35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

Compression based entropy estimation of heart rate variability on multiple time scales

Mathias Baumert, *Member, IEEE*, Andreas Voss, Michal Javorka

***Abstract*—Heart rate fluctuates beat by beat in a complex manner. The aim of this study was to develop a framework for entropy assessment of heart rate fluctuations on multiple time scales. We employed the Lempel-Ziv algorithm for lossless data compression to investigate the compressibility of RR interval time series on different time scales, using a coarse-graining procedure. We estimated the entropy of RR interval time series of 20 young and 20 old subjects and also investigated the compressibility of randomly shuffled surrogate RR time series. The original RR time series displayed significantly smaller compression entropy values than randomized RR interval data. The RR interval time series of older subjects showed significantly different entropy characteristics over multiple time scales than those of younger subjects. In conclusion, data compression may be useful approach for multiscale entropy assessment of heart rate variability.**

I. Introduction

Heart rate variability (HRV) refers to the beat-to-beat fluctuations in cardiac cycle and is modulated by the autonomic nervous systems [1]. Reduced HRV has been reported in clinical conditions such as diabetes mellitus and myocardial infarction and is clinically relevant for diagnosing diabetic autonomic neuropathy and predicting the risk for sudden cardiac death [1].

Assessment of HRV is based on short- or long-term ECG recordings in time and frequency domains [1]. Complexity assessment of heart rate fluctuations is thought to provide additional diagnostic information [2]. Among the various approaches that have been put forward for the quantification of HRV complexity, entropy measures have gained a significant interest. Shannon entropy, conditional entropy, approximate entropy and sample entropy, respectively are approaches that are frequently used for entropy estimation [3]. Entropy assessment over multiple time scales has been proposed to quantify complexity of HRV over different time regimes, thereby enabling complexity assessment of different sources of HRV acting at diverse time scales [4].

Previously, we proposed an algorithm to assess the entropy of HRV based on its compressibility [5, 6]. According to information theory, the simplest algorithm that is able to generate a time series is equal to the entropy of that time series. Based on the Lempel and Ziv algorithm for loss-less data compression (LZ77) we estimated redundancy in HRV and demonstrated the usefulness of this method in various clinical research studies [2].

The aim of this study was to develop the compression approach for multiple time scale assessment and investigate compressibility of HRV in young and old healthy subjects.

II. Methods

A. Data compression

From the point of information theory the smallest algorithm that produces a string is the entropy of that string (Chaitin-Kolmogorov entropy) [7]. Although it is theoretically impossible to develop such algorithm, data compression techniques may provide a reasonable approximation. Assuming that the source that generates the string is an ergodic process, the entropy per character x is the length of the compressed string divided by the length L of the original string, if $L \to \infty$.

In 1977, Lempel and Ziv developed a universal algorithm for lossless data compression (LZ77) [8], using string-matching on a sliding window that is nowadays implemented in many tools including 'zip' and 'Stacker'. The algorithm is briefly explained here:

A sequence of symbols $x = x_1, x_2, \ldots$ of length L from some given alphabet Θ of size $\Phi = |\Theta|$ is to be compressed. Subsequences $[x_m, x_{m+1},\ldots,x_n]$ of x will be denoted by x_m^n.

The algorithm keeps the w most recently encoded source symbols (sliding window of size w). The not-yet-encoded sequence of symbols is stored in the look-ahead buffer of size b. The encoder positioned at p looks for the longest match of length n between the not-yet-encoded n-string x_p^{p+n-1} in the look-ahead buffer and the already encoded string $x_{p-w+v}^{p-w+v+n-1}$ in the window beginning at position v. Thus, the matching string of n symbols is simply encoded by encoding the integer numbers n and v, i.e. a pointer to the previous occurrence of this string in the sliding window. In other words, the LZ77 algorithm operates in the following steps:

1) Encode the first w symbols without compression,
2) Set the pointer $p=w+1$,
3) Find for some v in the range of $1 \le v \le w$ the largest n in the range of $1 \le n \le b$ such that $x_p^{p+n-1} = x_{p-w+v}^{p-w+v+n-1}$,

Research supported by the Research Council Australia (DP 110102049).

M. Baumert is with the University of Adelaide, Adelaide, SA, 5005, Australia (corresponding author to provide phone: +61-88313-1616; fax: +61-8313-4360; e-mail: mathias.baumert@adelaide.edu.au).

A. Voss is with the Department of Medical Engineering and Biotechnology, University of Applied Sciences Jena, 07745 Jena, Germany (e-mail: voss@fh-jena.de).

M. Javorka is with the Department of Physiology, Comenius University, Martin, 03754, Slovakia (e-mail: Michal.javorka@jfmed.uniba.sk).

4) Encode the integers n and v into unary-binary code and the symbol $x_{p+n} \in \Theta$ without compression,

5) Set the pointer p to $p = n+1$ and go to step 3 (iterate).

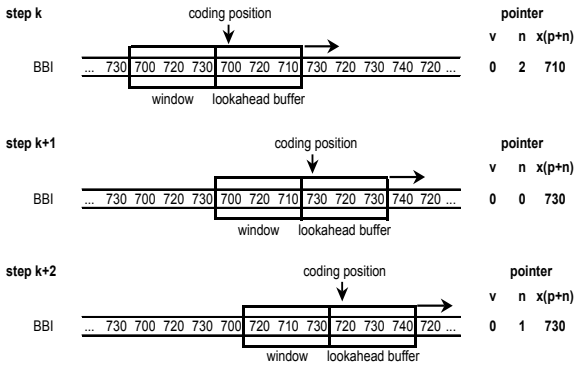

Figure 1 Illustration of the compression algorithm. Window (size three) and lookahead buffer (size three) are shifted through the stream of beat-to-beat intervals (BBI). The pointer stores position v, length of the matching string n, and the consecutive BBI $x(p+n)$.

When considering data compression for HRV analysis, the source of symbols is the sinus node of the heart that emits a sequence of RR intervals. Theoretically, RR intervals can take any value, albeit constrained by physiological limitations. To limit the alphabet Θ a coarse-graining procedure needs to be performed. Here, the sampling rate of ECG provides a natural limitation to the alphabet. In a further attempt to limit the alphabet and moreover, to remove the influence of RR interval variance from entropy estimation, we divided all RR intervals by the standard deviation of the RR interval time series, multiplied by a coefficient c, to weigh the influence of the standard deviation on the coarse graining. Subsequently, this normalized RR interval time series is rounded to integer values, which are subsequently assessed by the data compression algorithm.

In the implementation of the compression algorithm, a matrix $\boldsymbol{M}$ of size 3-by-K, which contains n_k, v_k, and $x_{k\,p+n}$ for each pointer k is obtained. This matrix fully describes the RR interval time series. The compression entropy of heart rate H_c is then defined as

$$H_c = \frac{K}{L} \tag{1}$$

The compression algorithm has two parameters, which influence the compressibility, i.e. the window size w and look-ahead buffer size b. Based on previous studies [5], we set $w = 7$ and $b = 3$.

B. Multiscale analysis

For investigating RR time series' compressibility on multiple time scales, we applied the compression algorithm outlined in the previous section ten times to calculate the compression entropy on ten time scales – scale one corresponding to original time series, higher scales (2-10) obtained by coarse-graining procedure, using the procedure proposed by Costa as part of multiscale entropy algorithm [4]:

Given a one-dimensional discrete time series, $\{x_1,...,x_i,...,x_N\}$, the coarse-grained time series $\{x(\tau)\}$ determined by the scale factor τ is constructed according to the equation:

$$x_j^{(\tau)} = 1/\tau \sum_{i=(j-1)\tau+1}^{j\tau} x_i, \tag{2}$$

where τ represents the scale factor and $1 \le j \le L/\tau$. In other words, coarse-grained time series for scale τ were obtained by taking arithmetic mean of τ neighboring original values without overlapping. For scale 1, the coarse grained time series is simply the original time series.

We restricted multiscale assessment to ten scales to take into account the finite length of RR interval time series, postulating that at least 200 samples are required for reliable entropy estimation.

C. Data

We analyzed ECG of 20 young (age: 21-34 years) and 20 old subjects (age: 68-85 years) from the Physionet database that were recorded during rest, while the subjects were watching the Disney movie "Fantasia" [9,10]. From these ECG recordings, RR intervals were extracted and visually checked for artifacts, using the LabChart® computer software (ADInstruments, NSW, Australia). Ectopic beats were filtered, using an algorithm that has been described in detail elsewhere [11]. For further processing we considered the last 2000 RR intervals from each recording.

As a first step towards the development of a multiscale compression analysis, we generated randomly shuffled surrogate data for all RR interval time series and compared the compressibility between original time series and surrogates on multiple time scales for different weighing coefficients c. After identifying a suitable choice for c, we investigated the effect of age on of heart rate complexity by performing multiscale compression analysis of RR interval time series of the 20 old subjects versus 20 young subjects.

D. Statistical analysis

Statistical analyses were carried out in GraphPad Prism® 6 (GraphPad Software, Inc., USA). Multiscale compression entropy functions of original versus surrogate data and of young versus old subjects, respectively, were carried out using the two-way ANOVA test for repeated measurements. Values of p less than 0.05 were considered to be statistically significant.

III. Results

A. Surrogate analysis

Surrogate analyses were performed for values of $c = 0.2$, 0.6, 1, 1.5, 2 and results are displayed in Figure 2. With increasing values of c, H_c values became larger for both, original and randomized RR time series. Two-way ANOVA

comparison of original and surrogate data, performed independently for different values of c, demonstrated a significant effect of scale on entropy estimation for all investigated choices of c (ANOVA $p < 0.0001$). Randomization had a significant effect on multiscale compression for values of c higher than 0.2 (ANOVA $p < 0.0001$). Interaction effects between scale and group were statistically significant for values of $c > 0.2$ and presumably caused by the relatively lower entropy values of original data for scales one and two compared to larger scales. Repeated measurement one-way ANOVA showed significant scale effects for both, original and surrogate data.

B. Age effect on multiscale compression

Comparison of multiscale compression entropy functions between young subjects and old subjects showed a significant scale effect as well as a significant scale x age group interaction effect (ANOVA $p < 0.0001$) for $c = 0.6, 1, 1.5, 2$, but no statistically significant group effect (ANOVA $p > 0.05$). Although the group average values of compression entropy over all ten scales were not significantly different between younger and older subjects, visual inspection of the compression functions, displayed in Figure 3, suggest group differences in the behavior of compressibility as a function of time scale, as demonstrated by the significant interaction effect. For further statistical analysis we considered compression with $c = 1.5$, since the age effect was most visible at this setting. At the same time, the compression functions were qualitatively comparable for all other values of c, except from $c = 0.2$. Scale-wise comparison of compression entropy, using the student's t-test, implied increased entropy in young subjects on scale one ($p < 0.05$) and decreased entropy on scales three, four and five ($p < 0.05$).

IV. Discussion

The main purpose of this study was to develop a framework for investigating the entropy of heart rate over multiple time scales based on a data compression algorithm. We were able to demonstrate significant differences in the complexity of RR intervals on different scales compared to that of randomized RR interval fluctuations.

Lempel-Ziv compression has been previously proposed as a tool for assessing the complexity of biomedical signals, but the assessment of dynamics using the original approach requires binary encoding of the process' dynamics [12]. Our approach followed a somewhat different strategy and is able to consider time series that consist of integer values. This comes at the expense of a notable deviation from the information theoretic foundation. Our measure of entropy does not describe the information contained in the time series in units of bit, but as a ratio of original time series length to compression matrix length. Despite this departure from this definition of entropy originating in information theory, our approach is in line with the underlying concept and provides a framework for practical implementation.

Multiscale assessment of compression entropy of RR interval time series demonstrated that the complexity of heart rate variability extends to multiple time scales, which is in line with previous studies that were based on sample entropy [13]. Surrogate analysis points to the presence of non-random dynamics over at least ten time scales, where the compressibility was highest on the original time scale. Surrogate data analysis further demonstrates that the compression approach gives relatively consistent entropy estimates for all time scales that were under investigation, although there was a small, but statistically significant time scale effect on the surrogate entropy estimate.

Age seemingly affects compression entropy of RR time series, but the effect size appears to be rather small. The RR time series of young subjects had higher entropy values on the original time scale compared to those of old subjects, but smaller entropy values on the intermediate time scales. Reduced heart rate variability in elderly subjects is well known and presumably caused be reduced vagal heart rate modulations [14]. Our data indicate that reduced heart rate variability is paralleled by reduced irregularity on a beat-to-beat basis, but increased irregularity on higher scales, possibly reflecting the lack of vagal heart rate control.

There are several limitations associated with this study. Window and buffer sizes were chosen based on previous investigations [5]. We cannot exclude that different settings might have affected our finding. Rather than studying a wide range of values for c we considered only five different values as a proof of concept, rather than optimizing significance levels for this particular data set.

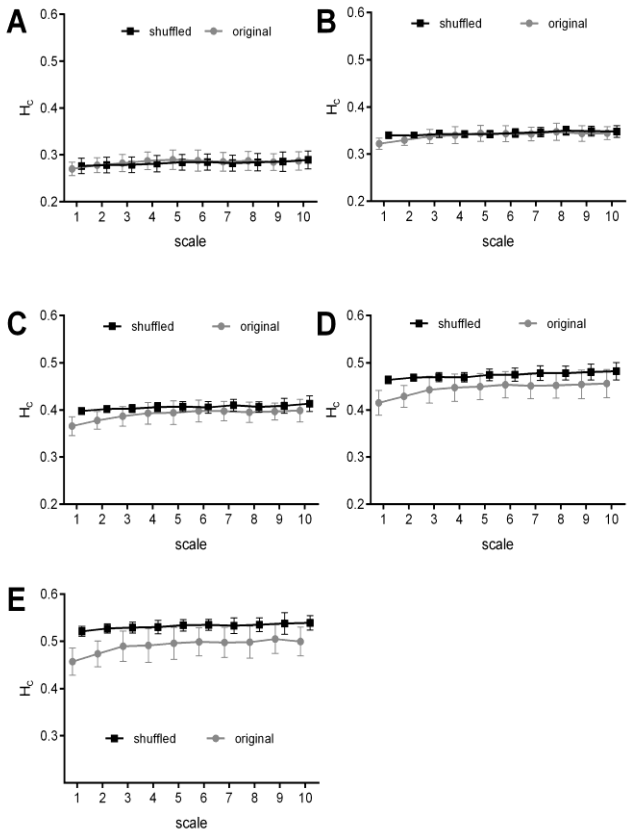

Figure 2 Compression entropy of randomized surrogate data in comparison to original RR interval time series of healthy volunteers as a function of time scale for different levels of coarse-graining (A – $c = 0.2$; B – $c = 0.6$; C – $c = 1$; D – $c = 1.5$; E – $c = 2$).

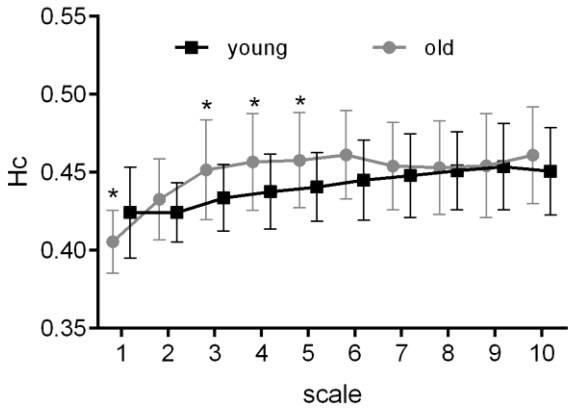

Figure 3 Compression entropy of RR time series of healthy young subjects in comparison to those of healthy old subjects as a function of time scale for c =1.5. Asterisks indicate significant differences between the two groups.

V. Conclusion

Multiscale investigation of compressibility may provide an alternative method for entropy assessment of biomedical signals over different time scales and be particularly useful for heart rate complexity analysis.

References

[1] A. J. Camm, M. Malik, J. Bigger, G. Breithardt, S. Cerutti, R. Cohen, P. Coumel, E. Fallen, H. Kennedy, and R. Kleiger, "Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology," *Circulation*, vol. 93, pp. 1043-1065, 1996.

[2] A. Voss, S. Schulz, R. Schroeder, M. Baumert, and P. Caminal, "Methods derived from nonlinear dynamics for analysing heart rate variability," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 367, pp. 277-296, 2009.

[3] A. Porta, S. Guzzetti, N. Montano, R. Furlan, M. Pagani, A. Malliani, and S. Cerutti, "Entropy, entropy rate, and pattern classification as tools to typify complexity in short heart period variability series," *IEEE Transactions on Biomedical Engineering*, vol. 48, pp. 1282-1291, 2001.

[4] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series," *Physical Review Letters*, vol. 89, p. 068102, 2002.

[5] M. Baumert, V. Baier, J. Haueisen, N. Wessel, U. Meyerfeldt, A. Schirdewan, and A. Voss, "Forecasting of life threatening arrhythmias using the compression entropy of heart rate," *Methods of Information in Medicine-Methodik der Information in der Medizin*, vol. 43, p. 202, 2004.

[6] M. Baumert, V. Baier, A. Voss, L. Brechtel, and J. Haueisen, "Estimating the complexity of heart rate fluctuations—an approach based on compression entropy," *Fluctuation and Noise Letters*, vol. 5, pp. L557-L563, 2005.

[7] G. J. Chaitin, Information, randomness & incompleteness: papers on algorithmic information theory vol. 8: World Scientific Publishing Company, 1990.

[8] J. Ziv and A. Lempel, "A universal algorithm for sequential data compression," *IEEE Transactions on Information Theory*, vol. 23, pp. 337-343, 1977

[9] N. Iyengar, C. Peng, R. Morin, A. Goldberger, and L. A. Lipsitz, "Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics," *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, vol. 271, pp. R1078-R1084, 1996.

[10] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, pp. e215-e220, 2000.

[11] N. Wessel, A. Voss, H. Malberg, C. Ziehmann, H. U. Voss, A. Schirdewan, U. Meyerfeldt, and J. Kurths, "Nonlinear analysis of complex phenomena in cardiological data," *Herzschrittmachertherapie und Elektrophysiologie*, vol. 11, pp. 159-173, 2000.

[12] R. Nagarajan, "Quantifying physiological data with Lempel-Ziv complexity-certain issues," *IEEE Transactions on Biomedical Engineering*, vol. 49, pp. 1371-1373, 2002.

[13] Z. Trunkvalterova, M. Javorka, I. Tonhajzerova, J. Javorkova, Z. Lazarova, K. Javorka, and M. Baumert, "Reduced short-term complexity of heart rate and blood pressure dynamics in patients with diabetes mellitus type 1: multiscale entropy analysis," *Physiological Measurement*, vol. 29, p. 817, 2008.

[14] A. Voss, A. Heitmann, R. Schroeder, A. Peters, and S. Perz, "Short-term heart rate variability—age dependence in healthy subjects," *Physiological Measurement*, vol. 33, p. 1289, 2012.