Forecasting of Life Threatening Arrhythmias Using the Compression Entropy of Heart Rate

M. Baumert¹, V. Baier¹, J. Haueisen², N. Wessel³, U. Meyerfeldt⁴, A. Schirdewan⁴, A. Voss¹

¹University of Applied Sciences Jena, Department of Medical Engineering, Jena, Germany
²Friedrich Schiller University Jena, Department of Neurology, Biomagnetic Center, Jena, Germany
³University of Potsdam, Institute for Physics, Potsdam, Germany
⁴Humboldt-University Berlin, Charité, Franz-Volhard-Hospital, Berlin, Germany

1. Introduction

Sudden cardiac death (SCD) is a leading cause of mortality in the developed countries with an incidence of 3 million cases per year worldwide [1, 2]. SCD is usually caused by a malignant tachyarrhythmia. In clinical trials, implantable cardioverter defibrillators (ICD) have been the most successful therapy to prevent SCD in high risk patients [3, 4]. The detection of ventricular tachycardia (VT) depends on a single ventricular rate sensing signal, a set of programmable detection criteria, and the resulting detection algorithm. Modern ICD offer enhanced detection criteria [5, 6]. However, inappropriate shocks or antitachycardiac pacing remain an important clinical problem in the ICD therapy as they cause unnecessary pain and sometimes proarrhythmic effects [7, 8].

Third generation ICD are capable of storing the beat-to-beat intervals (BBI) before VT. Therefore, an assessment of the autonomous nervous system (ANS) by means of heart rate variability (HRV) analysis has become available opening a completely new perspective on arrhythmogenesis. The ANS tone seems to have direct impact on the VT development [9, 10]. However, studies analyzing in ICD stored BBI data led to different results [11-19]. On the one hand it was discovered that mean heart rate, low frequency power and the low to high frequency power ratio, respectively, were increased before VT, on the other hand no significant HRV changes were found. The discrepancy is probably caused by a modified study design as well as different methods for HRV analysis.

HRV analysis has demonstrated to be a potential risk predictor in cardiac patients and is widely performed in time and frequency domain specified by the Task Force of the European Society of Cardiology and the North-American Society of Pacing and Electrophysiology [20].

As a novel approach we assessed the complexity of BBI time series based on its compressibility. According to Shannon’s information theory there is a limit in encoding a given sequence, its entropy [21]. Beneditto et al. applied zip-coding for entropy estimation in language trees [22]. Hypothesizing that a robust compressibility based entropy measure is suitable to detect HRV changes before the onset of VT we conducted a study in patients with ICD in order to forecast life-threatening arrhythmias.

2. Methods

2.1 Data and Preprocessing

Fifty patients with severe congestive heart failure were enrolled at the Franz-Volhard-Hospital Berlin. No patient received a class I or III antiarrhythmic drug prior to the study. All patients had an implanted ICD (PCD 7220/ 7221 Medtronic) capable of storing 1024 BBI before the onset of a VT with a resolution of 10 ms. HRV analysis of VT time series was performed in comparison with individually acquired and arbitrarily selected control time series (CON) without arrhythmic events that were stored just before a regular ICD follow-up examination (Fig. 1). Artifacts and ectopic beats were
removed and interpolated by an algorithm using a local variance estimation [23].

2.2 Data compression

In 1977 Ziv and Lempel [24] developed an universal algorithm for lossless data compression (LZ77) using string-matching on a sliding window that is nowadays implemented in many tools including ‘zip’ and ‘Stacker’. The algorithm is briefly explained here (Fig. 2):

A sequence of symbols $x = x_1, x_2, \ldots$ of length $L$ from some given alphabet $\Theta = \{\theta\}$ is to be compressed. Subsequences $(x_{m}, x_{m+1}, \ldots, x_{n})$ of $x$ will be denoted by $x_{n}^{m}$.

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 lookahead buffer of size $b$. The encoder positioned at $p$ looks for the longest match of length $n$ between the not-yet-encoded $n$-string $x_{p+n-1}^{p}$ in the lookahead buffer and the already encoded string $x_{p-w+v+n-1}^{p-w+v}$ in the window beginning at position $v$. Thus, the matching string of $n$ symbols is simply encoded by encoding the integers $n$ 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:

a) Encode the first $w$ symbols without compression
b) Set the pointer $p = w+1$
c) Find for some $v$ in the range of $1 \leq v \leq w$ the largest $n$ in the range of $1 \leq n \leq b$ such that $x_{p+n-1}^{p} = x_{p-w+v+n-1}^{p-w+v}$
d) Encode the integers $n$ and $v$ into unary-binary code and the symbol $x_{p+n} \in \Theta$ without compression
e) Set the pointer $p$ to $p = n + 1$ and go to step 3 (iterate)

From the point of information theory the smallest algorithm that produces a string is the entropy of that string (Chaitin-Kolmogorov entropy [25, 26]). Although it is theoretically impossible to develop such an algorithm data compression techniques might be a good approximation. Assuming that the source is an ergodic process the entropy per character $x$ is the length of the compressed string divided by the length of the original string if $L$ tends to $\infty$. 

2.3 Compression Entropy of Heart Rate

Utilizing data compression for HRV analysis the source is the sinus node emitting a sequence of BBI. Hence, the alphabet $\Theta$ consists of different BBI whereas $\Phi$ is mainly affected by the sampling rate $s$. Consequently, the implementation has to consider integer numbers. Furthermore, a transformation of the BBI string into a unary-binary code is unnecessary for entropy estimation. A matrix $M$ of length $K$ storing
Computed $H_c$ values ranged from $H_{c,100,1,1} : 0.77 (0.68-0.85)$ to $H_{c,100,20,10} : 0.39 (0.34-0.49)$. Group VT showed a similar behavior ranging between $H_{c,100,1,1} : 0.75 (0.67-0.82)$ and $H_{c,100,20,10} : 0.39 (0.33-0.43)$.

Figure 4 shows the results of the signed Wilcoxon rank test between CON and VT depending on window and lookahead buffer size. An optimum separation between CON and VT was achieved at a window size of 7 and a lookahead buffer size of 3 ($H_{c,100,7,3} : 0.48 (0.41-0.61)$ vs. $0.49 (0.41-0.53)$; $p = 0.007$). Here, the median relative difference of $H_c$ between CON and VT was 8% (Table 2). Significance by chance due to multiple testing is implausible as similar results were achieved within an area of $2 < w < 10$ and $2 < b < 9$.

### 3.2 Standard HRV Analysis

Table 2 displays the medians of CON and VT, interquartil ranges, relative differences and signed Wilcoxon rank test results of the calculated HRV measures. Except for meanNN that was slightly significantly decreased before VT, the analysis revealed no other significantly changed standard HRV parameter, whereas the new $H_c$ parameter was highly significantly reduced before the onset of VT. The reduction of meanNN is in accordance with findings of other authors [11, 19]. As $H_c$ has some correlation ($r = 0.63$) with meanNN and $H_c$ was in $r = 0.63$.}

#### 3.1 Compression Entropy Depending on Window Size and Lookahead Buffer Size

Figure 3 displays the mean $H_c$ of CON depending on window size and lookahead buffer size. $H_c$ depended in a logarithm-like behavior from window size. Although $H_c$ decreased with an increasing lookahead buffer size its influence was noticeable up to a lookahead buffer size of five. The computed $H_c$ values ranged from $H_{c,100,1,1} : 0.77 (0.68-0.85)$ to $H_{c,100,20,10} : 0.39 (0.34-0.49)$. Group VT showed a similar behavior ranging between $H_{c,100,1,1} : 0.75 (0.67-0.82)$ and $H_{c,100,20,10} : 0.39 (0.33-0.43)$.

Figure 4 shows the results of the signed Wilcoxon rank test between CON and VT depending on window and lookahead buffer size. An optimum separation between CON and VT was achieved at a window size of 7 and a lookahead buffer size of 3 ($H_{c,100,7,3} : 0.48 (0.41-0.61)$ vs. $0.49 (0.41-0.53)$; $p = 0.007$). Here, the median relative difference of $H_c$ between CON and VT was 8% (Table 2). Significance by chance due to multiple testing is implausible as similar results were achieved within an area of $2 < w < 10$ and $2 < b < 9$.

### 3. Results

#### 3.1 Compression Entropy Depending on Window Size and Lookahead Buffer Size

Table 3 displays the medians of CON and VT, interquartil ranges, relative differences and signed Wilcoxon rank test results of the calculated HRV measures. Except for meanNN that was slightly significantly decreased before VT, the analysis revealed no other significantly changed standard HRV parameter, whereas the new $H_c$ parameter was highly significantly reduced before the onset of VT. The reduction of meanNN is in accordance with findings of other authors [11, 19]. As $H_c$ has some correlation ($r = 0.63$) with meanNN and $H_c$ was in $r = 0.63$.

### 4. Discussion

In this paper we introduced a novel method for nonlinear HRV analysis using compression entropy of heart rate. To prove whether HRV analysis provides suitable markers for short-term forecasting of life-threatening arrhythmias in ICD-patients, a set of 50 BBI time series before VT was analyzed. Except for meanNN that was slightly significantly decreased before VT, the analysis revealed no other significantly changed standard HRV parameter, whereas the new $H_c$ parameter was highly significantly reduced before the onset of VT. The reduction of meanNN is in accordance with findings of other authors [11, 19]. As $H_c$ has some correlation ($r = 0.63$) with meanNN and $H_c$ was in $r = 0.63$.
data with a meanNN differing more than 200 ms were rejected (40 data pairs remained) and the statistical analysis was repeated subsequently. As a result, the significant difference in meanNN was lost whereas the significance in $H_c$ was preserved ($p = 0.03$) demonstrating the potential contribution of $H_c$.

The intracardiac electrograms preceding a VT were recorded under daily life conditions. Thus, a non-stationary BBI behavior negatively influencing standard HRV parameters has to be considered. Besides ectopic beats, the BBI detection algorithm of ICD might also cause artifacts. Therefore, the originally acquired BBI time series had to be filtered subsequently. Consequently, future ICD performing HRV analysis should necessarily feature a powerful filter algorithm that minimizes risk of erroneous BBI estimation to avoid inappropriate shocks. However, due to the sliding window technique, $H_c$ is relatively insensitive to artifacts and instationarity which have only a locally limited influence to the window and lookahead buffer size.

The variations of lookahead buffer size and window size revealed an optimum separation between VT and CON, i.e. the highest significance, at relatively small sizes ($w = 7, b = 3$). Anyhow, the analyses showed that within a narrow range both parameters might be varied without a major loss of significance (see Fig. 4). Apparently, reduced short-term fluctuations of HRV resulted in an increased compression in VT. This might be in consistence with the findings by Wessel et al. [12] and is supposedly caused by the shift of sympathovagal balance toward sympathetic predominance and reduced vagal tone of the ANS [11].

Furthermore, the gradient of $H_c$ in dependency of $w$ and $b$ (Fig. 4) suggests a sensitivity of $H_c$ to vagally rather than sympathetically mediated components of HRV. Since compression entropy analysis represents a short-term method longer lasting patterns reflecting especially sympathetically mediated influences are not or only partially considered for compression.

Besides window size and lookahead buffer size, $H_c$ probably depends on the sampling rate and should therefore be a subject of further investigations. Furthermore, a larger data set should be acquired to validate the presented results but also to quantify and classify different mechanisms of arrhythmia development. Future analyses should also investigate the influence of VT cycle length and different arrhythmia substrates on its predictability [13].

In conclusion, the nonlinear parameter $H_c$ revealed significant changes in HRV before the onset of VT and might be suitable...
for the short-term forecasting of life-threatening arrhythmias in ICD in order to improve VT sensing but also to enable an early patient warning of forthcoming shocks.

Acknowledgments
This study was supported by grants from the Deutsche Forschungsgemeinschaft (DFG Vo505/3-1 and DFG Vo505/4-2).

Table 2
Results of HRV analysis. Median, interquartile range and Wilcoxon test results (n.s. – not significant)

<table>
<thead>
<tr>
<th>parameter</th>
<th>CON median [ms]</th>
<th>CON interquartile range</th>
<th>VT median [ms]</th>
<th>VT interquartile range</th>
<th>relative difference [%]</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>meanNN [ms]</td>
<td>745</td>
<td>656–853</td>
<td>699</td>
<td>585–802</td>
<td>10</td>
<td>-9.52</td>
</tr>
<tr>
<td>sdNN [ms]</td>
<td>50</td>
<td>11–28</td>
<td>70</td>
<td>23–56</td>
<td>15</td>
<td>25</td>
</tr>
<tr>
<td>rmsd [ms]</td>
<td>17</td>
<td>12.5–132.7</td>
<td>17.7</td>
<td>4.9–114.8</td>
<td>23</td>
<td>-100</td>
</tr>
<tr>
<td>P [ms²]</td>
<td>36.1</td>
<td>12.5–132.7</td>
<td>7.0</td>
<td>2.3–71.2</td>
<td>36</td>
<td>-89</td>
</tr>
<tr>
<td>VLF [ms²]</td>
<td>16.1</td>
<td>5.1–52.1</td>
<td>3.3</td>
<td>0.5–15.5</td>
<td>30</td>
<td>-33</td>
</tr>
<tr>
<td>LF [ms²]</td>
<td>4.2</td>
<td>1.4–10.8</td>
<td>1.0</td>
<td>0.4–5.2</td>
<td>6</td>
<td>-65</td>
</tr>
<tr>
<td>HF [ms²]</td>
<td>1.4</td>
<td>0.4–5.2</td>
<td>2.0</td>
<td>1.1–3.2</td>
<td>21</td>
<td>-23</td>
</tr>
<tr>
<td>LF/HF [n.u.]</td>
<td>2.6</td>
<td>1.8–4.1</td>
<td>0.49</td>
<td>0.41–0.53</td>
<td>8</td>
<td>-6</td>
</tr>
<tr>
<td>H. [n.u.]</td>
<td>0.48</td>
<td>0.42–0.61</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

References