MEASURING THE COMPLEXITY OF NON-STATIONARY TIME SERIES — NONLINEAR INTERPRETATIONS OF SELECTED PHYSIOLOGICAL PROCESSES*

J.J. Żebrowski[a], W. Popławska[b], R. Baranowski[b] and T. Buchner[a]

[a]Institute of Physics, Warsaw University of Technology
Koszykowa 75, 00-662 Warszawa, Poland
e-mail: zebra@if.pw.edu.pl
[b]National Institute of Cardiology
Alpejska 42, 04-628 Warsaw, Poland
e-mail: baranows@ikard.waw.pl

(Received February 24, 1999)

A general method of analysis of non-stationary time series (time intervals of the human electrocardiogram) is presented: a short sliding window is used in conjunction with two different complexity measures. The first — a modified Shannon entropy called pattern entropy — quantifies the level of statistical order. The second is based on a symbolic dynamics in delay coordinate space and quantifies the level of sequential order by means of an estimator of algorithmic complexity. The sliding window procedure maps the original time series into a time series of the given complexity measure. The global state of the system is then characterized by the properties of the distribution of the resultant complexity measure. To characterize states on a local time scale the distribution of symbolic words is used. The method is applied to different data on heart rate variability, heart acceleration/deceleration and on the repolarization processes in the heart. We show that the nonlinear methods described may be applied to the analysis of the interaction of autonomic nervous system and the repolarization processes in the heart. This research was initiated to find new ways of prognosis the risk of sudden cardiac death. Below we show how the methods developed unveil new images of some physiological processes.

PACS numbers: 05.45.+b, 87.10+e, 87.80.+s

* Invited paper presented at the XI Marian Smoluchowski Symposium on Statistical Physics, Zakopane, Poland September 1–5, 1998

1. Introduction

This paper discusses two issues at the frontier between different disciplines of science: physics and medicine. On the one hand, we summarize and also present new results of our research on the techniques for measuring the complexity of nonlinear states of systems which have to be considered non-stationary. On the other hand, we demonstrate how these techniques may be used — in conjunction with medical knowledge — to assess the state of the autonomic nervous system. In the latter case, the ultimate goal is to find new diagnostic methods allowing to predict the risk of sudden cardiac death.

One of the topics discussed at this meeting was fuzzy logic and its use in bio-physical research. It is often acknowledged that medicine is not an 'exact' science. Fuzzy terms such as 'an ordered arrhythmia', 'a disordered arrhythmia', 'heart rate variability becomes inflexible in such and such circumstances' are used in day to day descriptions of particular cases and supplement the many precise bio-chemical and electrophysiological measurements helping to form a cardiologic diagnosis. One of the purposes of the research described below was to find non-linear dynamical methods to quantify these fuzzy terms.

In studying heart rate variability, specific patterns visible in the ECG trace are looked for. Linear spectral analysis and time domain analysis [1] have made a significant contribution to the understanding of the pathophysiology of heart rate variability. Linear analysis methods have severe limitations mostly due to the non-stationarity of the system studied. Recently, significant research [2–8] has been reported that non-linear methods derived from chaos theory perform better in assessing the risk of cardiac arrest. The issue whether heart rate variability is truly a deterministic chaotic state is hotly discussed, however [9, 10]. This is a strong indication that methods must be chosen which will be appropriate both for deterministic and for stochastic processes.

The time scale on which events which may be significant for sudden cardiac death occur is important. One of the reasons of the relative failure of such linear methods as the power spectrum in correctly predicting the risk of sudden cardiac death may be due to the number of data the method requires thus limiting the time scale from below. Woo *et al.* [11] and our own research [12] indicate that 24-hour measurements of heart rate variability may be mapped into 2 or 3 dimensional phase portraits. Certainly the 'torpedo shape' found by Woo *et al.* [11] may be associated with 24 hours of the natural heart rhythm (sinus rhythm). It is true also that various types of pathology distort the phase space trajectory producing a wide variety of shapes [12, 11]. It is however difficult to associate the level of risk with

these shapes alone. On the other hand, we have shown that on a short time scale of a few hundred heart beats (1.5-4.5 minutes) specific patterns in 3-dimensional phase space may be found [7, 12]. The dominant are the spiral and the radial patterns as discussed in [12–14]. In some cases a limit cycle may be obtained [15]. An elaborate embedding allowing to extract such patterns from the 24-hour time series was devised by Babloyantz *et al.* [16]. In our research, we have focused on finding the proper complexity measures to characterize the patterns obtained in phase space.

The data analyzed in this paper is discussed in Section 2. The methods used to analyze regularity in the statistical and in the sequential sense in phenomena occurring on a long time scale are presented in Section 3 while Section 4 discusses analysis of the short time scale effects both in heart rate variability and in the interaction of the heart rate with the repolarization processes the heart tissue. A summary of our results is given in Section 5.

2. The data

Most of this paper discusses heart rate variability which is measured as a series of time intervals between specific points of the ECG (the RR intervals as marked in Fig. 1). The R peak represents the moment in time when the ventricles contract. The contraction of heart tissue is associated with a depolarization phase superceded by a repolarization phase. The repolarization time (the RT interval[1] in Fig. 1) is an important diagnostic parameter.

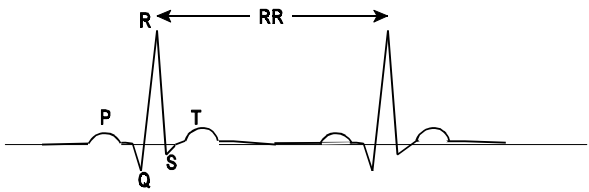

Fig. 1. Schematic of two cycles of an ECG trace with the characteristic points and time intervals marked.

Within the last 5 years our group has analyzed over 300 24-hour ECG recordings in the context of using nonlinear dynamics methods for medical prognosis of the risk of cardiac arrest. The bulk of the heart rate variability analysis was performed using the Del Mar Strata Scan 563 software at a 128 Hz sampling frequency. The repolarization processes were analyzed

[1] Although it is more usual to consider the QT interval as the repolarization time, because the point Q of the ECG trace is difficult to extract automatically and reliably from a Holter recording, we prefer to use the RT interval. It is well known that, for a given individual, the QR interval has constant length of approximately 10–20 ms.

by extracting at 256 Hz both the RR and RT intervals from the 24-hour recordings using a custom software designed at the Institute of Cybernetics of the Politechnica de Catalunya, Barcelona. [17]. All our recordings were carefully checked for artifacts and for arrhythmia by a qualified cardiologist. No kind of arrhythmia filtering was applied to the data.

Below we present results for 60 cases of apparently healthy individuals 16-64 years of age (only 7 of them were women), and 86 patients with hypertrophic cardiomyopathy. This disease caused by genetic mutations results in an abnormal structure of the heart muscle with a wide spectrum of changes ranging from small abnormalities in the ECG trace through different types of cardiomyopathy including even a dilation of the heart (dilated cardiomyopathy). At any stage of the disease the main risk is sudden cardiac death. So far there is no method to predict which patient would be prone to this risk *i.e.* which patient should be equipped with an automatic defibrillator. The group of patients with hypertrophic cardiomyopathy chosen for analysis were chosen precisely because of the failure of the standard time domain and spectral methods of the analysis of heart rate variability in defining the patients with the highest risk of sudden cardiac death.

In 7 older men, we performed additional testing of the autonomic nervous system functions including tilting of the whole body to the vertical position which stimulates the sympathetic part of this system. Before and after this maneuver, in addition to the ECG, the plasma levels of the sympathetic neurotransmitters norepinephrine and dopamine were also measured.

3. Long time scale

3.1. Statistical complexity

To analyze the heart rate variability, we have used pattern entropy [12, 14, 15, 18–21]. This is a complexity measure derived from the Shannon entropy:

$$S = -\sum_{i=1}^{N} P_i(k) \log P_i(k)\,,$$

where N is the number of bins in the histogram of RR intervals, is the probability distribution (a normalized histogram) of RR intervals within the time window, i is the bin index and k is the index of the RR interval at the end of the time window. Pattern entropy is obtained by substituting for the usual one dimensional probability density $P_i(k)$ the incomplete joint probability density:

$$P_i = p_i(k)p_i(k+\tau)p_i(k+2\tau)\,,$$

where is the delay of the time delay reconstruction. In all our computations we use $\tau = 2$ beats [12, 14]. The use of the incomplete joint probability (in place of the full joint probability in 3-dimensions) with the Shannon entropy (1) implies that *pattern entropy will be large for highly ordered time series* — contrary to the properties of Shannon entropy itself. Note, that pattern entropy does not have all the properties of conventional entropy (*e.g.* it is not additive). For convenience, throughout our work we use arbitrary units which are obtained by multiplying pattern entropy given by the above described equations by 10^4.

If the length of the time window is small compared to the length of the time series analyzed, window pattern entropy WPE is obtained. The length of the time window may be varied between 50 and 400 beats (for the effect of window size see [18]). Note that a window length measured in integer time (beats) rather than a fixed window size in real time (*e.g.* seconds) is preferred as for a real time length window — due to heart rate variability — the number of beats would change drastically from window to window. Typically a window length of 50 beats is equivalent to 20–35 s of real time while 400 beats is equivalent to approximately 3–4 Min.

Window pattern entropy fluctuates as the window sweeps through time series. Initially we calculated the minimum, maximum and average of WPE and compared these values with the risk of cardiac arrest for individual cases [12]. We now see that the full distribution of WPE of heart rate yields a better image of the given case studied [18–21]. We find, for example, that the most probable pattern entropy value measured in a 5 Min. epoch seems to correlate strongly with the plasma level of norepinephrine when that is measured simultaneously [15]. Although the shape of the distribution of WPE in a 24-hour heart rate time series is Poisson-like (see Figs 8–10 below) in a predominant number of cases, the most probable WPE value emerges as an indicator of the state of the system. Similar results were obtained for the dynamic states of the logistic map [19].

All definitions of the entropies used in chaos theory such as the Kolmogorov–Sinai entropy and generalized Renyi entropies [22] avoid the problem of fluctuations by assuming ergodicity and taking the limit of infinite time. For this, an infinitely long time series is required so that in practical applications one only calculates an estimator of the entropy. In our case, the problem is made more complex by the non-stationarity of the system: no restrictions are made on the behavior of the human subject during the 24-hour measurement of heart rate. In stationary state, window pattern entropy is a decreasing function of the time *i.e.* of the length of time window (*cf.* similar discussion in [3]). In a non-stationary state, we assumed that the best approximation of pattern entropy with the limit of $t \to \infty$ may be obtained by the following procedure: we calculate cumula-

tive pattern entropy CPE as pattern entropy with the length of the window gradually expanding to span the whole 24-hour RR interval time series. The minimum of CPE we assume to be the best approximation of the pattern entropy at $t \to \infty$. Although this is a purely empirical approach, we found that, often, CPE is a monotonically decreasing function of the time and only rarely does it increase during the given 24-hour recording of heart rate. In other words, observation of a large number of cases shows that instances of true non-stationarity due to the activity of the subject which cause a change of state (thereby causing an increase of the cumulative pattern entropy) are relatively rare. Note that we have often found that the 24-hour minimum of CPE remains the best indicator of the risk of cardiac arrest [12, 14].

It has been shown before [23] that the differences of RR intervals and the respective pattern entropies (denoted WPD and CPD) [18] are a measure of the activity of the parasympathetic nervous system. On the other hand, heart rate variability itself together with the resultant pattern entropies WPE and CPE reflect the activity of both the parasympathetic and the sympathetic nervous systems [18].

When studying simultaneously pairs of the time series — the RR intervals and their differences — we found that even for healthy persons the relation between the pattern entropies calculated from these time series is not always the same. This effect is seen both in the dependence on the time of WPE and WPD as well as in the shape of the distributions of these measures of statistical complexity.

Fig. 2 depicts an example of the dependence of RR intervals (part a) and their successive differences (part b) as functions of the time during a exercise stress test. In Fig. 2 the stress test lasted from about $t = 1000$ s to about $t = 2700$ s. This test manifests itself by a period of linear decrease of the RR interval length (during which the load is increased) followed by a recovery period. The variance of the RR interval decreases during the test and does not regain the value from before the test for a long time. It can be seen that window pattern entropy WPE (thick curve in part a) of Fig. 2) rises sharply at the beginning of the test and decays slowly afterwards. Pattern entropy of the RR interval differences WPD (thick curve in part b) of Fig. 2) follows the changes of WPE during the initial part of stress test but its value remains constant on the average at a moderate level for a long time after the test. In other examples, of the same type of behavior, WPD decreases sharply as soon as the load is reduced. This type of heart rate variability may be dubbed 'flexible' .

An example of the 'inflexible' heart rate variability is seen in Fig. 3 during a stress test conducted on a different (and much older) healthy subject. In this case, the stress test lasted from 600 s to 1900 s. It can be seen that, for this subject, reaction to the stress test occurs only briefly during the highest

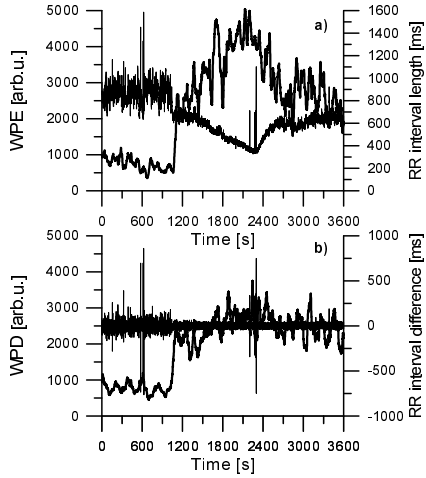

Fig. 2. An example of the 'flexible' heart rate variability: the dependence of RR intervals and the corresponding WPE as functions of the time (part a) compared with the RR interval differences and the corresponding WPD (part b) measured in a healthy subject during a stress test performed between about $t = 1000$ s to about $t = 2700$ s.

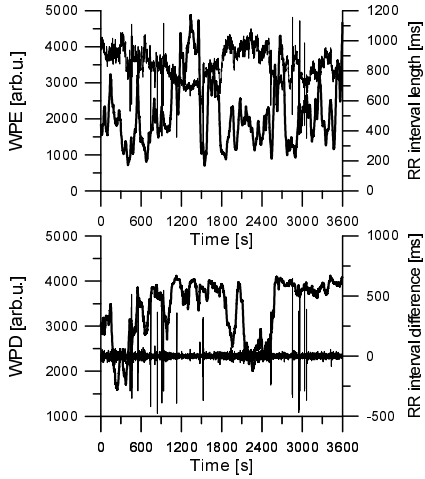

Fig. 3. An example of the 'inflexible' heart rate variability: the dependence of RR intervals and the corresponding WPE as functions of the time (part a) compared with the RR interval differences and the corresponding WPD (part b) measured during a stress test performed between $t = 600$ s and $t = 1900$ s on a healthy subject.

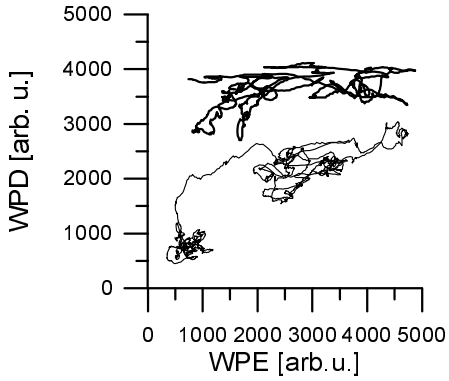

Fig. 4. The dependence of pattern entropy of RR interval differences WPD on the pattern entropy of RR intervals during the stress test itself. Thick curve — 'inflexible' heart rate variability; thin curve — 'flexible' heart rate variability. The raw data for the stress test is the same as in Fig. 2 and Fig. 3, respectively.

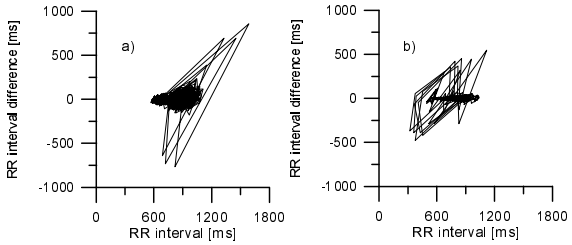

Fig. 5. The dependence of the RR interval differences on the RR interval lengths during period of the stress tests for the 'flexible' (part a) and the 'inflexible' heart rate variability (part b). This figure depicts the raw data used to calculate Fig. 4.

load phase and that the decrease in the level of statistical complexity of the RR intervals during recovery after the stress test occurs much faster. At the same time, the behavior of the complexity of the RR interval differences as a function of the time seems not to reflect the stress test at all.

Fig. 4 depicts WPD as a function of WPE during the stress test for the two examples just described. It can be seen that in the 'flexible' heart rate variability case (thin curve in Fig. 4) the stress test increases the level of the statistical order of both RR intervals and their differences. On the other hand, for the example of the 'inflexible' kind (thick curve in Fig. 4) the statistical order of the heart rate fluctuates in a wide range throughout the test while the complexity of heart acceleration/deceleration stays constant on a high level. Fig. 5 demonstrates that the differences between the examples are not visible in the raw data itself.

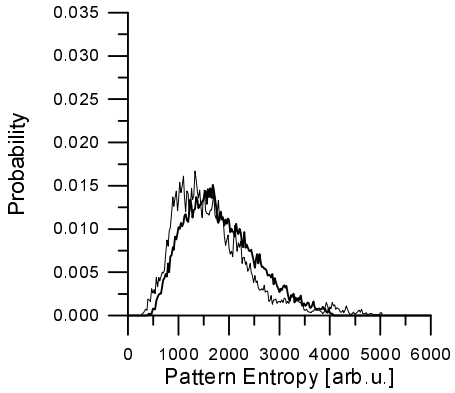

Fig. 6. The distributions of pattern entropy for the RR intervals (WPE — thin curve) and for the RR interval differences (WPD — thick curve) calculated for the full 24-hour time series for the 'flexible' heart rate variability. The raw data in Fig. 2 is a fragment of the recording used to calculate these distributions.

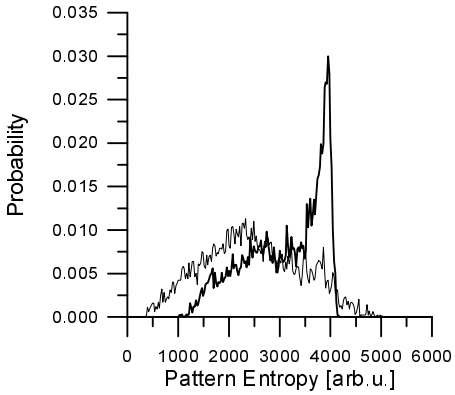

Fig. 7. The distributions of pattern entropy for the RR intervals (WPE — thin curve) and for the RR interval differences (WPD — thick curve) calculated for the full 24-hour time series for the 'inflexible' heart rate variability. The raw data in Fig. 3 is a fragment of the recording used to calculate these distributions.

Figs 2–5 depict effects seen only during the short period of the stress test. However, the property of 'flexibility' of the heart rate variability may be seen also in the shape of the 24-hour distributions of the respective pattern entropies. For the flexible heart rate variability, it can be seen in Fig. 6 that the distributions of both WPE and WPD coincide. In the case of the 'inflexible' heart rate variability (Fig. 7) an extremely large peak in the distribution of WPD (thick curve) appears at a high value of the complexity

measure indicating that heart rate variability with a highly ordered mode of rate change is dominant during the whole 24-hours of the recording.

4. Sequential complexity

Symbolic dynamics [24] is obtained by introducing a coarse grained partitioning and analyzing the way in which the given system visits each partition. Although it may be argued that the partitioning may be arbitrary, it is accepted that the most clear results are obtained when physical properties of the system are taken into account. Thus, in the symbolic dynamics of one dimensional maps [24] the partition border is given by the critical points of the map and the symbols are uniquely associated with its branches. In the case of heart rate variability, the dynamics is too complicated to model by a simple map for which such branches may defined. On the other hand, the spiral shape of three dimensional trajectories of RR intervals [12] — a dominant feature of heart rate variability in healthy individuals — indicates that an unstable fixed point may play an important role in the dynamics. For this reason, we introduced [15] [25] the following symbolic coding which takes advantage of the Takens delay coordinate reconstruction.

As a surrogate of the critical point of 1-dimensional maps, we used the average interval for each time window (*i.e.* the focal point of the spiral trajectory). All intervals were compared with this the reference level and a symbol was assigned: if the RR(i) value was less than the average the symbol was "L" and "R" if the opposite was true. Similar comparisons were carried out for RR($i+\tau$) (symbols "D" or "U") and RR($i+2\tau$) (symbols "T" or "B"). If the value of the RR intervals was closer to the average than the sampling error (7.5 ms) the symbol "C" was written. Thus, a given pattern in 3-dimensional space was mapped to a sequence of 3 letter words composed of 7 different symbols. To quantify the complexity of the sequence of symbolic words, the Lempel-Ziv algorithmic complexity [26] was then used by means of an algorithm implemented after Kaspar and Schuster [27]. This algorithm counts the number of unique k-symbol strings (k =1,2,...,K, with K the length of symbol sequence) into which the given sequence of symbols may be decomposed. By definition, algorithmic complexity is obtained in the limit of an infinite length of the symbol sequence analyzed. Since, here, we calculated algorithmic complexity of a finite sequence of symbols (three times the number of intervals per time window), the value calculated is only an estimator. A similar estimator was used by Witt *et al.* [28] except that there the normalized estimator of [27] reflecting algorithmic complexity per symbol was used. Here, with the length of the time window held constant at 100 beats, the estimator that we used is not normalized. The numbers obtained directly from the Lempel–Ziv algorithm were easier to interpret being related to the number of unique sequences within a window.

Similarly as in [7], we also found that better results are obtained if the reference level for symbolic coding is slightly shifted with respect to the average of the RR intervals in a time window. In our case, we multiplied the average by a constant $a = 1.01$. The reason for the use of such a constant with heart rate variability is the natural asymmetry of the distribution of the RR intervals themselves.

Since the 100 interval time window used here throughout the symbolic dynamics analysis is relatively short, sweeping the 24-hour time series with the time window results in a distribution of the algorithmic complexity values. The distributions discussed below were constructed only from local extrema of algorithmic complexity or from constant values of the complexity at which the system stayed for at least 3 window positions.

Fig. 8 depicts 24-hour distributions of the extrema of the algorithmic complexity for 3 healthy persons 25 years of age. Below them the corresponding distributions of the local extrema of window pattern entropy are plotted. The central example kndt in Fig. 8 depicts a case which deviates from the general character of other persons belonging to this age group. Patients chm and pzr are both very typical: the distribution of algorithmic complexity is Gaussian-like in appearance with a small negative skewness and a moderate positive kurtosis. The corresponding distributions of window pattern entropy are much more Poisson-like and relatively narrow with a long tail extending into high entropy values indicating that episodes of high statistical order are relatively rare. By contrast, for kndt both distributions are broader and a distinct peak of probability close to 4000 window pattern entropy is found.

When such distributions were constructed for persons just above 40 years of age (Fig. 9) the effect of age could be seen. The two typical examples for this age group (stra and kczk) again have a Poisson-like distribution of algorithmic complexity: skewed to the left and with a positive kurtosis. Now, however, the maximum of the distribution has shifted towards higher complexity values. For ttk — a case which just barely meets the classical medical criteria for cardiologic norm — the changes are even more apparent. The distributions of both measures are very broad and the distribution of window pattern entropy is skewed to the left with a large peak at the value of 4000. This is one example of only three such cases we found among the 60 healthy subjects reported here — for all others the most probable window pattern entropy was between 800 and 1600. Note that when studying the risk of cardiac arrest we found such characteristic peak at 4000 entropy value predominantly in the high risk group of patients after a myocardial infarction or with valvular heart disease [18, 19, 21].

When normals older than 50 were studied, it was found that the changes with age were not so large between this group and the 40 year old group.

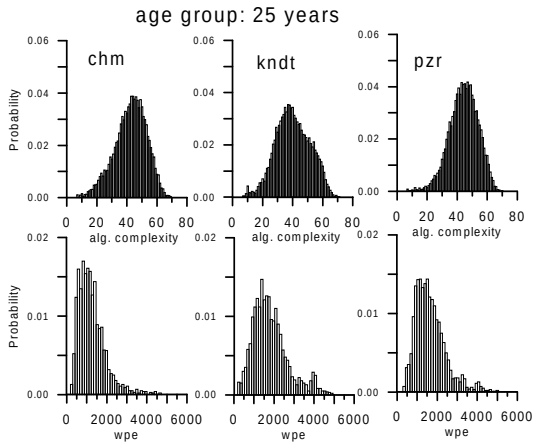

Fig. 8. Comparison of the distributions of algorithmic complexity (top row) and window pattern entropy wpe (bottom row) for three examples of 24-hour recordings of sinus rhythm measured in persons about 25 years of age.

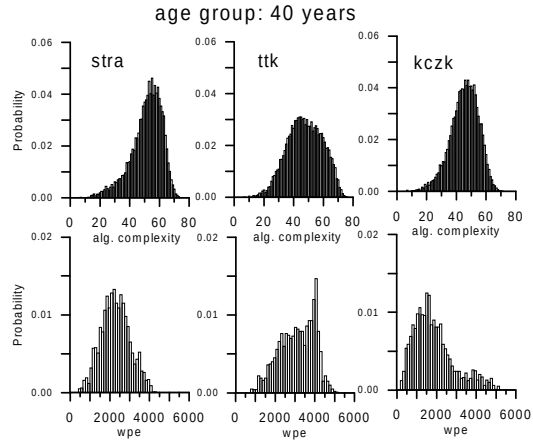

Fig. 9. Comparison of the distributions of algorithmic complexity (top row) and window pattern entropy wpe (bottom row) for three examples of 24-hour recordings of sinus rhythm measured in persons 1 or 2 years above 40 years of age.

Basically the distributions of both complexity measures were broader and a shift towards higher values was evident especially for window pattern entropy. Three examples for the oldest age group are shown in Fig. 10 where again the central case is borderline normal while the other two are typical. The shapes of the 24-hour probability distributions of complexity measures do not reflect the instantaneous relation between the entropy and algorithmic complexity measures as functions of the time. We showed elsewhere

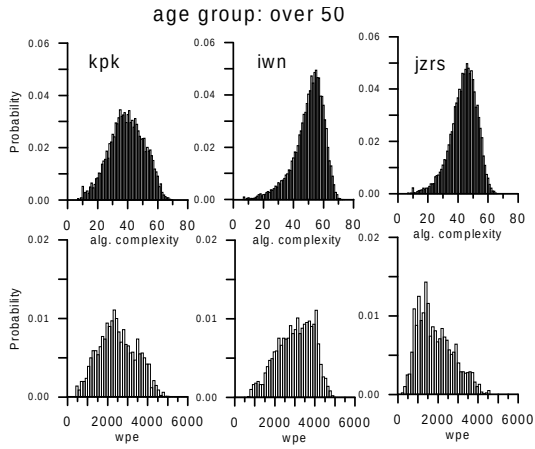

Fig. 10. Comparison of the distributions of algorithmic complexity (top row) and window pattern entropy wpe (bottom row) for three examples of 24-hour recordings of sinus rhythm measured in persons above 50 years of age.

for characteristic, specially chosen medical examples [15] that, over short periods of the time, sequential complexity of the heart rate variability for a healthy individual may repeatedly be larger than for the medical condition called atrial fibrillation. The latter is considered to be the most random type of heart rate variability. It seems that the notion of the randomness of heart rate in atrial fibrillation is due mostly to the widely used spectral analysis [1]. We have shown before [15, 25] that the level of order of heart rate variability for atrial fibrillation — as seen by both pattern entropy and by algorithmic complexity — was low as expected. In other cases, however, the dependence on the time of the simultaneously calculated window pattern entropy and algorithmic complexity is not trivial. Most of the time, a change of the value of pattern entropy is accompanied by a like change of the corresponding algorithmic complexity. Since pattern entropy is low for higher disorder the fact that the two complexity measures follow each other as functions of the time would indicate a compensation of an increase of statistical order in the series (less frequencies used) by a more complex sequence of RR intervals. There are both short periods of the time (up to 2 minutes) as well as long periods of the time (about 20 minutes) when the behavior of the two complexity measures is not correlated. An exceptionally clear case of such independent behavior was found [15] for the case of a person who suffered a cardiac arrest at the end of the 24-hour recording. Within that recording long periods of decorrelation between pattern entropy and algorithmic complexity occurred often while in a recording made for the same subject 9 months later and under medication no such periods could be found.

4.1. Short time scale — histograms of words

4.1.1. Heart rate variability

In this scale the sequence of events plays a dominant role. To characterize the short term effects within the 24-hour time series we used histograms of symbolic words [24] [7]. A typical example of such a histogram for a time series of RR intervals representing a sinus rhythm (*i.e.* the heart rhythm of a healthy person) is shown in Fig. 11. The reference level given by the horizontal dotted line at 1/27 represents the probability density of the words in the case of a complete lack of pattern. It can be seen that in spite of the coarse graining introduced by the symbolic coding, certain words are much more probable than others while some words have a probability much lower than the reference level. Note the large peaks of the distribution in Fig. 11 at the words '000' and '222'. These peaks represent the probability of a sustained (spanning 5 beats) higher than the window average heart beat rate and a sustained lower than the average heart rate, respectively. Typical for a healthy person is also the low level of probability density at the word '111' which represents a sustained very stable heart rate equal to the average within the sampling error.

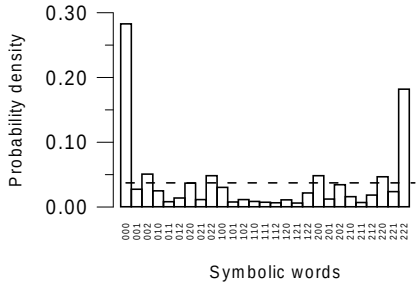

Fig. 11. Distribution of symbolic words calculated for a 24-hour time series of RR intervals measured in a healthy person (sinus rhythm). The dashed line at probability density 1/27 marks the level of a random distribution of words.

The distribution of symbolic words changes with the risk of cardiac arrest. These changes may be only quantitative as in the case depicted in Fig. 12 of the young patient without any symptoms except for a twice sustained cardiac arrest. Comparing this distribution with that of the control pair in Fig. 11, it can be seen that the most visible change is seen in the outlying peaks '000' and '222'. There is, however, a subgroup of patients at high risk of cardiac arrest, for which the words containing the symbol '1' signifying rate stability play a significant role. An example of such a case is seen in Fig. 13 in which the distribution of symbolic words, for a 2 hour recording which ended in cardiac arrest, is depicted.

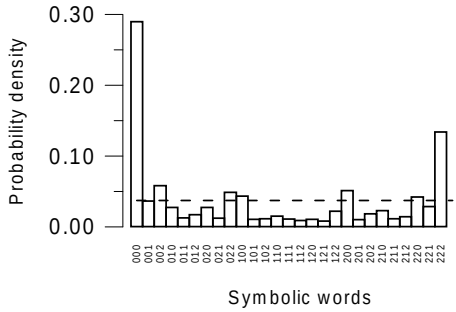

Fig. 12. Distribution of symbolic words calculated for a 24-hour time series of RR intervals measured in a person at risk of cardiac arrest. The dashed line at probability density 1/27 marks the level of a random distribution of words.

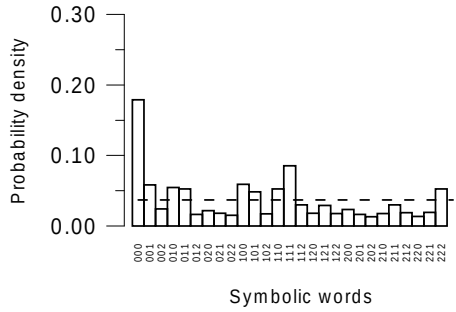

Fig. 13. Distribution of symbolic words calculated for a 2-hour time series of RR intervals which ended in a fatal cardiac arrest. The dashed line at probability density 1/27 marks the level of a random distribution of words.

We examined the probability densities of various words and their ratios and we found that the index of maximum variability ImaxVar *i.e.* the ratio of the probability densities $p(\text{'000'})/p(\text{'222'})$ correlates best with the risk of sudden cardiac death in the otherwise difficult to analyze group of patients with hypertrophic cardiomyopathy. The index ImaxVar was strongly related to certain echocardiographic parameters of hypertrophy (Pearson correlation coefficient $r = 0.75$ with $p < 0.001$). The word '111' expressing a lack of variability of the heart rhythm correlated strongly with such medical parameters that indicate a kind of ventricular overload. The index of maximal variability was found also to identify the high risk patients with a sensitivity, specificity and predictive accuracy of 81 %. This result is significantly better than the parameters of any other noninvasive technique so far.

4.1.2. Relation between the heart rate variability and the repolarization processes in the heart

The length of time needed for the repolarization of the heart tissue (the RT interval in Fig. 1) depends to a certain extent on the heart rate. Because of this, simple normalization formulas which allow to recalculate the repolarization interval taking into account the length of the RR interval are often used [29]. These formulas were derived for the sinus rhythm and usually with the assumption that a heart rate is limited to some range. Plotting the RT interval data as a function of the RR interval length shows that even if for a healthy person (Fig. 14) assuming that a simple relation may be justified, it is certainly not so for patients with hypertrophic cardiomyopathy (Fig. 15 and Fig. 16). We have previously reported that 24h repolarization variability expressed as standard deviation is lower in normal subjects in contrast to hypertrophic cardiomyopathy patients [29]. In contrast to heart rate variability, higher signal variance is a feature of normality. In fact, although from the statistical point of view the average linear correlation of the RT intervals with respect to the RR intervals is strong (Pearson correlation coefficient $r = 0.67$ in Fig. 14), note that during short duration changes of the heart rate this dependence is weak (Fig. 17). Only during a prolonged period of the increase in the heart rate (between 225 s and 325 s in the top part of Fig. 17) a complex dependence between the two processes may occur (bottom part of Fig. 17). Such behavior is most often seen in persons with a high risk of sudden cardiac death.

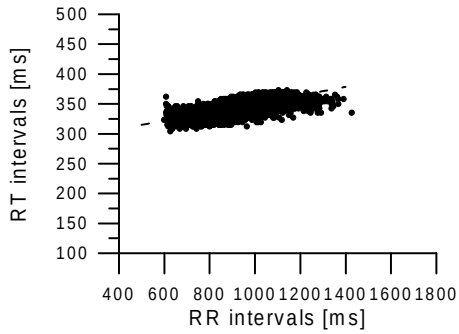

Fig. 14. The dependence of RT interval length on the heart rate for a healthy person. The dashed line depicts a linear regression fit to the data. The time series was 4 hours long and measured during the night.

To examine this dependence, we formed a 4-dimensional phase space $\{\mathrm{RR}(t), \mathrm{RR}(t+\tau\ ), \mathrm{RT}(t), \mathrm{RT}(t+2\tau)\}$. We next applied the coding algorithm described above to the RR intervals and to the RT intervals calculating their proper window averages as reference levels. Using data on the risk of

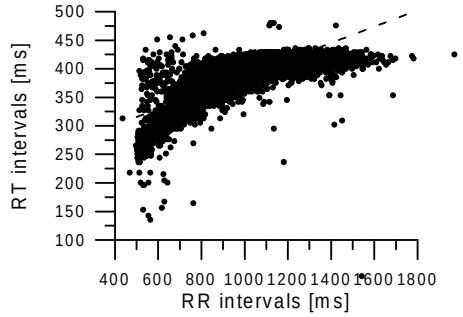

Fig. 15. The dependence of RT interval length on the heart rate for a person with hypertrophic cardiomyopathy who has since died. The dashed line depicts a linear regression fit to the data. The time series was 4 hours long and measured during the night.

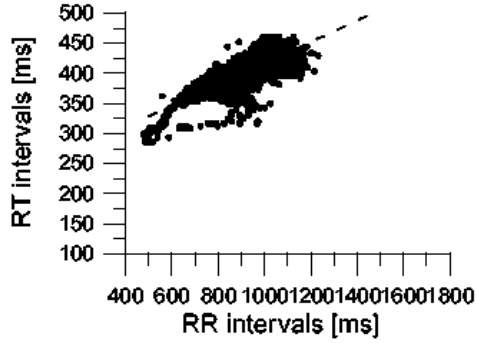

Fig. 16. The dependence of RT interval length on the heart rate for a person with hypertrophic cardiomyopathy who has since died — case different from that in Fig. 15. The dashed line depicts a linear regression fit to the data. The time series was 4 hours long and measured during the night.

cardiac arrest in hypertrophic cardiomyopathy, we studied the effect of different values of the tolerance parameters for RR intervals and RT intervals on the histograms of 4-dimensional symbolic words. We found that the best results were obtained when the tolerance parameter values were 7.5 ms for the RR intervals and 4 ms for the RT intervals. Note that here the optimum tolerance parameter for the former is about twice the sampling error and that, usually, the variance of the heart rate is also is much larger than the variance of the RT interval.

We stress that we do not know of any formal theory for such a hybrid phase space composed of two delay coordinate embeddings of different variables related to different processes occurring in a single system. The under-

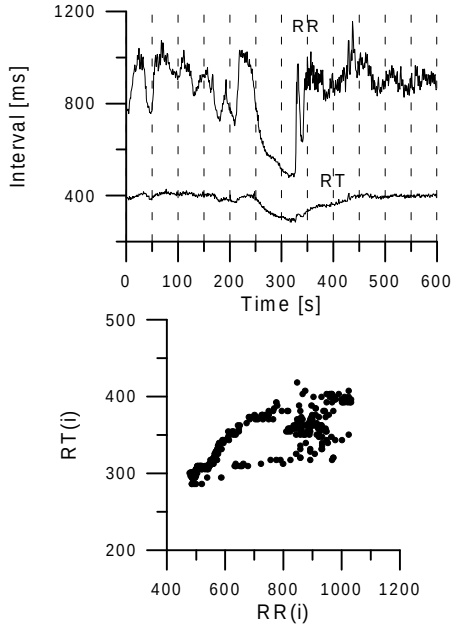

Fig. 17. The dependence of the RR and RT intervals on the time (top). The bottom figure depicts the dependence of the RT interval on the heart rate during the stress test visible in the top part approximately between 225 s and 325 s.

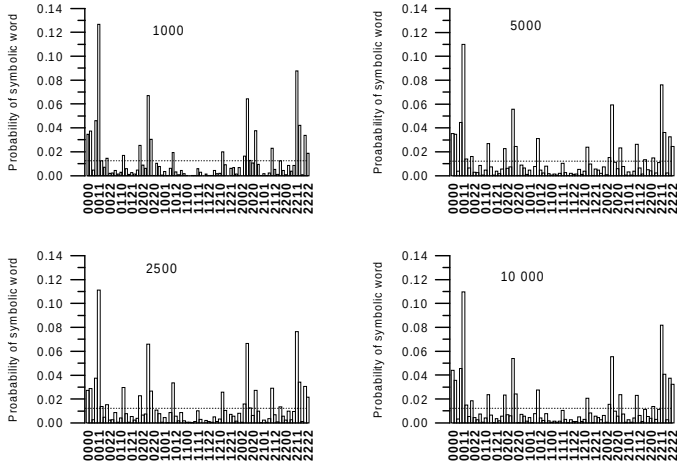

Fig. 18. The dependence of the distribution of the 4-dimensional symbolic words in {RR,RT} space on the length of the series as marked in each graph. The 10000 interval long data series lasted 4 hours.

lying heuristics is the following. By the Takens theorem, diffeomorphisms exist between the true trajectories for the heart rate and for the repolarization processes and their delay coordinate counterparts. We thus expect that the hybrid phase space composed of delay coordinate embeddings may be formed just as the proper phase space to examine the relation between heart rate and repolarization processes would be one formed by all the variables describing them. The approach described here is completely empirical of necessity as neither all such variables are known nor has the dimensionality of each of the processes been ascertained. We demonstrate below, however, that the empirical approach yields useful results.

Fig. 18 demonstrates the effect of the length of the data series on the shape of the distribution of 4D symbolic words. Fig. 18 is also an example of such a distribution for a healthy person obtained by sweeping through the time series of RR and RT intervals with a 100 data point window. It can be seen that there is some structure which is invariant with respect to the length of the time series and that even for a very short time series of 1000 data points this structure is recognizable.

To find out which of the two data series is responsible for the structure in Fig. 18, we performed a series of surrogate data tests. We compared the distribution of the symbolic words obtained with the original data with such a distribution obtained by selectively shuffling the data the RR interval time series only, by shuffling the RT data only and by shuffling both (Fig. 19).

It can be seen that selective shuffling only of the RR intervals retains the major structure of the original distribution while selective shuffling only of the RT intervals creates a distribution similar to that obtained when both types of data are shuffled.

An opposite effect was observed when data from some cases of high risk patients with hypertrophic cardiomyopathy were examined. An example of this is given in Fig. 20. It can be seen that the distribution obtained for the selectively shuffled RR intervals resembles that found when both RR and RT were shuffled. On the other hand, selective shuffling of RT only produces a symbolic word distribution which resembles the original one the main effect being the destruction of the large peaks at word '0000' and at word ' 2222', as expected. The two cases shown in Fig. 19 and Fig. 20 demonstrate again how complex and variable may be the relation between the heart rate and the repolarization processes. Global heart rate variability (expressed as standard deviation of RR intervals) in both cases is very similar — 112ms for the data of Fig. 19 versus 92 ms for that of Fig. 20 (*i.e.* a difference of about 10%). However, the variability of the RT intervals is larger by 30% in the latter case (18 ms) than in the former (12 ms). This explains why shuffling the RT series in Fig. 20 changed the basic features of the histogram more than shuffling the same type of data in case of the example of the sinus

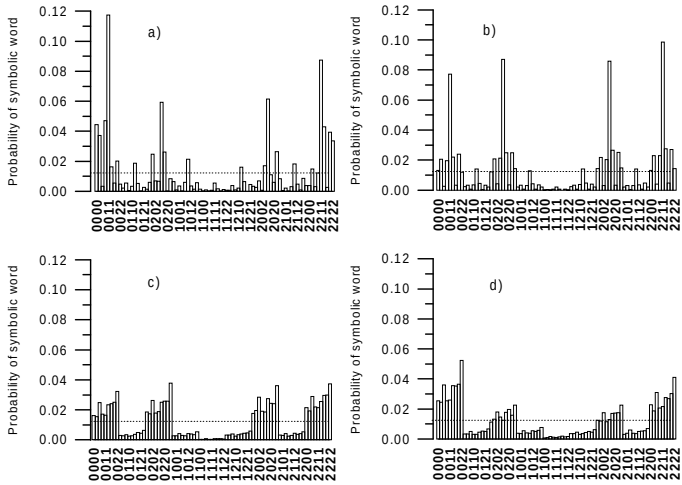

Fig. 19. The effect of data shuffling on the distribution of 4-dimensional symbolic words measured during sinus rhythm (*i.e.* in a healthy person): a) original data, b) only RR shuffled and RT as measured c) both time series surrogate d) only RT shuffled and RR as measured. The time series was 4-hours long and was measured during the night.

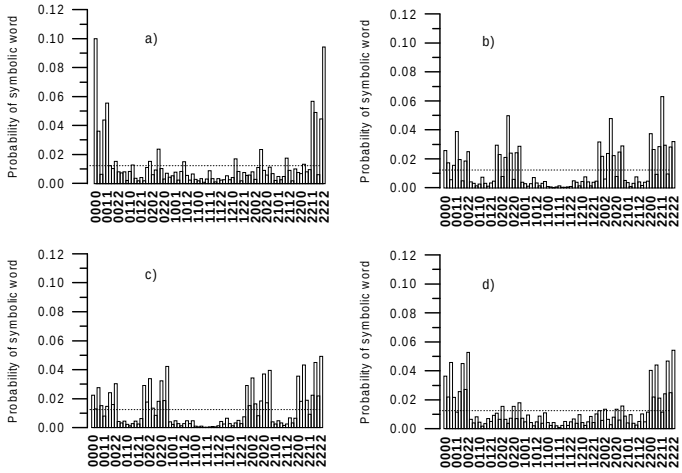

Fig. 20. The effect of data shuffling on the distribution of 4-dimensional symbolic words measured in a person with hypertrophic cardiomypathy at high of a cardiac arrest: a) original data, b) only RR shuffled and RT as measured c) both time series surrogate d) only RT shuffled and RR as measured. The time series was 4-hours long and measured during the night.

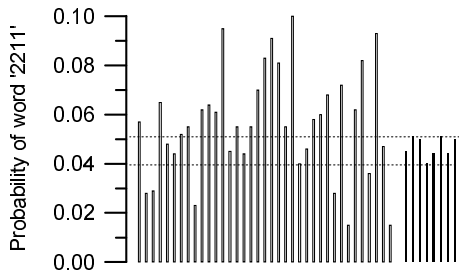

Fig. 21. Probability density of the symbolic word 2211 for 46 persons with hypertrophic cardiomyopathy: open bars — low risk cases, dark bars — persons who died due to cardiac arrest.

rhythm in Fig. 19. We see that the histograms of the 4-dimensional symbolic words allow a detailed analysis of repolarization dynamics as a function of different sequences of heart rate variability.

The significance of the individual symbolic words were analyzed using 8 sex, age and disease matched control pairs each containing a high risk and low risk patient with hypertrophic cardiomyopathy. We subtracted the symbolic word distributions of all controls from such distributions for the respective high risk patients. The differences between the probabilities of all words containing the symbols '11' on the last two positions in a word were significantly different from zero. The largest difference for all pairs was obtained for the word '2211'. This word may be interpreted: it represents the probability that during a period of sustained (5 beats) heart rate below the average in a given window the RT interval length remains stable. For cardiologic norm, this word achieves a probability density of 10% or more. The probability of obtaining the word '2211' for all patients with hypertrophic cardiomyopathy analyzed is shown in Fig. 21. It can be seen that for all high risk patients (they all had died) (black bars in Fig. 21) the probability of obtaining this word falls within a very narrow range. Note that one of the patients in the high risk group died during the study (the first dark bar from the left in Fig. 21) — he had been first classified as low risk by standard methods. The probability of the word 2211 for this person falls exactly within the specified range marked by the two horizontal lines in Fig. 21. If one assumes that the persons, for whom the probability of the word '2211' is within the range marked by horizontal broken lines in Fig. 21, then a criterion for the risk of sudden cardiac death in hypertrophic cardiomyopathy is obtained with rather high figures of merit: sensitivity (*i.e.* the ability of the method to separate the low risk group from the high risk group) of 85 % and specificity (*i.e.* the ability of the method to assign each individual patient to the low risk or the high risk group) of 80 %. Given the fact that

linear statistical methods such as analysis of standard deviation yield figures of merit of the order of 50–60 %, the result seems very encouraging. More research is needed analyze the properties of other symbolic words similar to '2211' in view of enhancing the method further. Also, although the group studied here is not small (46 patients), the sensitivity and specificity of the criterion proposed here need to be verified on a larger group.

5. Conclusions

Recently we have developed methods for the study of the level of regularity in the dynamics of complex systems which may be non-stationary. To study statistical properties of these systems we have used a sweeping time window in which pattern entropy [12] and Renyi entropy [14] was calculated. Furthermore, to study the importance of the sequence of events occurring in the dynamics of the system we have used the sweeping window with a symbolic dynamics scheme [15, 25] involving the calculation of the distribution of algorithmic complexity and, more recently, of the distribution of symbolic words.

In the system studied here *i.e.* the system which controls the variability of the rate of the human heart, statistical regularity and sequential regularity have different properties: they relate to different bio-chemical agents which control heart rate variability, the shape of the statistical distribution of the complexity measures is different and the effect of age on these dynamical properties of the system may be different. In particular, in a subgroup of the population studied by us (mostly younger persons) the level of statistical regularity is strongly correlated with that of sequential regularity. Most of the persons from the older subgroup show no correlation of the level of statistical regularity with sequential regularity. Not that for one dimensional maps and other simple models no distinction is usually made between statistical and sequential regularity.

Another aspect studied by us was the properties of statistical regularity of heart rate variability versus those of heart rate acceleration. It is well known that heart rate is a result of the balance between the activity of the sympathetic nervous system and that of the parasympathetic nervous system. Heart rate acceleration, on the other hand, is a function of the increased activity of the sympathetic nervous system combined with a decrease in the activity of the parasympathetic system. Comparing the 24-hour distributions of pattern entropy of hear rate and heart rate acceleration, we showed that healthy humans may be divided into two groups: in one the statistical properties of heart rate are similar to those of heart rate acceleration *i.e.* the positions of the maxima of both pattern entropy distributions indicate a statistically ordered time series. In the second, the position of the maximum

of one distribution indicates a statistically well ordered time series while the dominant level of entropy for the other distribution large indicates a low level of regularity in the time series. This type of behavior may also be seen in the behavior of the respective pattern entropies with the time. Both the properties of the distributions of pattern entropies of the appropriate signals and their behavior as functions of the time yield a non-linear dynamics image of the fuzzy categories of 'flexible' and 'inflexible' behavior as noticed often during diagnosis but unquantifiable up to now.

On the short time scale, we have studied the statistics of symbolic words obtained both from the heart rate variability and from a hybrid delay coordinate phase space composed both from heart rate variability and heart repolarization times data. We find that the distribution of symbolic words may be used as a diagnostic tool for the assessment of both the risk of cardiac arrest in certain types of heart disorders and as a method of the analysis of the interactions between two different processes occurring in a complex system.

The authors would like to thank J. Kurths, A. Malliani, A. Trzebski and A. Voss for important discussions during this research. The authors gratefully acknowledge the warm support of P. Caminal and of M. Vallverdú. The heart rate variability analysis by means of pattern entropy and algorithmic complexity was supported by KBN Grant 2P03B 020 09 (until June 1998), the study of hypertrophic cardiomyopathy was supported by KBN Grant P05B 062 08 (until February 1998) and the study of the interaction of repolarization times on the heart rate was supported by a Research Fellowship Grant of the European Society of Cardiology in 1996 and Priority Research Grant Program "Bioengineering" of the WUT (May 1996–May 1997).

REFERENCES

[1] P.K. Stein, M.S. Bosner, R.E. Kleiger, B.M. Conger, *Am. Heart J.* **127**, 1376 (1994).

[2] H. Kantz, J. Kurths, G. Mayer-Kress, (ed.), *Nonlinear Analysis of Physiological Data*, Springer Books of Synergetics, Springer, Berlin 1998.

[3] A. Porta, G. Baselli, D. Liberati, N. Montano, C. Cogliati , T. Gnecchi-Ruscone, A. Malliani, S. Cerutti, *Biol. Cyb.* **78**, 71 (1998).

[4] H. Dabire, D. Mestivier, J. Jarnet, M.E. Safar, N.P. Chau, *Am. J. of Phys.-Heart* **C44,** H1290 (1998).

[5] D. Mestivier, H. Dabire, M. Safar, N.P. Chau, *J. Appl. Physiol.* **84**, 1795 (1998).

[6] A. Posiewnik , J. Dąbkowski, *Z. Naturforsch.* **A53**, 112 (1998).

[7] A. Voss, J. Kurths, H.J. Kleiner, A. Witt, N. Wessel, P. Saparin, K.J. Osterziel, R. Schurath, R. Dietz, *Cardiovascular Res.* **31**, 419 (1996).

[8] S.M. Pincus, A.L. Goldberger, *Am. J. Physiol.* **266**, H1643 (1994).

[9] J.K. Kanters, N.H. Holstein-Rathlou, E. Agner, *J. of Card. Electrophys.* **5**, 591 (1994).

[10] M. Paluš, *Biol. Cybern.* **75**, 389 (1996).

[11] M.A. Woo, W.G. Stevenson, D.K. Moser, R.B. Trelease, R.M. Harper, *Am. Heart J.* **123**, 704 (1992).

[12] J.J. Żebrowski, W. Popławska, R. Baranowski, *Phys. Rev.* **50**, 4187 (1994).

[13] J.J. Żebrowski, W. Popławska, R. Baranowski, *Acta Phys. Pol.* **B26**, 1055 (1995).

[14] J.J. Żebrowski, W. Popławska, R. Baranowski, T. Buchner, *Nonlinear Analysis, Theory, Methods & Applications* **30**, 2, 1007 (1997).

[15] J.J. Żebrowski, W. Popławska, R. Baranowski, T. Buchner, Symbolic dynamics and complexity in a physiological time series, submitted to *Chaos, Solitons & Fractals.*

[16] A. Babloyantz , P. Maurer, *Phys. Lett.* **A221**, 43 (1996); A. Babloyantz, P. Maurer, *Nonlinear Analysis-Theory Methods & Applications* **30**, 943 (1997).

[17] P.E. Puddu, R. Jouve, S. Mariotti, S. Giampaoli, M. Lanti, A. Reale, A. Menotti, *J. Electrocardiol.* **21**, 219 (1988).

[18] J.J. Żebrowski, T. Buchner, R. Baranowski, W. Popławska, *Medical & Biological Engineering & Computing* **34**, Suppl. I, 379 (1996).

[19] T. Buchner, J.J. Żebrowski, W. Popławska, R. Baranowski, *Medical & Biological Engineering & Computing* **34**, Suppl. I, 389 (1996).

[20] T. Buchner, J.J. Żebrowski, *Chaos, Solitons and Fractals* **9**, 19 (1997).

[21] J.J. Żebrowski, W. Popławska, R. Baranowski, T. Buchner, *Computers in Cardiology* 377 (1995).

[22] E. Ott, *Chaos in Dynamical Systems*, Cambridge University Press, New York 1993.

[23] D.J. Eving, J.M.M. Neilson, P. Travis, *Br.Heart J.* **52**, 386 (1984).

[24] Hao Bai-Lin, *Elementary Symbolic Dynamics and Chaos in Dissipative Systems*, World Scientific, Singapore 1989.

[25] J.J. Żebrowski, W. Popławska, R. Baranowski, T. Buchner, *J. Techn. Phys.* **38**, 415 (1997).

[26] A. Lempel, J. Ziv, *IEEE Trans. Inf. Theory* **IT-22**, 75 (1976).

[27] F. Kaspar, H.G. Schuster, *Phys. Rev.* **A36**, 842 (1987).

[28] A. Witt, J. Kurths, F. Krause, K. Fischer, *Geophys. Astrophys. Fluid Dyn.* **77**, 79 (1994).

[29] R. Baranowski, M. Vallverdú, R. Jané, P. Caminal, W. Popławska, L. Małecka, L. Chojnowska, W. Rydlewska-Sadowska, X. Viñolas, A. Bayés de Luna, *Circulation* **96**, Suppl. I, 759 (1997).