a FIRSTBEAT Bodyguard 2 HRV recorder (Firstbeat Technologies Ltd., 2017), this is very different when using a heartbeat monitor such as the emWave 2 (Quantum Intech, Inc., 2014) which is limited to one (1) hour recordings per session, a relatively small sample size (~400 points). From our observations it seems desirable to find some conventions regarding sample size and sessions across studies since it will greatly simplify the comparison between studies and improve the utilization of data already obtained in previous studies. However, the normalization procedure will be of great value for comparing results when similar sample size across studies is unfeasible.

When we compare the results of the four (4) analyzed metrics in this study, we can draw some conclusions regarding the power of analysis associated to each metric. As we have shown in Figure 6, the Mean_S and its associated mean standard deviations across sessions produces, already with a sample size of 200 and with ten (10) sessions, results that are very close to the expected values. When comparing overall performance of participants Mean_S is an easy metric to compute that may give a very good initial impression about the similarities or differences between processes. This also applies to the ED_S metric which is in close

relation to the Mean_S. However, these two (2) metrics lack the ability to reflect the quality of the process itself when we need to make a distinction based on the type of trajectory with cumulative effects. If we are interested in capturing the nature of the process rather than overall tendencies, then the ED_CS metric is more appropriate. When applying this measure we can derive valuable information reflecting the participants' performance throughout the session. This kind of measure reveals which process is cumulatively better when compared with another one.

For example, when comparing a participant that started a session in red and then slowly recovered throughout the session (to green), with another participant whose process showed the opposite tendency, initially green and then deteriorating to red. The answer to this question is illustrated in Figure 9 and a thorough analysis awaits for future studies.

Finally, we conclude that the CS_final is the least useful of the four (4) metrics analysed in
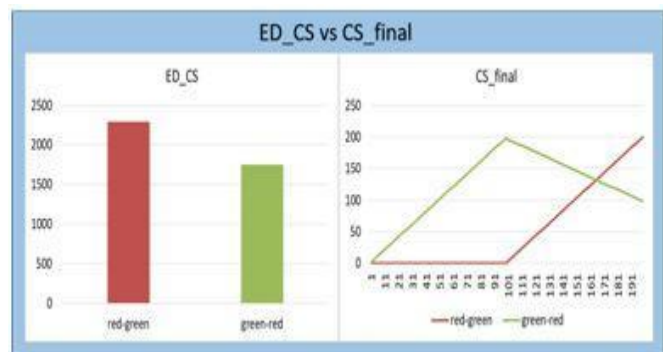
Figure 9

shows a comparison between a participant that starts in red and recovers to green (red line and bar), with another participant showing the opposite tendency, initially green and then deteriorating to red (green line and bar).

this study. This metric fails to capture the process as well as the fact that it may produce values that are non-comparable between participants and sessions. For example, if a participant spends the first half of a session in red and the second half in green, the CS_final computed from such a session will give a noticeably higher CS_final value than if the participant spent the first half in green and then the second half in red, completely ignoring that starting in green may mean more resilience and recovery capacity than starting in red when measuring psychophysiological coherence via HRV measurements. The intuitive minimum requirement would be equal values for both processes, otherwise the CS_final metric indicates that being first in red and then in green is better than vice versa. As already mentioned, it seems to us that the opposite is more biologically plausible, which means that it is better to initially be in green since when dropping we have already built resilience to efficiently address the drop into red. This issue surely awaits more research.

ACKNOWLEDGEMENTS

REFERENCES

1. Allen, A.O., 1978. Probability, Statistics, and Queueing Theory with Computer Science Applications. 1st ed. Orlando, FL: Academic Press.
2. Childre, D. & Cryer, B., 2008. From Chaos To Coherence - The Power To Change Performance. Revised ed. USA: HeartMath LLC.
3. Davis, J.J.J., Day, C. & Schübeler, F., 2019. A Study on the Behaviour of Heart Rate Variability (HRV) with the aid of Markov Chains Theory and Transition Probability Matrices. Journal of Modeling and Simulation. 1(26), pp. 1-19. https://scienchar.sanmed.ca//scienchar/showPage/a-study-on-the-behaviour-of-heart-rate-variability-hrv-with-the-aid-of-markov-chains-theory-and-transition-probability-matrices
4. Davis, J.J.J., Schübeler, F. & Kozma, R., 2019a. Psychophysiological Coherence in Community Dynamics - A Comparative Analysis between Meditation and Other Activities. OBM Integrative and Complementary Medicine, 4(1), pp. 1-24. https://doi.org/10.21926/obm.icm.1901015.
5. Davis, J.J.J. & Schübeler, F., 2019. A Stochastic Process Approach in Modelling the Behaviour of HRV as a Biomarker for Different Cognitive States. Journal of Modeling and Simulation. 1(25), pp. 1-13. https://scienchar.sanmed.ca/scienchar/showPage/a-stochastic-process-approach-in-modelling-the-behaviour-of-hrv-as-a-biomarker-for-different-cognitive-states
6. Firstbeat Technologies Ltd., 2017. Firstbeat Bodyguard 2 Guide. [Online] Available at:

https://www.firstbeat.com/app/uploads/2015/10/150811_Bodyguard2_Guide_ENG.pdf [Accessed 24 December 2018].

7. Fraleigh, J.B. & Beauregard, R.A., 1995. Linear Algebra. 3rd ed. Reading, MA: Addison-Wesley Publishing Company.

8. HeartMath, Inc., 2018. HeartMath Inner Balance Trainer Coherence Scoring System. [Online] Available at: https://www.heartmath.com/support/knowledgebase/?article=kA180000000XbZmCAK&t=HeartMath+Inner+Balance+Trainer+Coherence+Scoring+System [Accessed 24 December 2018].

9. Howard, R.A., 1971. Dynamic Probabilistic Systems, Volume II: Semi-Markov and Decision Processes. Toronto, Canada: John Wiley & Sons, Inc.

10. Law, A.M. & Kelton, W.D., 1991. Simulation Modeling and Analysis. 2nd ed. New York: McGraw-Hill.

11. McCraty, R., 2002. Heart Rhythm Coherence - An Emerging Area of Biofeedback. Biofeedback, 30(1), pp.23-25. https://www.heartmath.org/research/research-library/basic/heart-rhythm-coherence-an-emerging-area-of-biofeedback/.

12. McCraty, R., Atkinson, M. & Tomasino, D., 2001. Science of the Heart: Exploring the Role of the Heart in Human Performance. Boulder Creek, CA: HeartMath Research Center, Institute of HeartMath. https://www.heartmath.org/resources/downloads/science-of-the-heart/.

13. McCraty, R., Atkinson, M., Tomasino, D. & Bradley, R.T., 2006. The Coherent Heart: Heart-Brain Interactions, Psychophysiological Coherence, and the Emergence of System-Wide Order. Boulder Creek, CA: Institute of HeartMath.

14. Medicore, n.d. Heart Rate Variability Analysis System - Clinical Information (Version 3.0). [Online] Available at: http://medi-core.com/download/HRV_clinical_manual_ver3.0.pdf [Accessed 24 December 2018].

15. Quantum Intech, Inc., 2010. emWave Desktop Owner's Manual for PC and Mac. [Online] Available at: https://bio-medical.com/media/support/emwave-desktop-manual.pdf [Accessed 24 December 2018].

16. Quantum Intech, Inc., 2014. emWave 2 Quick Start Guide for PC and Mac. [Online] Available at: http://cdn.heartmath.com/manuals/emWave-2-QSG_3-6-14.pdf [Accessed 24 December 2018].

17. Ross, S.M., 1983. Stochastic Processes. Toronto, Canada: John Wiley & Sons.

18. Tarvainen, M.P. & Niskanen, J.-P., 2012. Kubios HRV version 2.1 User's Guide. Kuopio, Finland: University of Eastern Finland.