31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Discrimination of endocardial electrogram disorganization using a signal regularity analysis

D. Novák, V. Křemen, D. Cuesta, K. Schmidt, V. Chudáček, L. Lhotská

Abstract— **Measures from the theory of nonlinear dynamics were applied on complex fractionated atrial electrograms (CFAEs) in order to characterize their physiological dynamic behavior. The results were obtained considering 113 short term atrial electrograms (A-EGMs) which were annotated by three experts into four classes of fractionation according to A-EGMs signal regularity. The following measures were applied on A-EGM signals: General Correlation Dimension, Approximate Entropy, Detrended Fluctuation Analysis, Lempel-Ziv Complexity, and Katz-Sevcik, Variance and Box Counting Fractal Dimension. Assessment of disorganization was evaluated by a Kruskal Wallis statistical test. Except Detrended Fluctuation Analysis and Variance Fractal Dimension, the CFAE disorganization was found statistically significant even for low significant level $\alpha = 0.001$. Moreover, the increasing complexity of A-EGM signals was reflected by higher values of General Correlation Dimension of order 1 and Approximate Entropy.**

I. INTRODUCTION

Endocardial sites generating complex fractionated atrial electrograms (CFAEs) have been reported as ablative targets for the treatment of atrial fibrillation (AF) [1]. In order to identify those sites, a great effort has been made to describe patterns of activation in AF [2] and to quantify general characteristics of CFAEs either in time- or frequency-domain [3]. However, the process of a CFAE identification is highly dependent on the operator's judgment. Moreover, it is not clear if CFAE are a random process of a local atrial electrogram disorganization or a reproducible physiological effect [4]. This study is aimed at applying a signal regularity analysis for the description of spatio-temporal changes of the A-EGMs fractionation levels. This analysis enables to investigate A-EGMs nonlinear dynamics and confirm the hypothesis that high degrees of A-EGMs fractionation are also reflected by higher values of the regularity measures.

II. METHODOLOGY

A. Experimental Data Set

Atrial bipolar electrograms were collected during left-atrial endocardial mapping using 4-mm irrigated-tip ablation catheter (NaviStar, Biosense-Webster) in 12 patients (9 males, aged 56 $\pm$ 8 years) with persistent AF. Sampling frequency was 977Hz. 113 of the A-EGMs were manually selected and cropped by an expert. For the purposes of the study (see Fig. 1), four Classes of Fractionation (CF) were set. The dataset ranking, which was done by three independent experts, resulted in a total of 339 rankings (113*3 = 339). The three independent experts never disagreed in their ranking by more than one neighboring CF. Therefore, the most prevailing pattern was chosen as a final score. The four CFs enabled to get a uniform dataset of A-EGMs with a significant number of samples in each class ($class_1 : C_1 = 22$, $class_2 : C_2 = 42$, $class_3 : C_3 = 36$, $class_4 : C_4 = 13$), so that such dataset could be used in a regularity analysis [5],[6].

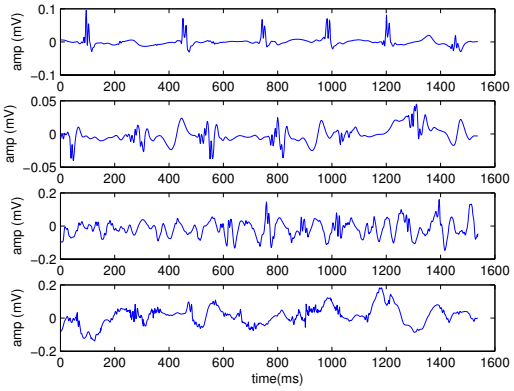

Fig. 1. Epochs of four complex fractionated atrial electrograms. These are representatives of each CF used in the study. From top to bottom: i) $class_1$: Organized activity, ii) $class_2$: Mild degree of fractionation. iii) $class_3$: Intermediate degree of fractionation. iv) $class_4$: High degree of fractionation.

B. Regularity Signal Measures

Before attempting to calculate the fractal dimension for endocardial electrograms, it is important to establish evidence that these waveforms may be characterized as fractals. Fractal dimension values are related to the regularity of a pattern, or the quantity of information embedded in a pattern in terms of morphology, entropy, spectra or variance [7]. Considering morphological properties of A-EGMs, it appears that these signals possess valid fractal dimension values, mainly because of two reasons. First, the signals do not self-cross. By looking at any one of A-EGMs waveforms in Fig.1, it is apparent that in order to scale it, a different scaling factor is required for each axis. This indicates that the A-EGM waveform is self-affine. Second, the signals exhibit clear quasi-periodicity because they emerge from natural repetitive processes (heart beats specifically). In the next subsections, we will describe the signal regularity measures used in this study.

D. Novák, V. Křemen, V. Chudáček, L. Lhotská are with Department of Cybernetics, Czech Technical University in Prague, Czech Republic, xnovakd1@labe.felk.cvut.cz. D. Cuesta is with Department of Computer Science,Polytechnic University of Valencia, Spain. K. Schmidt is with Faculty of Biomedical Engineering, Department of Biomedical Technology, Czech Technical University in Prague.

C. General Correlation Dimension

1) Parameter Selection and Signal Embedding: Only a time series of one variable was used as an input, but this time-series data was used to reconstruct a multidimensional embedding space [7]. It is necessary to determine the following three parameters i) *time delay* between sampled values, ii) *time interval* between successive vectors and iii) *embedding dimension*.

The *time delay* between sampled values was estimated using the auto mutual information. The first minimum of the auto mutual information was preferred value for attractor reconstruction from time series [7].

The *time interval between successive vectors* was set to be equal to the sampling time. In this case the number of vectors was equal to the number of samples in the time series minus the embedding dimension [7].

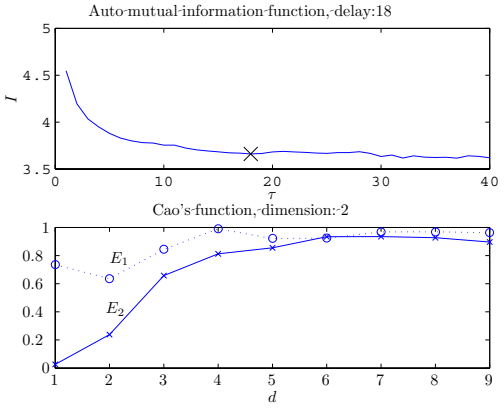

Fig. 2. Auto-mutual information and Cao's method functions $E_1(d)$ and $E_2(d)$ for a signal in $class_1$. The first minimum of the auto mutual function appears at time delay $\tau = 15$. regarding embedding dimension, the optimal value is found in two dimensional space.

To determine the optimal *embedding dimension* from the time series the reliable and commonly used Cao's method was used [8]. We used two measures: $E_1(d)$ which is the average measure of a ratio of the Euclidian distance between the reconstructed vector and its nearest neighbor and $E_2(d)$ which is useful to distinguish deterministic signals from stochastic signals. We were looking for the first minimum either in $E_1(d)$ or $E_2(d)$ as shown in Fig.2. In practical computations, it was difficult to resolve whether the $E_1(d)$ was slowly increasing or had stopped changing if d was sufficiently large. Therefore, the $E_2(d)$ quantity was considered [8].

2) Correlation Dimension: The correlation dimension is computed most efficiently by the correlation sum:

$$C(d,r) = \frac{1}{N_{pairs}} \sum_{i=d}^{N} \sum_{j<i-w} \Theta(r - (y_i - y_j)) \qquad (1)$$

where y_i are m-dimensional delay vectors. $N_{pairs} = (N-d-w)(N-d-w+1)/2$ is the number of pairs of points covered by the sums. Θ is the Heaviside step function: $\Theta(x)$ is zero for $x < 0$ and one for $x \geq 0$. The summation counts the number of pairs (y_i, y_j) for which the distance $|y_i - y_j|$ is less than r. All pairs of points in (3) whose time indices differ by less than w were ignored in order to exclude temporally correlated points. The fractal dimension of order 2,D_2, is than defined as:

$$D_2 = \lim_{r \to 0} \frac{\log C(r)}{\log r} \qquad (2)$$

3) General Correlation Dimension: GCD are a class of metrics to characterize the fractality [7]. It is based on counting the number of points in a box. Let B_i denote the ith box, and let $P_i = \mu(\mathcal{B}_i)/\mu(\mathcal{A})$ be the normalized measure of this box, where $\mathcal{A}$ is the fractal whose dimension has to be computed and μ is the population mean of a set. In other words, P_i is the probability for a randomly chosen point on the attractor to be in $\mathcal{B}_i$, and it is usually estimated by counting the number of points that are in the ith box and dividing by the total number of points. If the attractor had been embedded in dimension d, hypercubes of dimension d would be used. The generalized dimension is defined by:

$$GCD_q = \frac{1}{q-1} \lim_{r \to 0} \frac{\log \sum_i P_i^q}{\log r} \qquad (3)$$

D. Approximate Entropy

ApEn is a statistic measure that quantifies the unpredictability of fluctuations in a time series such as an instantaneous heart rate time series [9]. Informally, given N points, the family of statistics $ApEn(m, r, N)$ is approximately equal to the negative average natural logarithm of the conditional probability that two sequences that are similar for m points remain similar, that is, within a tolerance r, at the next point. Thus, a low value of ApEn reflects a high degree of regularity.

E. Detrended Fluctuation Analysis

DFA quantifies the presence or absence of long-range correlations. This technique is a modification of root-mean-square analysis of random walks applied to nonstationary data. Briefly, the time series to be analyzed (with N samples) is first integrated [9]. Next, the integrated time series is divided into boxes of equal length, n. In each box of length n, a least squares line is fit to the data (representing the trend in that box). The y coordinate of the straight line segments is denoted by $y_n(k)$. Next, the integrated time series, $y(k)$ is detrended by subtracting the local trend, $y_n(k)$, in each box. The root-mean-square fluctuation of this integrated and detrended time series is calculated.

F. Lempel-Ziv Complexity

LZC and its variants have been used widely to identify nonrandom patterns in biomedical signals obtained across distinct physiological states [10]. In general, LZ complexity measures the rate of generation of new patterns along a sequence and in the case of ergodic processes is closely related to the entropy rate of the source.

G. Katz-Sevcik Fractal Dimension

The Katz-Sevcik algorithm for obtaining a fractal dimension is largely based on morphology, and is calculated as [11] $KFD = \log_{10} n / \log_{10} n + log_{10} \frac{g}{L}$ where n is the number of increments between samples of the signal over which KFD is calculated; L is the sum of all the distances between successive increments; and g is the value of the maximum distance measured from the beginning of the first increment. It has to be pointed out that before calculating KFD, normalization along the y- and x-axes of the signal is performed.

H. Variance Fractal Dimension

The VFD is determined by the Hurst exponent, H, whose calculation was derived from the properties of fractional Brownian motion [12]. This calculation is based on the power law relationship between the variance of the amplitude increments of a signal, $\mathcal{C}(t)$, which was produced by a dynamical process over a time increment $\triangle t = \mid t2 - t1 \mid$, with $\mathcal{C}(t2) - \mathcal{C}(t1)$ denoted as $\triangle \mathcal{C}$. The power law is as follows: $Var[\triangle \mathcal{C}] \sim \triangle t^{2H}$ where the Hurst exponent is:

$$H = \lim_{\triangle t \to 0} \left(\frac{\log_2 Var[\triangle \mathcal{C}]}{\log_2 \triangle t} \right) \tag{4}$$

The VFD for a process with embedding Euclidean dimension, E (equal to 1 for A-EGMs signal), is determined by: $VDF = E + 1 - H$.

I. Box Counting method

An established approach to compute the fractal dimension of a set is the box-counting method (BFD) [13]. In detail, for a set of N points in $\mathcal{R}^d$, and a partition of the space in grid cells of length l, the fractal dimension DB can be derived from:

$$BFD = -\lim_{l \to 0} \frac{\log_{10} N(l)}{\log_{10}(l)} \tag{5}$$

where $N(l)$ represents the number of cells occupied by at least one point.

J. Statistical Evaluation

Non-parametric Kruskal Wallis test was applied for regularity measures comparison. The test was used in order to cope with a smaller number of A-EMG signals, specially in class 4, where 13 signals are only available.

III. Results and Discussion

The reliability of the A-EGM fractionation assessment using regularity measures is summarized in Table I. Mean and standard deviation values are reported for each class.

The optimum parameter setting for ApEn, DFA and BFD was found by a trial and error method for 10 runs of the experiment. In case of ApEn, tolerance parameter r was increased by 0.05 while keeping pattern length m fixed. Considering DFA, parameters for defining the slope of two DFA curves were increased by 4 in each run. Finally, BFD dimension dim_{BFD} was incremented by 1 in each run. The final parameters that resulted in minimum p-values across 10 runs were following: GCD: embedding dimension-variable for each signal, Cao's method was applied, time delay: variable for each signal, auto-mutual information was applied; ApEn: tolerance $r = 0.1$, pattern length $m = 2$; DFA: $fast = 2, mid = 32, slow = 64$; BFD: dimension $dim_{BFD} = 5$; LZC: binary coding was used.

TABLE I
Comparison of A-EGM fractionation

Method	C1	C2	C3	C4	Kruskal
GCD_{-6}	0.034 0.029	0.061 0.049	0.078 0.079	0.13 0.078	3.2e-7
GCD_{-5}	0.057 0.021	0.071 0.065	0.11 0.099	0.17 0.076	4.9e-6
GCD_1	**1.22** **0.27**	**1.50** **0.41**	**1.80** **0.30**	**2.16** **0.13**	**8.4e-11**
ApEn	**0.28** **0.088**	**0.46** **0.096**	**0.57** **0.070**	**0.71** **0.050**	**6.9e-13**
DFA	1.24 0.26	1.31 0.22	1.41 0.16	1.41 0.15	1.8e-3
LZC	0.37 0.040	0.47 0.061	0.51 0.071	0.60 0.054	9.1e-13
KFD	1.36 0.020	1.38 0.021	1.38 0.028	1.41 0.020	2.4e-8
VFD	1.78 0.17	1.67 0.17	1.64 0.14	1.65 0.13	0.1e-3
BFD	1.43 0.040	1.47 0.036	1.50 0.042	1.54 0.038	1.5e-9

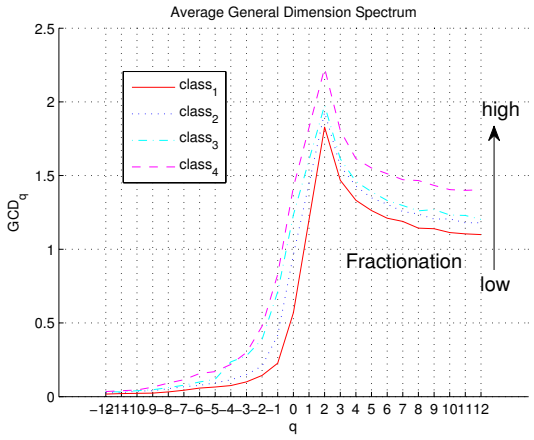

Fig. 3. The average generalized dimension spectrum for all classes of A-EGM fractionation.

Fig.3 shows the average generalized dimension spectrum. The spectrum should be convex, monotonically increasing [7]. The differences among classes are more apparent from general dimension $q \geq -2$. Especially, GCD values of $class_3$ are bigger than values of $class_4$ for generalized dimension $q = -4$. GCD measure reaches peak at $q = 2$ indicating that higher embedding dimensions $q > 2$ do not provide a reliable measure of discrimination due to the appearance of numerical errors in higher dimensional space. The GCD values for $q = -6, -5, 2$ were reported in Table I in order to point out the importance of using higher negative correlation dimensions.

Our assumption about fractal nature of A-EGM signals is confirmed by very low p-values in Table I. Even if significance level is set to $\alpha = 1e-3$, most fractal dimensions are statistically significant in differentiating A-EGM classes disorganization. The exception are DFA and VFD measures. Furthermore, LZC,VFD and BFD measures have lower discrimination capability compared to GCD_1 and ApEn measures as can be seen in Table I. The main reason is data insufficiency; A-EGMs are very short segments of 1.5s duration sampled with 977Hz which accounts for 1537 values. First, ApEn is able to work properly if input dataset contains at least 1000 samples [9]. Second, the good discrimination was mainly achieved by careful selection of algorithm parameters, especially in case of the GCD measure where auto-mutual infirmation and Cao's approach were applied.

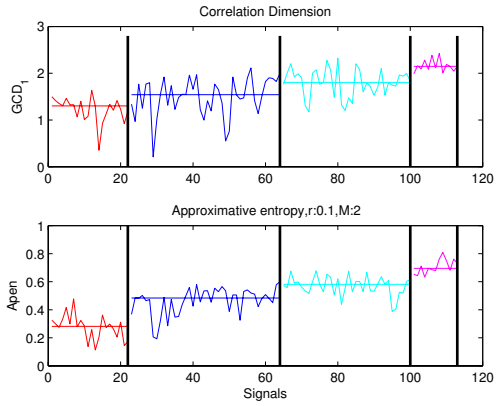

Fig. 4. Assessment of A-EGM fractionation using Correlation Dimension and Approximate Entropy. Horizontal line represents mean value for each class.

We compared two best performing dimensions according to the results of the statistical tests: General Correlation Dimension of order 1 and Approximate Entropy - see bold values in Table I. Fractal dimensions were calculated for all A-EGM signals in the dataset. General correlation dimension of order 1 is shown in upper part of Fig. 4. The first and the last classes are well separated by GCD_1 values. Note that intermediate classes 2 and 3 partially overlap. In case of ApEn, the separation of the 2nd and 3th class is slightly better compared to GCD_1 dimension as can be seen in lower part of Fig. 4.

IV. CONCLUSIONS

In the era of catheter ablation of AF, the initial attempts to describe A-EGMs during AF were predominantly based on frequency-domain analysis of atrial signals [3]. Not only dominant frequency (DF) but also the level of A-EGMs fractionation may be a clinically important descriptor of local atrial signal [1]. Sites with highly fractionated A-EGMs almost fully encompass the sites with high DF whereas the opposite is not always true. Therefore it is very important to develop other measures which could assess A-EGMs regularity, especially differentiating low A-EGMs fractionation (LF) from high fractionation (HF).

This study revealed the presence of nonlinear dynamics in A-EGM across all selected fractionation levels. Fig. 4 documents the fact that with increasing complexity of A-EGM's signals the value of regularity measures increases too. Across all levels of A-EGM fractionation, the discrimination using regularity measures was found statistical significant at significance level $\alpha = 0.001$. The proposed complex measures were successful in separating class C_1(LF) from class C_4 (HF) of the A-EGM signals. However, separation between intermediate classes 2 and 3 was not so clear. One of the reasons is that the classes were defined by averaging the classification by 3 experts to obtain semi-continuous scale of fractionation. In some cases it was very difficult to assign A-EGM signals either to class 2 or class 3.

Regarding future studies, the selected complex measures (GCD_1 and ApEn) can be used as additional features for automated and operator independent system that facilitates AF substrate ablation [5].

ACKNOWLEDGMENT

The project was supported by Ministry of Education, Youth and Sport of the Czech Republic with the grant No. MSM6840770012 entitled "Transdisciplinary Research in Biomedical Engineering II" and by Spanish Ministry of Science and Innovation, grant TEC2008-05871/TEC.

REFERENCES

[1] K. Nademanee and et al, "A new approach for catheter ablation of atrial fibrillation: mapping of the electrophysiologic substrate," *J Am Coll Cardiol.*, vol. 43, pp. 2044–2053, 2004.

[2] K. Kumagai, "Patterns of activation in human atrial fibrillation," *Heart Rhythm*, vol. 4, no. 3, pp. S7–S12, 2008.

[3] J. Ng, J. Gold, and J. Goldberger, "Understanding and interpreting dominant frequency analysis of af electrograms," *Journal of Cardiovascular Electrophysiology*, vol. 18, no. 6, pp. 680–685, 2007.

[4] B. H. et al, "Nonlinear analysis of epicardial atrial electrograms of electrically induced atrial fibrillation in man," *J Cardiovasc Electrophysiol.*, vol. 6, no. 6, pp. 419–440, 1995.

[5] V. Křemen, L. Lhotská, R.Čihák, V.Vančura, J. Kautzner, and D. Wichterle., "A new approach to automated assessment of endocardial electrograms fractionation in human left atrium during atrial fibrillation," *Physiol. Measurement*, vol. 29, pp. 1371–1381, 2008.

[6] V. Křemen, "Automated assessment of endocardial electrograms fractionation in human," Ph.D. dissertation, Czech Technical University in Prague, 2008.

[7] R. Hilborn, *Chaos and Nonlinear Dynamics*. Oxford University Press, 2000.

[8] L.Cao, "Practical method for determining the minimum embedding dimension of a scalar time series," *Physica D*, vol. 110, no. 1-2, pp. 43–50, 1997.

[9] K.Ho and et al., "Predicting survival in heart failure case and control subjects by use of fully automated methods for deriving nonlinear and conventional indices of heart rate dynamics," *Circulation*, vol. 96, no. 3, pp. 842–848, 1997.

[10] J.Hu, J.Gao, and J.Principe, "Analysis of biomedical signals by lemepl-ziv complexity: the effect of finite data size," *IEEE Trans. Biomed. Engg.*, vol. 53, no. 12, pp. 2606–2609, 2006.

[11] C. Sevcik, "On fractal dimension of waveforms," *Chaos, Solitons & Fractals*, vol. 28, no. 2, pp. 579–580, 2006.

[12] J. Gnitecki and Z.Moussavi, "Variance fractal dimension trajectory as a tool for heart sound localization in lung sounds recording," in *Proc. IEEE Eng. Med. Biol Soci. (EMBS)*, 2003, pp. 2420–2423.

[13] S. Liebovitch and T. Toth, "A fast algorithm to determine fractal dimensions by box-counting," *PHYS-LETT-A*, vol. 141, no. 8-9, pp. 386–390, 1989.