Approximate Entropy for Heart Rate Variability: Effect of Tolerance and Data Length

Jae Mok Ahn[1,*], Jeom Keun Kim[1]

[1] *Professor, College of Software, Hallym University, 1 Hallymdaehak-gil, Chuncheon-si, Gangwon-do, 24252 South Korea.*

[*]*Corresponding author*

ORCID: 0000-0002-1061-217X (Jae Mok Ahn)

ABSTRACT

Heart rate variability (HRV) can describe the clinical significance of the autonomic nervous system (ANS) through the calculation and combination of various HRV parameters. One of those parameters is approximate entropy (ApEn), which represents the similarity and complexity of heart rate rhythms in a time series. The ApEn provides quantitative information about similar patterns within short-term HRV data recordings. However, in calculating ApEn, the effect of the selection of r, the tolerance threshold for similarity, or of the data length is not well understood. To ensure consistency of the ApEn, a minimum data length, approximate dimensions, and the criteria for the tolerance value r must be determined appropriately. Thus, in this study three different tolerances (r= 0.1, 0.2, and 0.5 times both the standard deviation of normal-to-normal heartbeat (SDNN) and the root mean square of the successive differences (RMSSD) of the HRV dataset) and five data lengths were used to investigate variations of the ApEn parameters. The results showed that when either the data length or the tolerance increased, the ApEn value increased when both SDNN and RMSSD were used as a function of the tolerance threshold. Therefore, it is concluded that both the tolerance and data length should be carefully determined in performing irregularity analysis of HRV.

Keywords: Approximate entropy, heart rate variability, similarity, autonomic nervous system

1. INTRODUCTION

Heart rate variability (HRV) has been applied to analyze the physiological activity reflecting cardiovascular control by the autonomic nervous system (ANS) [1-3]. HRV is regulated by both the sympathetic and parasympathetic branches linked to sinoatrial node of the heart [4]. The ANS has been extensively implicated in the causes of sudden death by relative sympathetic dominance [5-6]. HRV consists of changes in the time intervals between successive heartbeats, called normal-to-normal (NN) intervals. The HRV patterns of a healthy heart are constantly changing to rapidly adjust to abrupt psychological and physiological challenges due to homeostasis, the self-regulating processes by which biological systems tend to maintain stability. The oscillations of a normal heartbeat in a time series can be described by mathematical expressions [7]. While healthy individuals show complexity and irregularity in their HRV rhythms, cardiovascular-related disease can involve either a loss or excessive increase of complexity [8]. HRV measures for analyzing the HRV time domain include the standard deviation of all NN intervals (SDNN), the root mean square of the sum of all differences between successive NN intervals (RMSSD), the percentage of successive intervals that differ by more than 50 ms (pNN50), and the standard deviation of the differences between successive NN intervals (SDSD). HRV parameters in the frequency domain comprise parametric and nonparametric power spectral estimation to assess very low frequency (VLF), low frequency (LF), and high frequency (HF) power bands [9-10]. In particular, a nonlinear analysis statistical method using a time series HRV dataset is used to analyze HRV complexity through approximate entropy (ApEn). Subjects have shown a higher complexity value in the awake state than in the sleep state [11-12]. Nonlinear ApEn values in the HRV analysis show that clinical indicators of the similarity in HRV patterns beginning at the sinoatrial node could be effectively investigated and detected through complexity thresholding, whereas Fourier transform (FT) technology has shown no meaningful information on investigating the regularity reflecting heart disease [11]. Therefore, the ApEn as a measure of complexity is more efficient for analyzing HRV parameters than FT technology. A statistic that quantifies the similarity of an HRV dataset reflects clinical information on ANS-related physiological activity by analyzing relatively short data recordings. To calculate the ApEn value of a heartbeat time series, an HRV dataset is first rearranged into a series of dimensional vectors; then m, the distances between two corresponding data points, is calculated; r, the tolerance threshold representing the similarity value for comparing reconstructed vectors is determined, and finally the similarity parameter for the reconstructed vectors under a given r value is obtained. Among these procedures, choosing the r from the range of 0.1 to 1.0 times the SDNN is a time-sonsuming process. Different threshold values for ApEn calculation were used to investigate the HRV between heart failure and healthy control groups [13]. The tolerance threshold value for accepting similar patterns was reported to be between 0.1 and 0.25 times the SDNN. and the dimension vector, m=2, determined the subsequent lengths to be compared to calculate the complexity of the HRV recording [14-15]. ApEn should be used with caution since different tolerance values could affect the estimation of the parameters of a stationary HRV time series and provide incorrect conclusions if r was selected inappropriately [16]. Some studies have shown that the selection of r in the estimation of ApEn values for short-term recordings is critical in HRV studies [17-18]. Therefore, the aim of this study was to understand the impact of changing the parameters r and the data length on the calculation of ApEn and to provide their optimal criteria to evaluate the activity of ANS as the ApEn value in healthy subjects. To implement this analysis, an HRV dataset obtained from photoplethysmogram

(PPG)-based measurement was analyzed. The ApEn was estimated for three different similarity thresholds, r=0.1, 0.2, 0.5 and for different short-term data lengths: 30s, 1min, 2min, 2min30s, and 5min recordings.
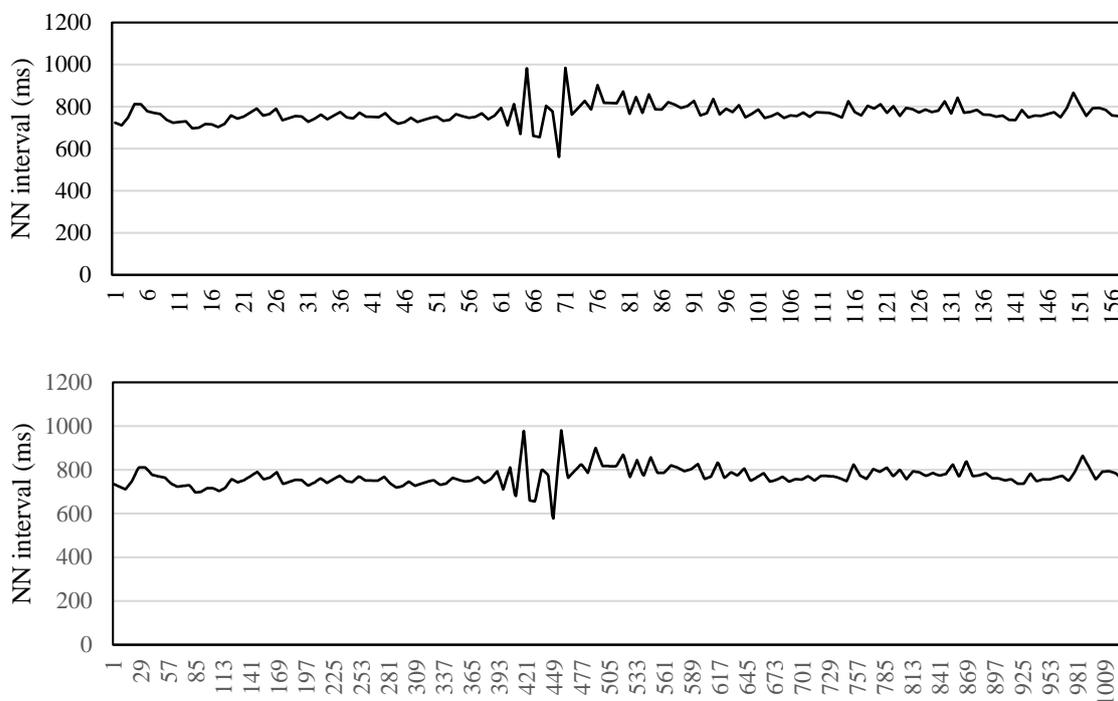
2. INTERPOLATION

We applied a linear interpolation method to heartbeats expressed as a uniform sequence of durations between two consecutive beats. This resampling method transforms a discrete signal in which sampling intervals along the x-axis are of different lengths to a discrete signal with evenly spaced sampling intervals. Resampling increases the number of points in the original dataset to improve its performance for HRV analysis in the frequency domain and time domain. The new discrete time sequence of values was derived at a resampling frequency of 3.3 Hz. Regardless of the length of HRV dataset measured, 1024 data points were created after resampling. Resampling was applied prior to HRV signal processing to obtain reliable HRV parameters. Linear interpolation for resampling draws a straight line passing through x_1 and x_2 to obtain some x value that is between x_1 and x_2. A y value was found on the line for the selected x through the following formula.

$$y = y_1 + (x - x_1)\frac{y_2 - y_1}{x_2 - x_1} \tag{1}$$

Some studies have demonstrated that resampling with linear interpolation is efficient and better than nonlinear interpolations for predicting missing values in the HRV dataset in a time series [19-20]. Figure 1 shows a representation of the beat-to-beat variability of each cardiac cycle with both axes representing the time between beats. The tachogram in Figure 1 (top) is inherently a discrete, uneven time series. However, every FT parameter and ApEn value require evenly sampled data. Thus, in this study, using linear interpolation, a resampled tachogram was obtained with regularly sampled data points in discrete time signal as shown in Figure 1 (bottom).

Figure 1. Resampling data to convert nonuniformly spaced NN intervals to uniformly spaced NN intervals in a time series for a short-term dataset: (top) a tachogram of 157 data points before linear interpolation and (bottom) a tachogram of 1024 data points after linear interpolation.

3. APPROXIMATE ENTROPY

The ApEn method is a mathematical calculation that represents the measure of complexity or similarity in the HRV dataset. Using the ApEn, HRV complexity can be evaluated over a relatively short time series for the analysis of ANS activity. Several studies have reported sickness and aging with significantly decreased ApEn values that reflect more regularly patterned HRV recordings [21-22]. The higher the values of ApEn, the more irregularities there will be and vice versa. For the calculation of ApEn, two parameters, m and r, must be fixed throughout the entire computation. M, which is defined as the embedding dimension is the value of the vector size for comparing selected segments of NN intervals. In a given HRV dataset with N total number of sample data points, an array vector of N-m+1 arrays consisting of m components each is created as follows;

$$NN(1) = \{nn(1), nn(2), \dots, nn(m)\}$$
$$NN(2) = \{nn(2), nn(3), \dots, nn(m + 1)\}$$

......

$$NN(N - m + 1) = \{nn(N - m + 1), nn(N - m + 2), \dots, nn(N)\} \tag{2}$$

$NN(N-m+1)$ represents a sequence of m consecutive NN interval values, nn. In Equation (2), each of the vectors is composed of m discrete data points in the time series. The distance between two vectors NN(i) and NN(j) can be defined as the maximum difference in their respective corresponding elements, as in Equation (3),

$$d[NN(i), NN(j)] = \max_{k=1,2,\dots,m} (|nn(i+k-1) - nn(j+k-1)|) \qquad (3)$$

where i=1, 2, …, N-m+1, j=1, 2, …, N-m+1, and N is a total number of HRV data points. Two vectors, NN(i) and NN(j), are similar if the distance, $d[NN(i), NN(j)]$, is less than r, a predetermined tolerance value defined in Equation (4-5),

$$r = k * SDNN \text{ or } k * RMSSD \qquad (4)$$

$$\alpha = \sum_{i \neq j} \Delta\{r - d[NN(i), NN(j)]\} \qquad (5)$$

where

$$\Delta\{x\} = \begin{cases} 1, x \geq 0, \\ 0, x < 0. \end{cases} \qquad (6)$$

The factor, k was selected among 10%, 20% to 50%, then multiplied to the SDNN and RMSSD calculated from the HRV dataset. The probability that describes the similarity between the vector NN(i) and all other vectors NN(j) can be constructed as

$$C_i^m(r) = \frac{1}{N-(m-1)} \alpha \qquad (7)$$

where $C_i^m(r)$ is the probability of finding a sequence of m heartbeats similar to the sequence NN(i). Therefore, the ApEn value of an infinite time series can be calculated as

$$ApEn(r,m) = \lim_{N \to \infty} \left[ln \frac{C_i^m(r)}{C_i^{m+1}(r)} \right]. \qquad (8)$$

For practical applications, a finite time series with N HRV data points can be defined as

$$ApEn(r,m,N) = ln \frac{C_i^m(r)}{C_i^{m+1}(r)}. \qquad (9)$$

ApEn can be thought of as a biased estimator that describes the similarity or complexity of patterns in a sequence of length N. A high degree of similarity means that discrete time data that are similar for m points are highly expected to be similar for the next m+1 point. The ApEn method is relatively easy to be performed and has been widely applied to numerous clinical cardiovascular studies. A time series data with sequences that are more similar have smaller ApEn values. Thus, the greater the regularity or similarity, the lower the ApEn value is. To ensure consistency of the ApEn estimation, a minimum data length N, embedding dimension m and tolerance value r must be determined approximately. To obtain the optimal data length, the ApEn value was calculated in a periodic time series with a single frequency component, resulting in a value close to zero due to the high similarity within the dataset, as shown in Figure 2.
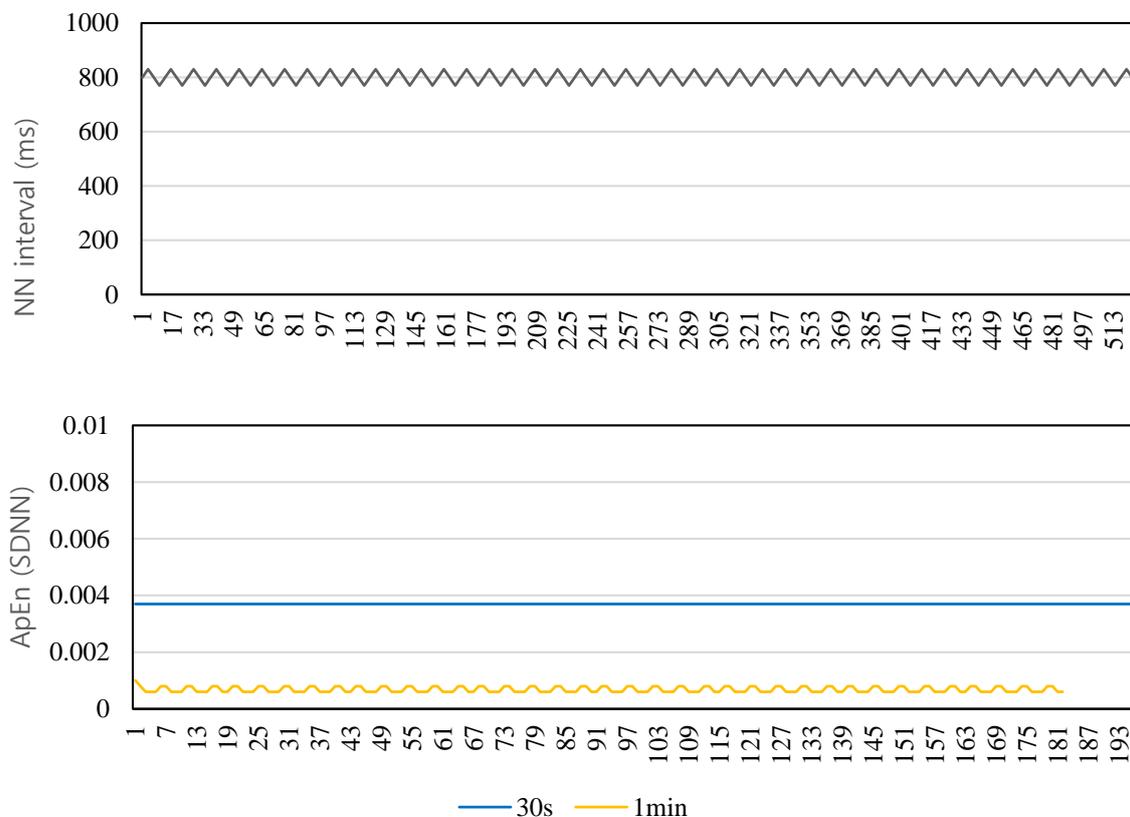
Figure 2. The ApEn of a discrete time signal with a single frequency component was calculated with respect to different lengths of data points at r=0.2*SDNN.

4. SPECTRAL ANALYSIS

Power spectral analysis for the HRV dataset was analyzed through fast Fourier transform (FT), which is relatively simple and requires little computational power. FT estimates HRV parameters in the frequency domain by analyzing the spectrum of short-term recordings from 30 s-5 min. For analysis of HRV frequency domain parameters, the strategy for the original 8-minute HRV recording was to divide this original dataset into five consecutive 30 s-5 min blocks, as shown in Figure 3. The individual HRV parameters of all these blocks were obtained by shifting each block forward. The frequency domain parameters ln HF and ln LF were calculated in real time by using the following equations:

$$ln \int_a^b |X(f)^2| df \qquad (10)$$

$$ln \int_c^d |X(f)^2| df \qquad (11)$$

where low frequency power (LF) is the frequency activity between a=0.04 and b=0.15 Hz, and high frequency power (HF) is the frequency activity between c=0.15 and d=0.1 Hz. The power spectrum, X(f) , was calculated as the squared magnitude of the fast FT of the HRV dataset in our previous study [23].

5. PROCESSING SCHEME

The ApEn calculation algorithm was embedded in the commercial TAS9VIEW (or CANOPY9 RSA) pulse analyzer (IEMBIO Co., Ltd., Chuncheon-si, Korea). In TAS9VIEW's research mode, all HRV parameters, including the ApEn value, were analyzed by shifting each block of five different data lengths (30 s, 2 min, 2 m 30 s, and 5 min) forward by 2 s during the entire 8-minute HRV tachogram. A block refers the moving window used for subsequent analyses of the HRV data points. All results were automatically stored in an Excel file and used for the comparison. The processing scheme is displayed in Figure 3. The reason that we used different short-term segments of the HRV dataset was to determine an appropriate length of HRV data for medical application of the ApEn result. To collect the original 8-minute HRV dataset, the participant was seated in a quiet room and was not allowed to talk or move while the measurement was carried out. In this study, fingertip PPG recordings were obtained, and PPG signals were sampled at 1000 samples s^{-1}.
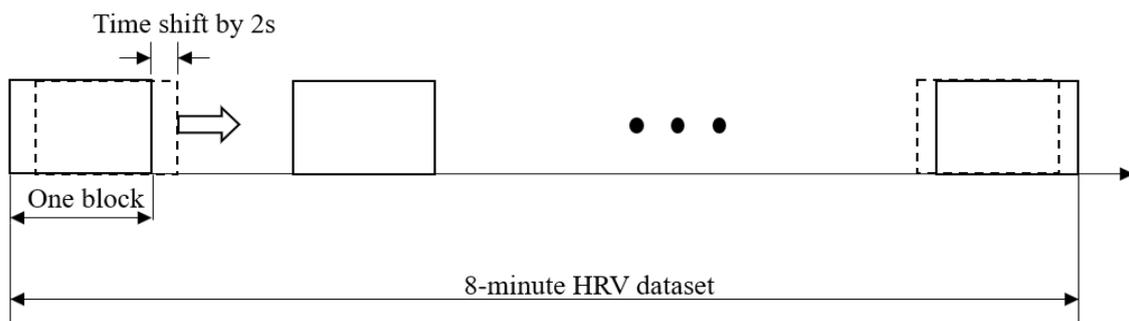

Figure 3. The processing scheme for ApEn analysis using time shifts in TAS9VIEW's research mode.

## 6. RESULTS

The more similar the HRV data is, the smaller the ApEn value, as shown in Figure 2. A sinusoidal signal with a single frequency was to determine the approximate data length for calculating the ApEn for a regular pattern; the ApEn values calculated were 0.0037 for a 30 s data length, 0.0008 for 1 min, and close to 0 for data lengths of 2 min, 2 m 30 s, and 5 min. Because the pure sinusoid with a data length of more than 1 minute had an ApEn value close to 0, at least 1-minute of data length is required to calculate the ApEn value. It is important to note that the test frequency of the sinusoid was approximately 0.2 Hz when the ApEn value of approximately 0 was obtained. Figure 4 shows how the ApEn of the HRV data fluctuates in terms of data length. There was no significant difference between ApEn (SDNN) and ApEn (RMSSD) when the tolerance was calculated with k=0.2 (or 20%). Except for the 5-minute data length, more fluctuations can be observed in the ApEn magnitude. This demonstrates how data length could affect the change of ApEn magnitude. In Figure 5, the spectral profiles (top, ln HF; bottom, ln LF) of the entire 8-minute global dataset were obtained through shifting local segments of the HRV dataset with 30 s-5 min blocks. The very short data length, 30 s is characterized by a low fluctuation in the HF component and a high fluctuation in the LF component, representing fast responses to environmental challenges. To analyze the effect of tolerance threshold and data length, the values of ApEn for three tolerances, e.g., when r=0.1, r=0.2, and r=0.5, and the five data lengths are illustrated in Figure 6. The ApEn value should not be used with a low tolerance and a data length of less than 1-minute in clinical application. Specifically, the use of a high threshold (r=0.5) leads to much larger complexity than use of r=0.1 or r=0.2 at the same data length. Meanwhile, the use of a 5 min data length at r=0.5 showed a lower complexity than that of a 2 m 30s data length at the same tolerance. In Table 1, the average results of the ApEn values show that a selection of r=0.1 for a 2 m 30 s (0.3571) data length resulted in a 76% decrease in the ApEn value compared to a 5 min data length (0.6305), whereas r=0.5 resulted in a 2% increase when r was calculated as a function of SDNN.
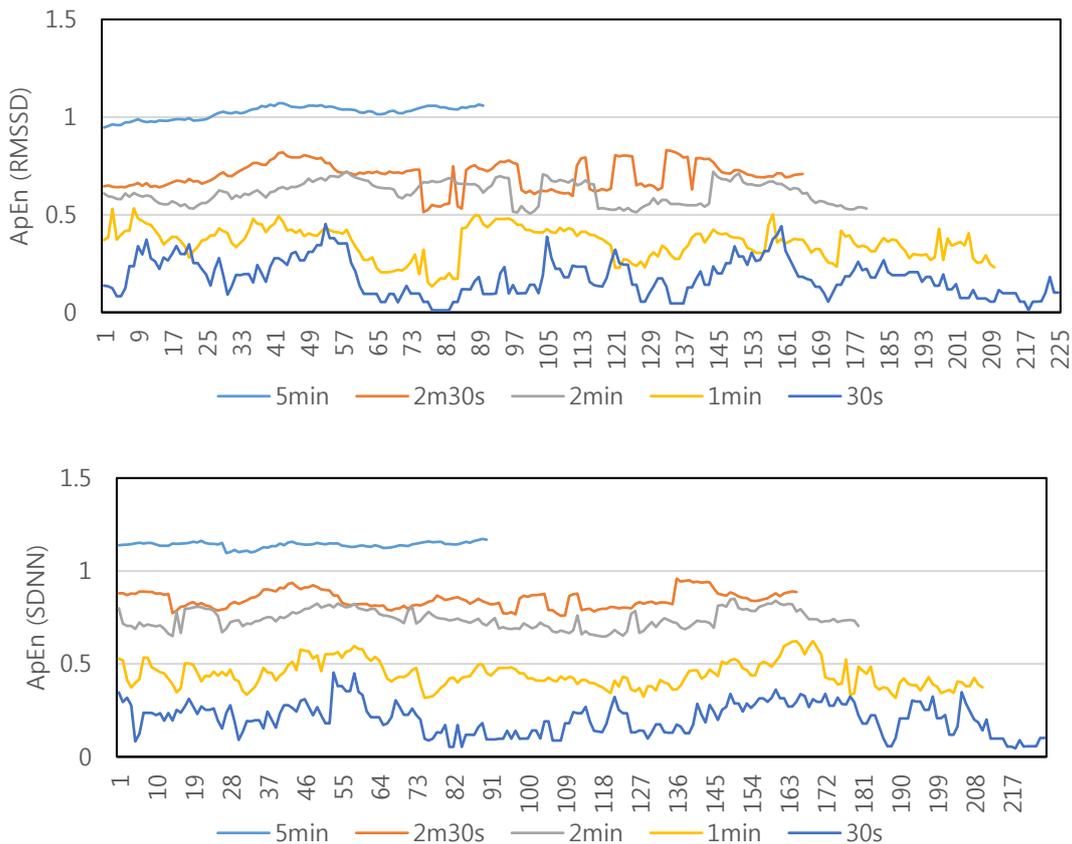
Figure 4. Fluctuations of ApEn for different thresholds and data lengths in terms of SDNN and RMSSD.
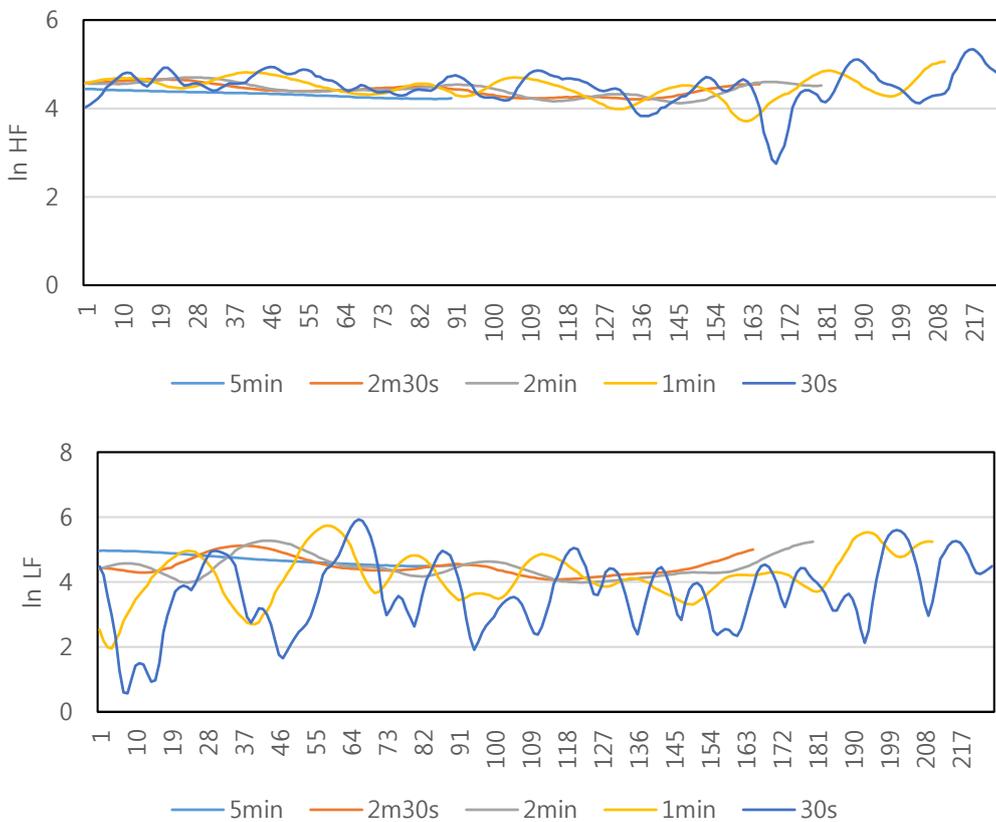
Figure 5. Fluctuations of ln HF and ln LF for five different data lengths
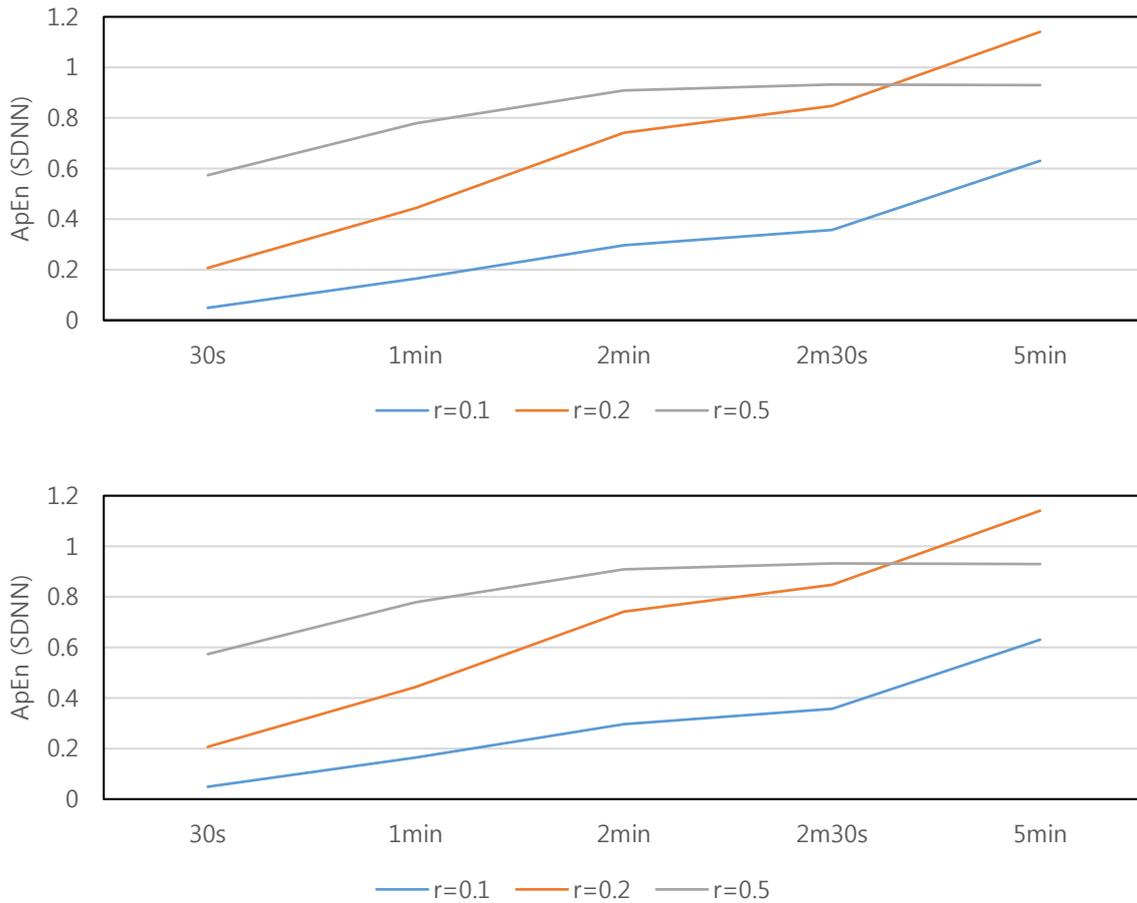
Figure 6. ApEn calculated for different tolerances in terms of SDNN and RMSSD.

Table 1. ApEn values in terms of r and data length for the entire 8-minute recordings of the HRV data: (top) SDNN, and (bottom) RMSSD.

Data length	SDNN		r=0.1		r=0.2		r=0.5	
	Mean	STDEV	Mean	STDEV	Mean	STDEV	Mean	STDEV
30 s	23.73	5.15	0.0491	0.064	0.2066	0.0854	0.5734	0.0819
1 min	25.32	3.43	0.1647	0.0605	0.4438	0.0703	0.7787	0.0507
2 min	26.55	2.01	0.2967	0.0538	0.7416	0.0502	0.9089	0.0414
2 m 30 s	26.78	2.01	0.3571	0.0588	0.8474	0.0454	0.9320	0.0373
5min	29.21	1.30	0.6305	0.0965	1.1406	0.0154	0.9298	0.0434

Data length	RMSSD		r=0.1		r=0.2		r=0.5	
	Mean	STDEV	Mean	STDEV	Mean	STDEV	Mean	STDEV
30 s	21.01	2.56	0.0385	0.0651	0.1792	0.0948	0.527	0.0946
1 min	20.92	1.58	0.1150	0.0629	0.3559	0.0819	0.7603	0.0532
2 min	20.75	0.93	0.2415	0.0690	0.6131	0.0580	0.9725	0.0548
2 m 30 s	20.70	0.77	0.2882	0.0907	0.7021	0.0709	1.0207	0.0548
5 min	20.67	0.30	0.5555	0.0263	1.0238	0.0326	1.1122	0.0291

7. CONCLUSIONS

In calculating the ApEn value, the tolerance threshold, r and the HRV data length have been shown to be important predetermining factors for determining an effective measure for indicating the activity of the cardiovascular system. Our results have shown that the similarity or regularity in short-term HRV recordings more than 1-minute long can be effectively identified through an ApEn value approaching zero in a pure sinusoidal signal. However, for a very short (30 s) data length with a single frequency, the ApEn value was 0.0037, not 0. The ApEn values were also shown to be sensitive to changes in data length and tolerance. The ApEn value increased by approximately 100% when the tolerance r increased from 0.1 to 0.5 time either SDNN or RMSSD using 5-minute data length, while the ApEn value increased by approximately 1200% and 300% for a 30 s data length and 1 min data length, respectively. Thus, an ApEn decrease was correlated with a decrease in data length and with a decrease in tolerance. There was no significant difference between SDNN and RMSSD in the change of ApEn magnitude for different tolerance variable. The standard deviations for all ApEn values averaged over all tolerances showed no appreciable change across data length. The results indicate that the similarity of the HRV dataset is higher for very short data lengths (less than 1-minute) than for longer data lengths and is higher for lower r values. A predetermined r significantly affected the ApEn value compared to data length. These results emphasized that the selection of r is critical for this kind of regularity analysis because an inappropriate selected r can lead to misleading conclusions about the similarity of the HRV dataset related to ANS imbalance.

ACKNOWLEDGEMENTS

REFERENCES

[1]   Task Force of the European Society of Cardiology the North American Society of Pacing Electrophysiology. Heart Rate Variability: standards of measurement, physiological interpretation, and clinical use. Circulation, 93, 1043-1065 (1996).

[2]   Malliani, A. Cardiovascular neural regulation explored in the frequency domain. Circulation, 84, 482-492 (1991).

[3]   Pomeranz, B., Macaulay, R.J., Caudill, M.A., Kutz, I., Adam, D., Gordon, D., Kilborn, K.M., Barger, A.C., Shannon, D.C., Cohen, R.J. Assessment of autonomic function in humans by heart rate spectral analysis. The American Journal of Physiology, 248(1), H151-H151 (1985).

[4]   Evans, S., Seidman, L.C., Tsao, J.C.I., Lung, K.C., Zeltzer, L.K., Naliboff, B.D. Heart rate variability as a biomarker for autonomic nervous system response differences between children with chronic pain and healthy control children. Journal of Pain Research, 6, 449-457, (2013).

[5]   Schwartz, P.J., La Rovere, M.T., Vanoli, E. Autonomic nervous system and sudden cardiac death. Experimental basis and clinical observations for post-myocardial infarction risk stratification. Circulation, 85(1), 177-191 (1992).

[6]   Verrier, R.L., Tan, A. Heart rate, autonomic markers, and cardiac mortality. Heart Rhythm, 6(11), S68-S75 (2009).

[7]   Goldberger, A.L. Is the normal heartbeat chaotic or homeostatic? News In Physiological Sciences, 6, 87-91 (1991).

[8]   Vaillancourt, D.E., Newell, K.M. Changing complexity in human behavior and physiology through aging and disease. Neurobiology of Aging, 23(1), 1-11 (2002).

[9]   Porta, A., Di Rienzo, M., Wessel, N., Kurths, J. Addressing the complexity of cardiovascular regulation. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 367(1892), 1215-1218 (2009).

[10]   Goldberger, A.L, Peng, C.K., Lipsitz, L.A. What is physiologic complexity and how does it change with aging and disease? Neurobiology of Aging, 23(1), 23-26 (2002).

[11]   Zhang, X.S., Roy, R.J. EEG complexity as a measure of depth of anesthesia for patients. IEEE Transactions on Biomedical Engineering, 48(12), 1424-1433 (2001).

[12]   Zhang, X.S., Zhu, Y.S., Thakor, N.V., Wang, Z.Z. Detecting ventricular tachycardia and fibrillation by complexity measure. IEEE Transactions on Biomedical Engineering, 46(5), 548-555 (1999).

[13]   Liu, C., Liu, C., Shao, P., Li, L., Sun, X., Wang, X., Lie, F. 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, 32(2), 167-180 (2011).

[14]   Pincus, S.M. Approximating Markov chains. Proceedings of the National Academy of Sciences of the United States of America, 89, 4432-4436 (1992).

[15]   Pincus, S.M. Approximate entropy as a measure of system complexity. Proceedings of the National Academy of Sciences of the United States of America, 88(6), 2297-2301 (1991).

[16]   Chon, K.H., Scully, C.G, Lu, S. Approximate entropy for all signals. IEEE Engineering in Medicine and Biology Magazine, 28(6), 18-23 (2009).

[17]   Pincus, S.M., Singer, B.H. Randomness and degree of irregularity. Proceedings of the National Academy of Sciences of the United States of America, 93(5), 2083-2088 (1996).

[18]   Pincus, S.M. Approximate entropy (ApEn) as a complexity measure. Chaos, 5(1), 110-117 (1995).

[19]   Lepot, M., Aubin, J.B., Clemens, F.H.L.R. Interpolation in Time Series: An Introductive Overview of Existing Methods, Their Performance Criteria and Uncertainty Assessment. Water, 9(796), 1-20 (2017).

[20]   Gnauck, A. Interpolation and approximation of water quality time series and process identification. Analytical and Bioanalytical Chemistry, 380(3), 484-492 (2004).

[21]   Kaplan, D.T., Furman, M.I., Pincus, S.M. Techniques for analyzing complexity in heart rate and beat-to-beat blood pressure signals. Proceedings Computers in Cardiology, 1, 243-246 (1990).

[22]   Kaplan, D.T., Furman, M.I., Pincus, S.M., Ryan, S.M., Lipsitz, L.A., Goldberger, A.L. Aging and the complexity of cardiovascular dynamics. Biophysical Journal, 59(4), 945-949 (1991).

[23]   Kim, J.K., Ahn, J.M. Effects of a spectral window on frequency domain HRV parameters. Advances in Computer Communication and Computational Sciences, 924, 697-710 (2019).