SYMBOLIC DYNAMICS ANALYSIS OF SHORT DATA SETS: AN APPLICATION TO HEART RATE VARIABILITY FROM IMPLANTABLE DEFIBRILLATOR DEVICES*

Jan J. Żebrowski†

Faculty of Physics, Warsaw University of Technology
Koszykowa 75, 00-662 Warszawa, Poland

Rafał Baranowski and Andrzej Przybylski

National Institute of Cardiology
Alpejska 42, 04-628 Warszawa-Anin, Poland

(Received December 2, 2002)

A method is described for the assessment of the complexity of short data sets by nonlinear dynamics. The method was devised for and tested on human heart rate recordings approximately 2000 to 9000 RR intervals long which were extracted from the memory of implantable defibrillator devices (ICD). It is, however, applicable in a more general context. The ICDs are meant to control life-threatening episodes of ventricular tachycardia and/or ventricular fibrillation by applying a electric shock to the heart through intracardiac electrodes. It is well known that conventional ICD algorithms yield approximately 20–30 % of spurious interventions. The main aim of this work is to look for nonlinear dynamics methods to enhance the appropriateness of the ICD intervention. We first showed that nonlinear dynamics methods first applied to 24-hour heart rate variability analysis were able to detect the need for the ICD intervention. To be applicable to future ICD use, the methods must also be low in computational requirements. Methods to analyse the complexity of the short and non-stationary sets were devised. We calculated the Shannon entropy of symbolic words obtained in a sliding 50 beat window and analysed the dependence of this complexity measure on the time. Precursors were found extending much earlier time than the time the standard ICD algorithms span.

PACS numbers: 05.45.–a, 05.45.Tp, 87.19.Hh, 05.10.–a

---

* Presented at the XV Marian Smoluchowski Symposium on Statistical Physics, Zakopane, Poland, September 7–12, 2002.

† e-mail: `zebra@if.pw.edu.pl`

1. Introduction

Sudden cardiac death is usually caused by malignant ventricular tachyarrhythmias such as ventricular fibrillation (VF) or tachycardia (VT) [1]. Clinical studies showed that implantable cardioverter-defibrillators (ICD) are superior to pharmacological therapy in patients who survived sudden cardiac death [2, 3]. Recent trials documented also their efficacy in patients at high risk of sudden cardiac death (*i.e.* after huge myocardial infarction) [4,5]. Despite the technological progress, inappropriate ICD interventions are still a very important side-effect of this kind of therapy. About 20% of therapies delivered by ICD has been estimated as inappropriate [6]. These are usually caused by supraventricular tachyarrhythmias, T-wave oversensing, noise or non-sustained ventricular arrhythmias [7, 8].

Because it is implanted and power consumption must be low, the ICD algorithms must be kept as simple as possible. Obviously, the algorithms must be fail-safe. The most important detection criterion used by ICD devices is simply the length of consecutive RR intervals *i.e.* the time intervals between heartbeats. Arrhythmia is considered to be detected if a certain number of RR intervals or if x out of consecutive y intervals are shorter than a preprogrammed value [2, 7]. In single chamber devices (with the electrode placed in the ventricles only), additional detection criteria like onset, stability and morphology enhance the specificity. Dual chamber ICDs are able to compare atrial and ventricular rhythm and classify arrhythmia using special built-in algorithms. These detection algorithms are, however, not available when ventricular fibrillation (VF) occurs [7, 8].

Recently recurrence plot analysis [9, 10], symbolic dynamics and short term growth rates [11] have been applied to data extracted from ICD devices with the purpose of enhancing the detection of life-threatening arrhythmias. The main difference in the methods applied as compared to standard algorithms in state of the art ICD devices is that the detection of arrhythmia was not based on any kind of average heart rate but on the heart rate variability preceding VT or VF. Since ICD devices are implanted in severely ill patients so that ventricular arrhythmias and atrial fibrillation are often present at all time, standard linear methods of heart rate variability analysis [12, 13] fail.

The aim of our work is to develop a symbolic dynamics method which may be applied to short, non-stationary data. We apply the method to time series extracted from the ICD. We do not use the length of RR intervals (which the ICD algorithm uses) but rather heart rate variability (HRV) to verify the classification of heart rhythm by the ICD. This paper presents our preliminary results. Since we had only access to RR intervals without ECG verification, it was impossible to use the HRV time or frequency domain methods now standard in cardiology. These methods require arrhythmia filtration and arrhythmia could not be identified without the full ECG. To demonstrate that nonlinear dynamics is applicable to ICD data, we used methods of nonlinear dynamics that were proposed and used by our group in studies of 24-h recordings of HRV [14,15]. We next developed a new method in which we calculate the Shannon entropy of the distribution symbolic words in a sliding 50 heartbeat window. We show that the average over the last 750 of this entropy is a good prognostic of the approaching malignant ventricular arrhythmia and that precursor dynamics is present in the data extending over much longer times than those analyzed by the present ICD algorithms.

2. The data

For preliminary analysis, twenty one recordings obtained from 14 patients were studied. All patients had single chamber devices: Phylax XM — 6 patients, MicroPhylax Plus — 6 patients (Biotronik, Berlin, Germany) and Microjewel II — 2 (Medtronic Inc. Minneapolis, USA). In all cases endocardial leads were used. There were 12 males and two females in the study group ranging in age from 24 to 78 years (mean 54.7 ± 14.3 y). Eleven patients had coronary artery disease, while in three patients hypertrophic cardiomyopathy, dilated cardiomyopathy and idiopathic ventricular fibrillation (VF) had been diagnosed. The left ventricular ejection fraction ranged from 15 to 70% (mean 40.1 ± 14.8%). Nine subjects had a history of VF, 4 a history of ventricular tachycardia (VT) and one patient had the device implanted prophylactically because of hypertrophic cardiomyopathy and a history of sudden cardiac deaths of family members. Patients who had a predominantly paced rhythm were excluded from the study.

RR intervals prior to interventions stored in the ICD memory were analyzed. The RR recordings were exported from the ICD using the PDM 2000 (Biotronik) and STDWIN (Medtronic) programs. The recordings were from 2049 to 9176 RR intervals long (mean 7312 ± 3039).

Episodes of ICD interventions were analyzed and qualified by one of the cardiologists (A.P.) on the basis of ICD Holter memory storage, primarily using intracardiac electrograms prior to arrhythmia detection and subsequent ICD intervention. Arrhythmia-related clinical symptoms were an additional differentiating factor. Judging from the stored intracardiac electrograms and clinical data, 7 of the 21 interventions were found to be inappropriate. They were caused by T-wave oversensing (3 recordings) and by atrial fibrillation (4 recordings).

For symbolic dynamics analysis on a short time scale, the main topic of this paper, the heart rate variability data was organized into sets. There were overall 62 data series from 18 patients. Only interventions found by the cardiologist to be correct were analyzed in this stage of research. Most of the recordings were slightly over 5500 and 9000 intervals in length while some were as short as 2000 intervals (including the data set discussed below). Each data set contained recordings from a single patient only. All were obtained from the Medtronic device and using the same software as described above. The data sets differed in the number of recordings they contained. There were 10 sets of 2 recordings, 3 of 3 recordings, and sets of 10, 8, 6, 5 and 4 recordings, each. Each set data set contained at least a single control recording — measured on demand when there was no need for an intervention — and at least a single recording of RR intervals prior to interventions stored in the ICD memory.

3. Preliminary analysis

In the preliminary stage of the analysis [16], two methods developed earlier for the RR interval time series extracted from 24-hour portable ECG recordings were used — window pattern entropy and algorithmic complexity [14, 15]. Both are based on 3D phase trajectories in delay coordinates. The shape of these trajectories was also analyzed.

Window pattern entropy is a statistical measure of signal variability and calculated [14, 15] as a modified Shannon information entropy: The sliding window length was 50 RR intervals throughout this paper and the window was shifted by one data point. Because an (incomplete) joint probability is used in its definition, pattern entropy is peculiar in that it increases when the time series is more ordered — contrary to the well known properties of Shannon entropy itself. Pattern entropy is given in arbitrary units and, for convenience, all values of pattern entropy were multiplied by 10^4.

Each point in the 3D delay coordinate space was represented by words of three symbols and each 50 beat sliding window — by a string 150 symbols long. Seven different symbols were used [14, 15]: for each of the three delay coordinates, a different pair of symbols indicated whether the RR interval was larger or smaller than the window average, respectively. The seventh symbol was assigned whenever the RR interval was within a prescribed range of the average (defined as the tolerance parameter). This symbol was common to all three delay coordinates and in all our earlier studies was equal to the sampling rate error (7.5 ms). To quantify the complexity of this string the Lempel–Ziv algorithmic complexity was applied [17]. This measure quantifies how complex is the sequence of phase space points within one sliding window. The relation between pattern entropy and algorithmic complexity of heart rate variability as well was discussed elsewhere [14].

In a blind test, the 21 recordings were analyzed by one of the cardiologists (R.B.) and the physicist (J.J.Ż) [16]. We observed an outstanding feature of a correct intervention: it was preceded by a sudden decrease of both algorithmic complexity and of window pattern entropy. The same behavior was observed to precede VT/VF episodes for patients with sinus rhythm as well as for atrial fibrillation cases. This result points to the usefulness of nonlinear dynamics algorithms as it is well known that conventional ICD algorithms may fail if atrial fibrillation occurs. The absence of a significant rapid reduction of algorithmic complexity and window pattern entropy was found to be an indicator of a false intervention. The use of window pattern entropy and algorithmic complexity allowed to identify all 7 of the spurious interventions. Two of the 14 RR interval recordings from appropriate therapies were mislabelled. Note, however, that one of these appropriate interventions was an aborted one *i.e.* the device sensed the conditions for intervention but by the time the condenser had been loaded the rhythm had changed and finally no intervention was delivered.

The high success rate in detecting inappropriate interventions without any changes made to methods initially designed for 24-hour time series analysis indicated that nonlinear dynamics methods may be appropriate to create algorithms for ICD use. The preliminary analysis showed also that precursors of ventricular tachycardia or ventricular fibrillation in the heart rate variability exist but that they are usually very short — most often not exceeding 2–3 minutes. Thus the need arose to develop methods for an analysis in a much shorter time scale than either window pattern entropy or algorithmic complexity allow.

4. Symbolic dynamics in short time scale

To study statistics of the instantaneous patterns in the dynamics of the RR intervals [15] (and also in studying the dynamics of the repolarization processes of the heart tissue [15,18]) we had used distributions of the words of three symbols obtained from the coding described in the preceding paragraph. Similar symbolic word distributions had been used in the past by other groups [11] for 30 minute RR interval time series with good results. In our heart rate variability studies, the symbolic word distributions were, however, averaged over 24 hours and perhaps that was the reason why we did not obtain significant results with this technique.

In the present study, we formed the distribution of words for each sliding window of 50 RR intervals and studied how the distribution changes with the time. We also assigned a complexity measure to the distribution for each window by calculating the Shannon entropy S:

$$S = \sum_{i=1}^{N} p_i \ln p_i \,,$$

where N is the number of different symbolic words (for word length 3 N is equal to 27) and p_i is the probability density of word i.

5. Results

For presentation in this paper, one representative set of 8 recordings of time series of RR intervals of a single patient was chosen. The set contained 2 recordings of control (#1: 2048 beats long and #7: 3726 beats long) — extracted from the device when there was no need for an intervention and six pre-intervention time series all 2048 beats long.

The results for the other sets were the same. However, since some of the data sets contained only two or three recordings (including at least one control) the presentation of a statical assessment of the effectiveness of our method is not possible at this time. What will be demonstrated is the ability of the method to distinguish between dynamics during control-like periods and dynamics during precursors to the triggering of the ICD intervention.

Fig. 1 depicts the raw data (top trace — solid curve) and the entropy (dotted curve) as a function of the RR interval index for the longest of the two control recordings #7. It can be seen that the entropy of the symbolic word distribution oscillates but that the average of the entropy is approximately constant (1.36±0.58 for the whole recording and 1.58±0.47 for the last 750 intervals). Fig. 2 depicts a three-dimensional projection of the distribution of symbolic words as a function of the RR interval index. The words are arranged equidistant in the following order 000, 001, 002, 010,... ,

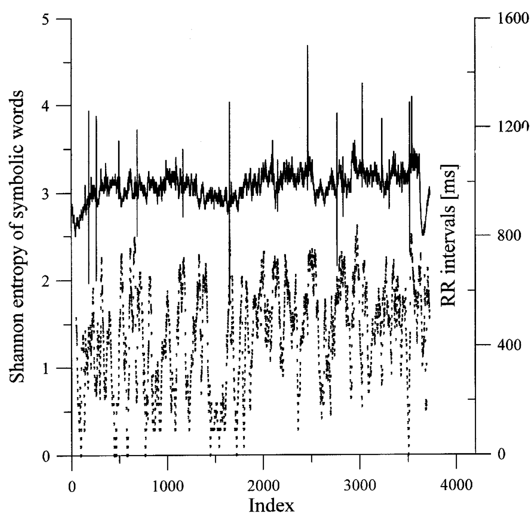

Fig. 1. RR intervals (upper trace) and the Shannon entropy of the symbolic words (dotted curve) as functions of the RR interval index for control recording #7. The index indicates the last RR interval in the sliding window used to construct the symbolic dynamics.

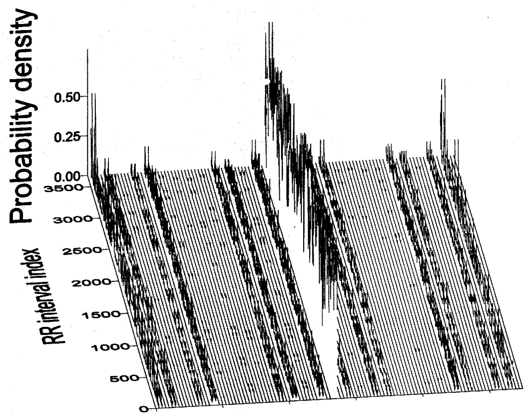

Fig. 2. The distribution of symbolic words as a function of the index for the last RR interval in the sliding window (left axis) for the control recording #7. Vertical axis — probability density of the symbolic words. For the sequence in which the words are depicted — see text.

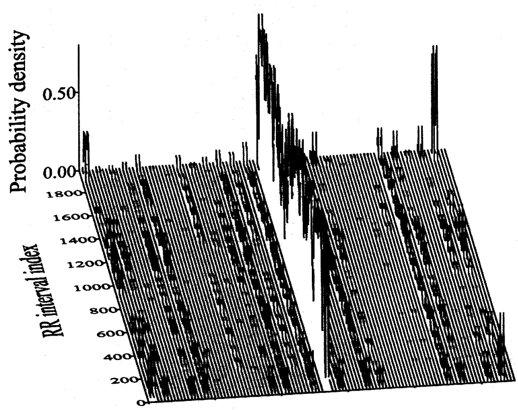

Fig. 3. The distribution of symbolic words as a function of the index for the last RR interval in the sliding window (left axis) for the control recording #1. All notation as in Fig. 2.

111, ... , 221, 222 along the horizontal axis at the bottom of the figure. It can be seen that, for the control recording #7, there are 12 of the 27 possible words which are forbidden: their probability density is less than 0.02. These are the words: 002, 012, 020, 021, 022, 102, 120, 200, 201, 202, 210 and 220.

The distribution of the symbolic words for the shorter of the two control recording (#1) is shown in Fig. 3. In this case two of the words forbidden for the case of #7 become active: these are 021 and 102. Otherwise the 3-dimensional image is strikingly similar. In Fig. 4 it can be seen that the entropy of symbolic words has a slightly lower average (1.08 ± 0.6). However, for the last 750 intervals this entropy was on average 1.2 ± 0.58. Fig. 5 depicts the raw data (top trace) and the entropy of the symbolic

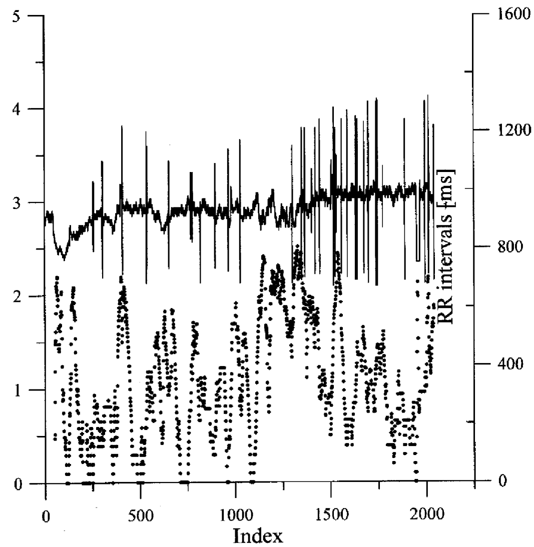

Fig. 4. RR intervals (upper trace) and the Shannon entropy of the symbolic words (dotted curve) as functions of the RR interval index for control recording #1. All notation as in Fig. 1.

word distribution (dotted curve) for a pre-intervention recording: #3. The average entropy of the words is now well below that for the controls (0.84 ± 0.65) while within the last 750 intervals this average becomes 0.58 ± 0.58. In Fig. 6 it can be seen that the decrease of the entropy of the distribution is associated with a sharp increase in the number of forbidden words which begins at approximately at an RR interval index slightly over 1200 and lasts a little past 1600. Within that period only the words around 111 appear. However, the low level of the average entropy in Fig. 5 extends well past that range of the index and lasts till the end of the recording. An even more severe loss of complexity in the symbolic dynamics of the RR intervals was obtained for the recordings #4 and #5. Fig. 7 depicts the raw data (upper trace) and the entropy as functions of the RR interval index (dotted curve) for the latter case — the average entropy was 0.45 ± 0.5 (0.36 ± 0.6 for the last 750 intervals). In the 3D projection in Fig. 8 it can be seen that the only word with a large probability density is 111 — the other words either have only a small probability density or are forbidden. The results for the recording #4 were very similar with the overall average equal to 0.25 ± 0.4.

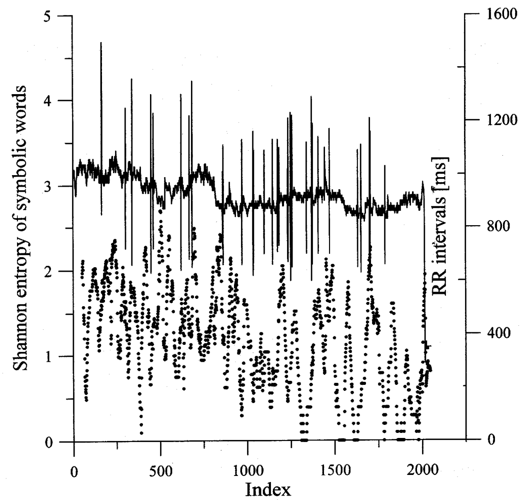

Fig. 5. RR intervals (upper trace) and the Shannon entropy of the symbolic words (dotted curve) as functions of the RR interval index for control recording #3. All notation as in Fig. 1.

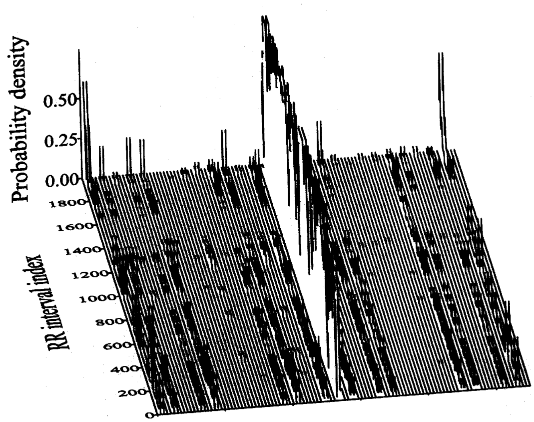

Fig. 6. The distribution of symbolic words as a function of the index for the last RR interval in the sliding window (left axis) for the control recording #3. All notation as in Fig. 2.

These results indicate that all three pre-intervention recordings may readily be recognized by means of the entropy of the symbolic words — it is enough to assume that all recordings for which the last 750 intervals have the entropy below 1 to be those that require the activation intervention procedure. The ease of recognition is due to the long precursor activity

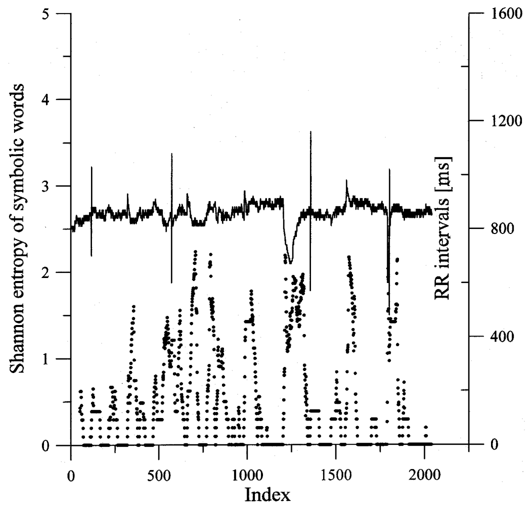

Fig. 7. RR intervals (upper trace) and the Shannon entropy of the symbolic words (dotted curve) as functions of the RR interval index for control recording #5. All notation as in Fig. 1.

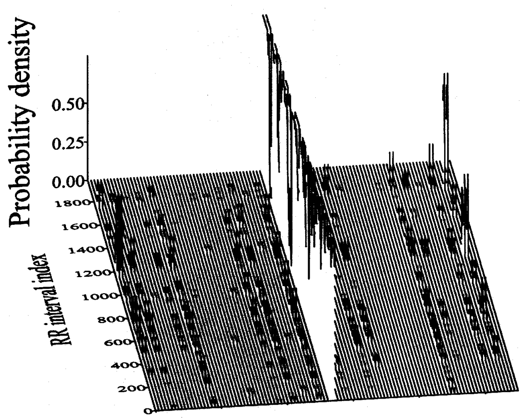

Fig. 8. The distribution of symbolic words as a function of the index for the last RR interval in the sliding window (left axis) for the control recording #5. All notation as in Fig. 2.

that precedes the activation of the standard ICD algorithm in these cases. The criterion is also good for the case #2 for which the precursor is short (Fig. 9). A look at the 3D projection of the distribution of the words as a function of the time (Fig. 10) shows that the decrease in the entropy in this case is due to the symbolic words on the extreme right and left of the figure becoming forbidden. Note that these are words associated with large changes in the RR interval length from beat to beat. Many of the other words present in the control recordings (Fig. 2 and Fig. 3) are, however, present also in Fig. 10.

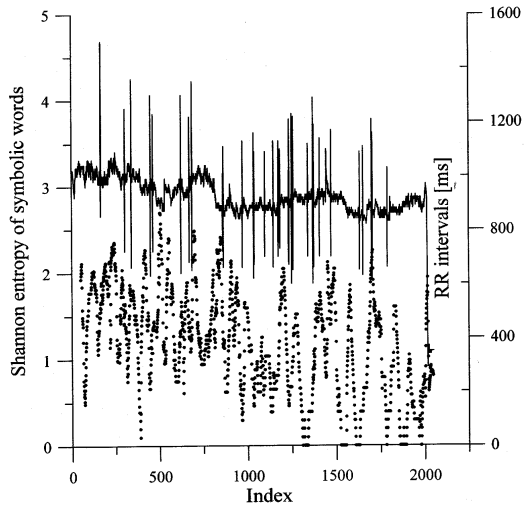

Fig. 9. RR intervals (upper trace) and the Shannon entropy of the symbolic words (dotted curve) as functions of the RR interval index for control recording #2. All notation as in Fig. 1.

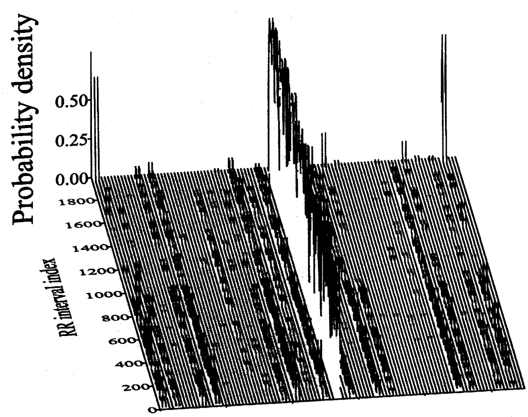

Fig. 10. The distribution of symbolic words as a function of the index for the last RR interval in the sliding window (left axis) for the control recording #2. All notation as in Fig. 2.

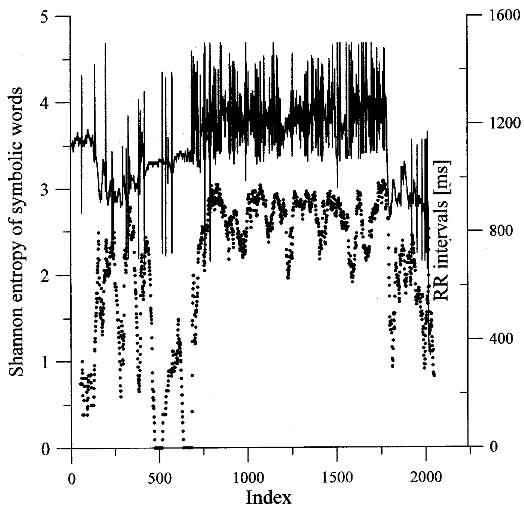

Fig. 11. RR intervals (upper trace) and the Shannon entropy of the symbolic words (dotted curve) as functions of the RR interval index for control recording #6. All notation as in Fig. 1.

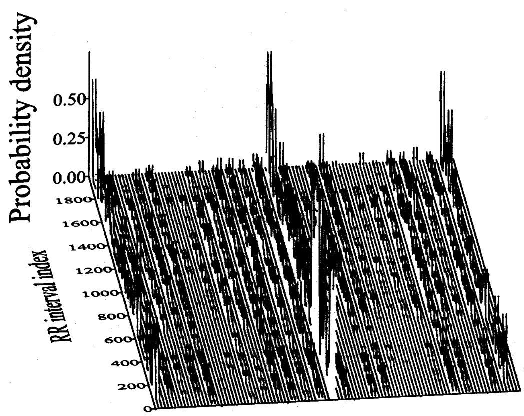

Fig. 12. The distribution of symbolic words as a function of the index for the last RR interval in the sliding window (left axis) for the control recording #6. All notation as in Fig. 2.

The most unexpected result was obtained for the recording #6 (Fig. 11) for which the entropy — except for a brief period staring close to the RR interval index 500 — was actually *much higher* than for both of the control recordings. That this is a case different from all the others in this set may be seen both in the behavior of the raw data (upper trace in Fig. 10) and in the 3D projection of the distribution of the words themselves (Fig. 12). For most of this recording the word 111 is no longer the most probable and the probability density is much more evenly distributed. The nature of the dynamics of the heart rate variability is not clear at this time and will be investigated separately.

6. Conclusions

We have demonstrated that nonlinear dynamics may be used as a tool to enhance the effectiveness of existing algorithms used in implantable cardioverter defibrillator devices. At first, we showed that, even without creating a dedicated algorithm, the nonlinear dynamics methods previously applied by our group to the analysis of 24-hour heart rate variability data when applied to time series extracted from the ICD give meaningful results. We next proposed a new algorithm which is much less computation intensive. This new method requires symbolic coding with respect to the average heart rate within a short (50 beat) sliding window and the assessment of the distribution of symbolic words as a function of the time. As a complexity measure we applied the Shannon entropy of the distribution.

We found that the average of the entropy of symbolic words over the last 750 evolutions is an effective way of predicting the need for an intervention by the ICD device. By this method we are able to demonstrate the existence of a precursor dynamics which precede the ventricular tachycardia or ventricular fibrillation the device senses. Such precursors are usually much longer than the range of heart beats (12 or 16) that the ICD usually analysis.

We found that there exists a range of the average of the entropy of symbolic words which may be considered safe for the patient. To determine whether this range is universal and may be applied to all ICD patients or must be found for each individual patient separately, much more data must be accumulated and analyzed.

This work was supported by the Polish State Committee for Scientific Research (KBN) grant no 6 P05B 085 21.

REFERENCES

[1] D.L. Bayes, P. Coumel, J.F. Leclerq, *Am. Heart J.* **117**, 151 (1989).

[2] Anonymous, *N Engl. J. Med.* **337**, 1576 (1997).

[3] G. Gregoratos, M.D. Cheitlin, A. Conill *et al.*, *J. Am. Coll. Cardiol.* **31**, 1175 (1998).

[4] A.J. Moss, J.W. Hall, D.S. Cannom *et al.*, *N Engl. J. Med.* **335**, 1933 (1996).

[5] A.J. Moss, W. Zareba, W. Jackson Hall *et al.*, *N Engl. J. Med.* **346**, 877 (2002).

[6] J. Brugada, *Am. J. Cardiol.* **83**, (5B0):40D-44D (1999).

[7] K. Nanthakumar, M. Paquette, D. Newmann *et al.*, *Am. Heart J.* **139**, 797 (2000).

[8] M. Rosenquist, T. Beyer, M. Block *et al.*, on behalf of the European 7219 Jewel ICD Investigators, Adverse Events with Transvenous Implantable Cardioverter- Defibrillator. A Prospective Multicenter Study. Circulation. 1998;98:663-670.

[9] N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, J. Kurths, *Phys. Rev.* **E66**, 026702 (2002).

[10] U. Meyerfeldt, N. Wessel, H. Schutt, D. Selbig, A. Schumann, A. Voss, J. Kurths, C. Ziehmann, R. Dietz, A. Schirdewan, *Int. J. Card.* **84**, 141 (2002).

[11] N. Wessel, C. Ziehmann, J. Kurths, U. Meyerfeldt, A. Schirdewan, A. Voss, *Phys. Rev.* **E61**, 733 (2000).

[12] A.L. Goldberger, D.R. Rigney, J. Mietus, E.M. Antman, S. Greenwald, Nonlinear Dynamics in Sudden Cardiac Death Syndrome: Heartrate Oscillations and Bifurcations, *Experientia*, **44**, 983-7 (1988).

[13] L. Glass, D. Kaplan, *Medical Progress Through Technology* **19**, 115 (1993).

[14] J.J. Żebrowski, W. Popławska, R. Baranowski, T. Buchner, *Chaos Solitons & Fractals* **11**, 1061 (2000).

[15] J.J. Żebrowski, W. Popławska, R.Baranowski, T.Buchner, *Acta Phys. Pol. B* **30**, 2547 (1999).

[16] A. Przybylski , R. Baranowski, J.J. Żebrowski, H. Szwed, Verification of Implantable Cardioverter Defibrillator (ICD) Interventions by Nonlinear Analysis of Heart Rate Variability — Preliminary Results, presented at XIII International Congress A World Congress in Cardiac Electrophysiology, Nice 19-22 June 2002, abstract in Europace 2002 vol. 3, Supplement A, p. A14. Paper in preparation.

[17] A. Lempel, J. Ziv, *IEEE Trans. Inf. Theory* **IT-22**, 75 (1976); F. Kaspar, H.G. Schuster, *Phys. Rev.* **A36**, 842 (1987).

[18] R. Baranowski, J.J. Żebrowski, *J. Elektrocardiol.* **35**, 95 (2002).