Ultra-short entropy for mental stress detection

Rossana Castaldo[1] [0000-0003-4143-1678], Luis Montesinos[1,2] [0000-0003-3976-4190] and Leandro Pecchia[1] [0000-0002-7900-5415]

[1] School of Engineering, University of Warwick, Coventry CV4 7AL, UK
[2] Tecnologico de Monterrey, Campus Ciudad de Mexico, Mexico City, 14380, Mexico
{r.castaldo, l.montesinos-silva, l.pecchia@warwick.ac.uk}

Abstract. Approximate Entropy (*ApEn*) and Sample Entropy (*SampEn*) are measures of signals' complexity and are widely used in Heart Rate Variability (HRV) analysis. In particular, recent studies proved that almost all the features measuring complexity of RR series statistically decreased during the stress and therefore, thus showing ability to detect stress. However, the choice of the similarity threshold *r* and minimum data length *N* required for their computation are still controversial. In fact, most entropy measures are considered not reliable for recordings shorter than 5 minutes and different threshold values *r* have shown to affect the analysis thus leading to incorrect conclusions.

Therefore, the aim of this study was to understand the impact of changing parameters *r* and *N* for the computation of *ApEn* and *SampEn* and to select the optimal parameters to detect stress in healthy subjects. To accomplish it, 84 RR series, extracted from electrocardiography signals acquired during real-life stress, were analyzed. *ApEn* and *SampEn* were estimated for two different values of *r* computed using previously published methods and for N={100, 200, 300, 400, 500) data points. The statistical significance for the differences in mean *ApEn* and *SampEn* values was assessed by non-parametric tests.

The two methods used to compute *r* produced entropy values significantly different over different *N* values. In contrast, *ApEn* and *SampEn* showed consistency in differentiating rest and stress conditions for different input parameters. More specifically, $ApEn_{Chon}$ and $SampEn_{Chon}$ showed to have a better discrimination power between stressed subjects and resting subjects on ultra-short recordings ($N < 500$).

Keywords: Entropy, Heart rate variability, Ultra-short term

1 Introduction

Heart rate variability (HRV), the variation of the time interval between consecutive heartbeats (i.e. R-to-R intervals), is a consequence of the dynamical and complex regulation of the heart rate. Since the overall cardiac response to external stimuli and the related state of the autonomic nervous system can be investigated noninvasively by HRV, a large number of indices to characterize the latter have been developed [1]. In particular, entropy measures have shown great potential for physiological time-series

analysis [2]. Hence, they have been widely used to quantify HRV [3], with the hypothesis that decreasing entropy values reveal perturbations of the underlying physiological mechanisms or disease. Moreover, recent studies have proved that almost all measures of complexity of RR series statistically decreased during stress, therefore were deemed able to detect it [3].

Generally speaking, Approximate Entropy (*ApEn*) and Sample Entropy (*SampEn*) measure the probability that vectors of length *m* built from a time-series of length *N* that are similar within a tolerance range given by ±*r* times the standard deviation of the time-series, remain similar for vector of length *m*+1. Hence, for any fixed *m*, their computation requires the selection of parameters *N* (data length) and r (similarity threshold). The use of *m*=2 has been previously suggested [2, 4]. As for *N*, values normally range between 100 and 5000, whereas for *r* values usually range between 0.1 and 0.25 [2, 4]. However, there are still open questions about the minimal data length (*N*) and the optimal threshold value *r* required to compute *ApEn* and *SampEn* measures. In fact, some studies have shown that *ApEn* values for recordings shorter than 3 minutes are considered unreliable [7, 8]. Additionally, some studies have shown that the selection of *r*, the similarity threshold, is critical in human HRV studies [5, 6]. In this regard, a study recommended that the threshold value *r* is the one that provides the maximum *ApEn* value [5], whereas another study recommended to compute *r* using a formula proposed by its authors [9].

Therefore, this study aimed to understand the impact of changing parameters *N* and *r* for the computation of *ApEn* and *SampEn* and to select the best parameters to detect stress in healthy subjects based on ultra-short recordings ($N < 500$).

2 Methods and Materials

2.1 Data description

Eighty four stationary RR series extracted from electrocardiographic recordings acquired during real-life stress were analyzed. The dataset consisted of 42 students with an age range from 18 to 25 years old. The data were acquired using a commercial electrocardiograph (Easy ECG Pocket. ATES MEDICA Device s.r.l., Verona, Italy), which allows 3-lead clinical research ECG acquisitions, with a sampling frequency of 500 Hz and a resolution of 12 bits. The data were acquired on two different conditions: rest and stress. The stress session was recorded during a university verbal examination. The participants were examined under standard conditions during rest and stress phases: in the same quiet room, at a comfortable temperature, while sitting. From each record, subsequent RR time series of 5-minute length were extracted. A detailed description of the protocol can be found in [3].

2.2 Algorithms for ApEn and SampEn computation

A detailed description of the algorithms for the computation of *ApEn* and *SampEn* can be found elsewhere [3]. Briefly, given a RR time series of length N, such as RR_1, RR_2, …, RR_N, a sequence of vectors of length m: X_1, X_2, …, X_{N-m+1} is constructed as follows:

$X_i=[RR_i, RR_{i+1}, \ldots, RR_{i+m-1}]$. The distance $d[X_i , X_j]$ between vectors X_i and X_j is defined as the maximum absolute difference between their respective scalar components. For each vector X_i, the number of vectors X_j for which $d[X_i , X_j]<r$ is computed as

$$C_i^m(r) = \frac{number\ of\ \{d[X_i, X_j] \le r\}}{N-m+1} \quad \forall j \tag{1}$$

Then, the index $\Phi^m(r)$ is computed by taking the natural logarithm of each $C_i^m(r)$ and averaging them over *i*.

$$\Phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N.-m+1} \ln C_i^m(r) \tag{2}$$

Finally, the approximate entropy is computed as:

$$\mathrm{ApEn(m,r,N)} = \Phi^{m}(r) - \Phi^{m+1}(r) \tag{3}$$

In this study, we computed the *ApEn* for N={100, 200, 300, 400, 500} samples, m=2 and two different values of the threshold r:

- $r = r_{Max}$, that is, the value of r in the interval (0.01 * SDNN, 1.0 * SDNN) which maximizes the *ApEn*;
- $r=r_{Chon}$ that is the value computed according to the formula proposed by Chon [9]:

$$r_{Chon} = (-0.036 + 0.26\sqrt{SDDS/SDNN}) / \sqrt[4]{N/1000} \tag{4}$$

where SDDS and SDNN are the short-term and long-term variability of the RR sequence, respectively. Formally, SDDS is the standard deviation of the difference sequence of the series RR, that is, $[RR_{i+1} - RR_i, RR_{i+2} - RR_{i+1}, \ldots, RR_N - RR_{N-1}]$, and; SDNN is the standard deviation of the RR series.

To compute *SampEn*, $C_i^m(r)$ is computed as reported in equation 5, $\Phi^m(r)$ as reported in equation 2 and finally *SampEn* as in equation 6.

$$C_i^m(r) = \frac{number\ of\ \{d[X_i, X_j] \le r\}}{N-m+1} \quad \forall j \ne i \tag{5}$$

$$SampEn(m,r,N) = \log \frac{\Phi^m(r)}{\Phi^{m+1}(r)} \tag{6}$$

Note that *ApEn* and *SampEn* differ in that the latter does not take into account vector self-matches. Additionally, the dependence on the parameter *r* is different: *SampEn* decreases when increases. On the other hand, it has been shown that *SampEn* and *ApEn* often provide comparable results for large values of N and r [10].

2.3 Statistical analysis

Since a previous study showed that *ApEn* and *SampEn* did not follow normal distribution [11], the following descriptive statistics were computed: median (MD), standard deviation (SD), and the 25th and 75th percentiles. The statistical significance of the differences in median values estimated using the two methods to compute r for N={100, 200, 300, 400, 500} were assessed by a non-parametric statistical test (namely, the Wilcoxon signed rank test). Moreover, Spearman's correlation coefficients (rho) and their associated p-values (p_{rho}) were computed between the estimates of *ApEn* and *SampEn* varying N and r for rest and stress. The differences between *ApEn* and *SampEn* values for different N and r were also investigated to assess whether *Apen* and *SampEn* calculated for different N and r could discriminate between rest and stress conditions.

In-house Matlab scripts were used to compute *ApEn* and *SampEn* and perform the statistical analysis.

3 Results

Table 1 and 2 show summary statistics for *ApEn* computed for N={100, 200, 300, 400, 500} and r={r_{Chon}, r_{Max}} during rest and stress, respectively. Moreover, Table 1 and 2 also report the p-values calculated using Wilcoxon signed rank and the Spearman's correlation coefficient (rho). Statistically significant differences ($p<0.001$) were observed between the $ApEn_{Max}$ and $ApEn_{Chon}$, as shown in Table 1 and 2 for rest and stress respectively. These results were supported by rho values below 0.7, which demonstrate a very low correlation. Moreover, Fig. 1 shows the median and standard deviation for $ApEn_{Max}$ and $ApEn_{Chon}$ during rest and stress, over different N values with m=2.

Table 1. *ApEn* during rest computed for N={100, 200, 300, 400, 500} and r={r_{Chon}, r_{Max}}

	r_{Chon}				r_{Max}				r_{Chon} VS r_{Max}	
N	MD	SD	25th	75th	MD	SD	25th	75th	p-value	rho
100	0.350	0.901	0.001	0.278	0.323	0.088	0.260	0.381	<0.001	0.232*
200	0.820	1.341	0.003	1.161	0.538	0.098	0.477	0.599	<0.001	0.070*
300	1.301	1.692	0.042	2.411	0.730	0.091	0.685	0.790	<0.001	-0.108*
400	1.821	1.953	0.212	2.642	0.897	0.101	0.831	0.931	<0.001	-0.283*
500	1.896	1.954	0.219	2.662	0.896	0.101	0.836	0.931	<0.001	-0.283*

* $p_{rho} < 0.05$

Table 2. *ApEn* during stress computed for N={100, 200, 300, 400, 500} and r={r_{Chon}, r_{Max}}

	r_{Chon}				r_{Max}				r_{Chon} VS r_{Max}	
N	MD	SD	25th	75th	MD	SD	25th	75th	p-value	rho
100	0.021	0.131	0.000	0.002	0.398	0.142	0.322	0.496	<0.001	-0.075*
200	0.305	0.912	0.012	0.06	0.609	0.124	0.510	0.691	<0.001	0.195*
300	0.523	1.070	0.044	0.376	0.772	0.123	0.671	0.854	<0.001	0.151*
400	0.716	1.532	0.021	0.766	0.853	0.125	0.759	0.920	<0.001	-0.019*
500	0.895	1.586	0.21	1.041	0.954	0.101	0.905	1.016	<0.001	0.027*

* $p_{rho} < 0.05$

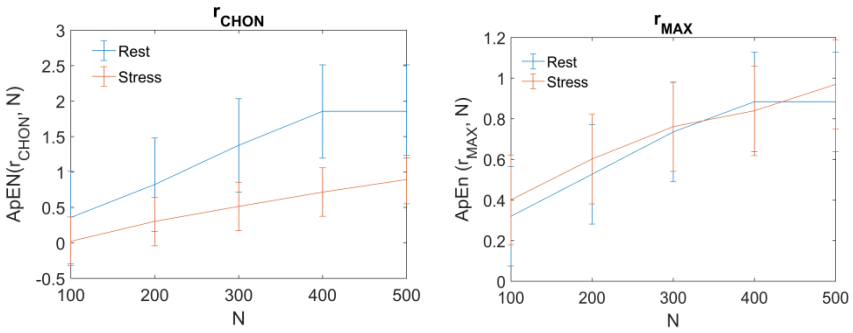

Fig. 1. *ApEn*$_{Max}$ and *ApEn*$_{Chon}$ for N={100, 200, 300, 400, 500} with m=2. Error bars represent the standard deviation.

The same analysis was run also for *SampEn*. Table 3 and 4 show summary statistics for *SampEn* evaluated for N={100, 200, 300, 400, 500} and r={r_{Chon}, r_{Max}} during rest and stress, respectively. Moreover, Table 3 and 4 also report the p-values calculated using Wilcoxon signed rank and the Spearman's correlation coefficient (rho). Statistically significant differences ($p<0.001$) were observed between the *SampEn*$_{Max}$ and *SampEn*$_{Chon}$, as shown in Table 1 and 2 for rest and stress respectively. However, rho showed to be above 0.7 highlighting a correlation between *SampEn*$_{Max}$ and *SampEn*$_{Chon}$ over different N values.

Fig. 2 shows the median and standard deviation of *SampEn*$_{Max}$ and *SampEn*$_{Chon}$ during rest and stress over different N values with m=2.

Table 3. *SampEn* during rest computed for N={100, 200, 300, 400, 500} and r={r_{Chon}, r_{Max}}

	r_{Chon}				r_{Max}				r_{Chon} VS r_{Max}	
N	MD	SD	25th	75th	MD	SD	25th	75th	p-value	rho
100	1.130	0.203	1.005	1.250	2.134	0.398	1.788	2.319	<0.001	0.716*
200	1.290	0.166	1.169	1.394	2.120	0.361	1.907	2.386	<0.001	0.861*
300	1.403	0.156	1.321	1.497	2.082	0.266	1.859	2.284	<0.001	0.621*
400	1.447	0.142	1.319	1.542	2.075	0.281	1.847	2.294	<0.001	0.704*
500	1.457	0.148	1.329	1.543	2.073	0.283	1.848	2.294	<0.001	0.704*

* $p_{rho} < 0.05$

Table 4. *SampEn* during stress computed for N={100, 200, 300, 400, 500} and r={r_{Chon}, r_{Max}}

	r_{Chon}				r_{Max}				r_{Chon} VS r_{Max}	
N	MD	SD	25th	75th	MD	SD	25th	75th	p-value	rho
100	0.845	0.292	0.668	1.059	1.876	0.503	1.557	2.174	<0.001	0.490*
200	0.962	0.327	0.683	1.171	1.898	0.425	1.595	2.247	<0.001	0.702*
300	0.946	0.301	0.746	1.148	1.777	0.469	1.477	2.131	<0.001	0.806*
400	1.028	0.308	0.833	1.236	1.785	0.466	1.490	2.073	<0.001	0.769*
500	1.068	0.343	0.852	1.347	1.692	0.503	1.385	2.180	<0.001	0.854*

* $p_{rho} < 0.05$

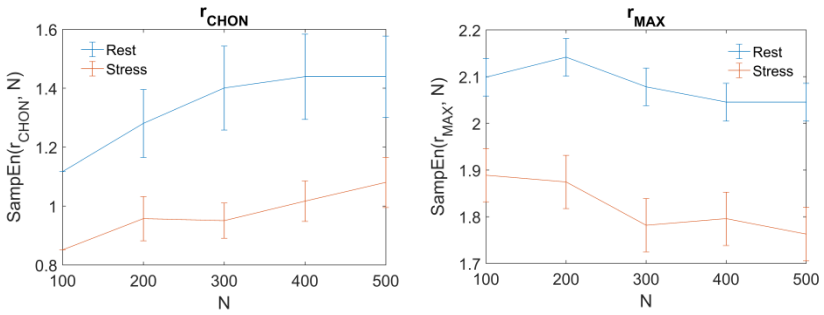

Fig. 2. *SampEn*$_{MAX}$ and *SampEn*$_{Chon}$ over different *N* values with m=2. Error bars represent the standard deviation.

Table 5 presents the p-values for differences in *ApEn*$_{Chon}$, *ApEn*$_{Max}$, *SampEn*$_{Chon}$ and *SampEn*$_{Max}$ values between rest and stress conditions for different lengths N. *ApEn*$_{Chon}$ showed ability to discriminate between rest and stress for N={200, 300, 400}. *ApEn*$_{Max}$ could not discriminate between rest and stress conditions for N < 500. *SampEn*$_{Chon}$ and *SampEn*$_{Max}$ showed discriminative power between rest and stress conditions for all data lengths analyzed in this study.

Table 5. Wilcoxon signed rank test between Rest and Stress for *ApEn*$_{Chon}$, *ApEn*$_{Max}$, *SampEn*$_{Chon}$ and *SampEn*$_{Max}$

	***ApEn*$_{Chon}$**	***ApEn*$_{Max}$**	***SampEn*$_{Chon}$**	***SampEn*$_{Max}$**
N	**p-value**	**p-value**	**p-value**	**p-value**
100	0.001	0.120	<0.001	0.031
200	<0.001	0.180	<0.001	0.004
300	<0.001	0.254	<0.001	0.002
400	<0.001	0.088	<0.001	0.006
500	0.002	<0.001	<0.001	0.004

4 Discussion and conclusion

In this paper, we reported the methods and results of an analysis performed on 84 RR series to assess the appropriateness of using two different values of the parameter r, namely r_{Chon} and r_{max}, for the computation of *ApEn* and *SampEn* on ultra-short HRV time series. *ApEn*$_{Chon}$ was significantly different from the *ApEn*$_{Max}$ over different N for both rest and stress conditions. These findings were consistent with those of previous studies on smaller time series (N=120) [6, 11] and larger time series (N=500) [12]. On the other hand, *SampEn*$_{Chon}$ was significantly different from but highly correlated to *SampEn*$_{Max}$ over different N for both rest and stress conditions. These results make evident that entropy values computed using different *r* parameter values should be carefully compared.

Additionally, the $ApEn_{Chon}$ and $SampEn_{Chon}$ appeared to be able to discriminate better than $ApEn_{Max}$ $SampEn_{Max}$ between rest and stress in ultra-short recordings (N < 500). Consequently, this may lead to the conclusion that the $ApEn_{Chon}$ and $SampEn_{Chon}$ have a good discrimination power in distinguishing stressed subjects from resting subjects.

Acknowledgment.
R.C. thanks the Institute of Advanced Study Early Career Fellowship at University of Warwick, UK.

Conflict of Interest.
The authors declare no conflict of interest.

References

1. Force, T., Heart rate variability guidelines: Standards of measurement, physiological interpretation, and clinical use. European Heart Journal, 1996. 17: p. 354-381.
2. Pincus, S., Approximate entropy as a measure of system complexity. Proc Natl Acad Sci USA, 1991. 88: p. 2297 - 2301.
3. Melillo, P., M. Bracale, and L. Pecchia, Nonlinear Heart Rate Variability features for real-life stress detection. Case study: students under stress due to university examination. Bio-Medical Engineering OnLine, 2011. 10(1): p. 1-13.
4. Pincus, S. and A. Goldberger, Physiological time-series analysis: what does regularity quantify? Am J Physiol, 1994. 266: p. H1643 - H1656.
5. Lu, S., et al., Automatic selection of the threshold value r for approximate entropy. IEEE Transactions on Biomedical Engineering, 2008. 55(8): p. 1966-1972.
6. Castiglioni, P. and M. Di Rienzo. How the threshold "r" influences approximate entropy analysis of heart-rate variability. in Computers in Cardiology, 2008. 2008.
7. Salahuddin, L., et al., Ultra short term analysis of heart rate variability for monitoring mental stress in mobile settings. Conf Proc IEEE Eng Med Biol Soc, 2007. 2007: p. 4656-9.
8. Yentes, J.M., et al., The appropriate use of approximate entropy and sample entropy with short data sets. Annals of biomedical engineering, 2013. 41(2): p. 349-365.
9. Chon, K.H., C.G. Scully, and S. Lu, Approximate Entropy for all Signals Is the Recommended Threshold Value r Appropriate? Ieee Engineering in Medicine and Biology Magazine, 2009. 28(6): p. 18-23.
10. Rajendra Acharya, U., et al., Heart rate variability: a review. Med Biol Eng Comput, 2006. 44(12): p. 1031-51.
11. Liu, C.Y., et al., Comparison of different threshold values r for approximate entropy: application to investigate the heart rate variability between heart failure and healthy control groups. Physiological Measurement, 2011. 32(2): p. 167-180.
12. Melillo, P., et al. Approximate Entropy for Short-Term Heart Rate Variability Analysis: Is the Threshold Value Computed by Chon's Formula Appropriate? in The International Conference on Health Informatics. 2014. Springer.