

Optimal Parameters Study for Sample Entropy-Based Atrial Fibrillation Organization Analysis

Raúl Alcaraz¹, Daniel Abásolo², Roberto Hornero², José J. Rieta³

¹ Innovation in Bioengineering Research Group, University of Castilla-La Mancha, Spain.

² Biomedical Engineering Group, University of Valladolid, Spain.

³ Biomedical Synergy, Electronic Engineering Department, Universidad Politécnica de Valencia, Spain

Abstract

Sample Entropy (SampEn) is a nonlinear regularity index that requires the *a priori* selection of three parameters: the length of the sequences to be compared, m , the patterns similarity tolerance, r , and the number of samples under analysis, N . Appropriate values for m , r and N have been recommended and widely used in the literature for the application of SampEn to some physiological time series, such as heart rate, hormonal data, etc. However, no guidelines exist for the selection of that values in other cases. Therefore, an optimal parameters study should be required for the application of SampEn to not previously analyzed biomedical signals. In the present work, a thorough analysis on the optimal values for m , r and N is presented within the context of atrial fibrillation (AF) organization estimation, computed from surface electrocardiogram recordings. Recently, the evaluation of AF organization through SampEn, has revealed clinically useful information that could be used for a better treatment of this arrhythmia. The present study analyzed optimal SampEn parameter values within two different scenarios of AF organization estimation, such as the prediction of paroxysmal AF termination and the electrical cardioversion outcome in persistent AF. As a result, interesting recommendations about the selection of m , r and N , together with the relationship between N and the sampling rate (f_s) were obtained. More precisely, (i) the proportion between N and f_s should be higher than one second and $f_s \geq 256$ Hz, (ii) overlapping between adjacent N -length windows does not improve AF organization estimation with respect to the analysis of non-overlapping windows, and (iii) values of m and r maximizing successful classification for the analyzed AF databases should be considered within a range wider than the proposed in the literature for heart rate analysis, i. e. $m = 1$ and $m = 2$ and r between 0.1 and 0.25 times the standard deviation of the data.

Keywords: Atrial Fibrillation, Electrocardiogram, Main Atrial Wave, Sample Entropy, Parameters Selection, Nonlinear Analysis

Address for correspondence:

Raúl Alcaraz Martínez, E. U. Politécnica, Campus Universitario, 16071, Cuenca, Spain.

Phone: +34-969-179-100 Ext. 4847, 4768; Fax: +34-969-179-119.

e-mail: `raul.alcaraz@uclm.es`

1 Introduction

The application of nonlinear regularity metrics to physiological signals is a valuable tool because “hidden information” related to underlying mechanisms can be obtained [1, 2]. To this respect, the employment of sample entropy (SampEn) to estimate non-invasively atrial fibrillation (AF) organization has revealed clinically useful information, which could be used for a better treatment of the arrhythmia [3–5]. The organization estimation of AF is a relevant aspect to improve its comprehension, since its mechanisms are still unexplained [6] despite of affecting up to 1% of the general population [7, 8].

Given a time series with N data points, the a priori selection of two unknown parameters, m and r , is required to compute SampEn [2]. The parameter m determines the length of the sequences to be compared and r is the tolerance for accepting similar patterns between two segments. Although these parameters are critical in determining the outcome of SampEn, no guidelines exist for optimizing their values. Typically recommended m and r values are $m = 1$ and $m = 2$ and r between 0.1 and 0.25 times the standard deviation (SD) of the data [9]. This recommendation is largely based on the application of approximate entropy (ApEn) to relatively slow dynamic signals such as heart rate [1, 9, 10] and hormone secretion data [11]. Given that SampEn is a modified version of ApEn to solve its shortcomings, such as bias, relative inconsistency and dependence on the sample length [2], these values are also applicable to SampEn.

However, a recent work has demonstrated that the typically recommended values for ApEn are not always appropriate for fast dynamic signals [12]. As a consequence, since only few values, within the range suggested in the literature [9], were tested in the previous works where SampEn was applied to AF organization estimation [3–5], the main goal of the present study is to carry out an in depth analysis on SampEn parameters able to achieve optimized classification of AF events which are directly dependent on AF organization. Thus, several combinations of m and r , over a range wider than the one typically recommended, together with the number of analyzed samples, N , and the overlapping effect between adjacent N -length windows have been analyzed. Additionally, since in AF organization estimation, SampEn is applied to a time series that depends on the original ECG

sampling rate (f_s), the relationship between N and f_s is also addressed.

Due to the lack of a method able to generate AF signals with a priori controlled organization, the use of simulated signals is not possible. As a consequence, real AF signals with organization-dependent events were selected. Two different scenarios, such as the prediction of paroxysmal AF termination and the electrical cardioversion (ECV) outcome in persistent AF, in which organization plays an important role, as corroborated by invasive recordings [13], were analyzed.

2 Materials

In the present work, two databases were used with the aim of providing general recommendations, for optimal SampEn computation, that are applicable to a wide range of AF studies in which organization has to be estimated, like ablation guiding or drug effects. Firstly, a set of paroxysmal (i.e., spontaneously terminating) AF recordings were analyzed to predict the termination of the arrhythmia and, secondly, a set of persistent AF recordings (i. e., requiring external intervention for termination) were studied to predict ECV outcome. The recordings belonging to each data set were sampled at different rates, as will be described in the next subsections.

2.1 Paroxysmal AF Database

Fifty Holter recordings of 30 seconds in length and two leads (II and V1) available in Physionet [14] were analyzed. The database included non-terminating AF episodes (group N), which were observed to continue in AF for, at least, one hour following the end of the excerpt, and AF episodes terminating immediately after the end of the extracted segment (group T). These signals were digitized at a sampling rate of 128 Hz and 16-bit resolution. Nevertheless, they were upsampled to 1024 Hz in order to allow better alignment for QRST complex subtraction, such as Bollmann et al suggested [13]. This step is necessary to extract the atrial activity (AA) from surface ECGs, see section 3.1. A cubic splines interpolation method was used because it provided the best resolution (lower than 1 ms) in the R peak detection in comparison with other methods analyzed for the same purpose [15]. After the AA extraction, the residual signal was downsampled back to 128 Hz.

2.2 Persistent AF Database

Sixty-three patients (20 men and 43 women, mean age 73.35 ± 9.02 years) with persistent AF lasting more than 30 days, undergoing ECV were followed during four weeks. A standard 12-lead ECG was acquired for each patient during the whole procedure and a segment of 30 seconds in length was extracted from each recording for the analysis. All the signals were digitized at a sampling rate of 1024 Hz and 16-bit resolution.

After the ECV, 22 patients (34.93%) maintained normal sinus rhythm (NSR) during the first month. On the contrary, in 31 patients (49.20%), NSR duration was below one month and, in the remaining 10 (15.87%), AF recurred immediately after ECV. These 41 patients constituted the group of AF recurrence. All the patients were in drug treatment with amiodarone. The median arrhythmia duration was 10.58 months (range 1–47.22), echocardiography demonstrated a mean left atrium diameter of 45.82 ± 6.93 mm and 20.63% of the patients presented underlying heart disease. No significative differences were found in the aforementioned clinical parameters between the patients who maintained NSR and relapsed to AF.

2.3 Data Preprocessing

In both databases, lead V_1 was chosen for the analysis because previous works have shown that AA is prevalent in this lead [16]. The recordings were preprocessed in order to improve later analysis. Firstly, baseline wander was removed making use of bidirectional high pass filtering with 0.5 Hz cut-off frequency [17]. Secondly, high frequency noise was reduced with an eight order bidirectional IIR Chebyshev low pass filtering, whose cut-off frequency was 70 Hz [18]. Finally, powerline interference was removed through adaptive notch filtering, which preserves the ECG spectral information [19].

3 Methods

3.1 Strategy to estimate AF organization

The proper application of SampEn to the surface ECG in AF requires the fulfillment of several steps. Firstly, the ventricular activity has to be removed making use of a cancellation technique. Next, the main atrial wave (MAW) has to be extracted from the AA and, finally, SampEn computation can be applied to this wave.

Regarding the first step, AA analysis from surface ECG recordings is complicated by the simultaneous presence of ventricular activity, which is of much greater amplitude. Whereby, the AA signal has to be firstly extracted. Although a variety of different techniques exist for this purpose, a QRST cancellation method was used, since only one lead was considered in the study. Thus, the highest variance eigenvector of all the ECG beats was considered as the ventricular template for the cancellation. This QRST template was selected because it provided a higher quality AA extraction in short AF recordings, such as the analyzed in this work, than the obtained by averaging all the beats [20].

Nevertheless, since QRST morphology is affected by respiration, patient movement, etc., QRST residua and noise are often present in the extracted AA signal [16]. These nuisance signals degrade the AA organization estimation using non-linear regularity indexes, which provokes unsuccessful results [21]. To overcome this problem, the MAW has to be gained from the AA [21]. The MAW can be considered as the fundamental waveform associated to the AA, as Fig. 1 shows, its wavelength being the inverse of the AA dominant frequency [22].

In order to extract the MAW, the AA power spectral density (PSD) was firstly computed using the Welch Periodogram. A Hamming window of 4096 points in length, a 50% overlapping between adjacent windowed sections and a 8192-points Fast Fourier Transform (FFT) were used as computational parameters, as suggested by previous works [23]. The highest amplitude frequency within the 3–9 Hz range was selected as the dominant atrial frequency (DAF) [22], and the MAW was obtained by applying a selective filtering to the AA centered on the DAF. A linear phase FIR filter was used to prevent distortion [24]. The filter was designed by the Chebyshev approximation, with 3 Hz bandwidth and 768 coefficients, as detailed in previous works [3,25]. Finally, SampEn was computed over

the MAW in order to estimate AF organization.

3.2 Sample entropy definition

SampEn examines a time series for similar epochs and assigns a non-negative number to the sequence, with larger values corresponding to more irregularity in the data [2]. This number is the negative natural logarithm of the conditional probability that two sequences similar for m points remain similar for $m + 1$ points, within a tolerance r , excluding self-matches. Thus, for a time series of N points, $\{x(n), n = 1, \dots, N\}$, the $k = 1, \dots, N - m + 1$ vectors of length m are formed as $X_m(k) = \{x(k + i), i = 0, \dots, m - 1\}$. The distances among vectors are calculated as the maximum absolute distance between their corresponding scalar elements. The vectors number of which distance with the vector i is below a tolerance r is counted as B_i . The counting number of different vectors is calculated and normalized as

$$B^m(r) = \frac{1}{N - m} \sum_{i=1}^{N-m} \frac{B_i}{N - m - 1}. \quad (1)$$

Repeating the process for vectors of length $m + 1$, $B^{m+1}(r)$ can be obtained and SampEn can be defined as

$$SampEn(m, r, N) = -\ln \left[\frac{B^{m+1}(r)}{B^m(r)} \right]. \quad (2)$$

A more detailed description of SampEn, from a mathematical point of view, can be found in several previous works [2, 26, 27].

3.3 Selection of N and f_s

Previous works, in which heart rate and hormonal data regularity were analyzed [1, 9, 11, 28], have shown that for $m = 2$, values of r from 0.1 to 0.25 times the SD of the data and values of N between 100 and 5000 samples produce good statistical validity of ApEn. Additionally, Pincus and Goldberger [1] suggested that N should be at least 10^m and, preferably, at least 30^m when heart rate regularity is analyzed. However, this recommendation is not applicable to the case studied in this work, because the MAW length depends on the ECG sampling rate. To this respect, the selection of

a specific number of samples N will stretch out more or less MAW information as a function of the sampling rate. Hence, for high f_s , a low MAW time interval would be analyzed and the opposite also holds. As a consequence, the selection of a sampling rate, as the one defined by the Nyquist criterion, is not a guarantee of accurate organization estimation in AF.

Determining the optimal value of N has important practical relevance. When small N values are used to compute SampEn, the estimates can be inaccurate, because they can present a large variance. On the contrary, large N values may contain abrupt changes in amplitude that could result in inaccurate estimates due to normalization errors in the parameter r . However, it is noteworthy that N depends on f_s , since $N = T \cdot f_s$, where T denotes the time interval (in seconds) on which SampEn is computed. Thereby, in addition to the optimal value of N , it is important to determine the most appropriate sampling rate for the MAW.

To obtain the optimal values of these parameters, the MAW sampling rate was varied and, for each case, SampEn was computed with different N values. More precisely, the selected N values were 30, 60, 120, 240, 480, 960, 1920, \dots , L samples, where L is the closest series length lower than the analyzed MAW segment. Given that the minimum considered sampling rate was 64 Hz and that the ECG lengths were 30 seconds, their combination yields 1920 samples. As a consequence, all the selected N values took that number as reference, divided or multiplied by a power of two. For each N , the mean and SD of SampEn values corresponding to all the AF episodes were obtained. It should be noted that SampEn was computed with $m = 2$ and $r = 0.25$ times the SD of the analyzed segment. These values were chosen because they provided the best classification results in previous works dealing with AF organization estimation through SampEn [3–5, 25]. Additionally, the discriminative abilities for both predictions of paroxysmal AF termination and ECV result were calculated making use of the receiver operating characteristic (ROC) curves. The ROC curve is a graphical representation of the trade-offs between sensitivity and specificity. Sensitivity was considered as the proportion of non-terminating paroxysmal AF episodes correctly discerned, whereas specificity represented the percentage of terminating episodes properly identified for the spontaneous AF termination prediction. Similarly, for the prediction of ECV outcome, sensitivity was the proportion of patients relapsing to AF appropriately classified and specificity was the percentage of patients resulting in NSR after ECV

accurately predicted. The total number of paroxysmal AF patients and ECVs precisely classified was considered as the diagnostic accuracy corresponding to each prediction. As the number of episodes included into each database was not notably long, to evaluate the statistical robustness of the accuracy obtained for each prediction, a leave-one-out cross-validation scheme was used. Finally, significant differences between terminating and non-terminating AF episodes and between patients who resulted in NSR and relapsed to AF were evaluated making use of Student's t -test. All the groups had a normal and homoscedastic distribution as the Shapiro–Wilks and Levene tests proved, respectively. A two-tailed value of $p < 0.01$ was considered statistically significant.

The MAW was downsampled or upsampled to 64, 128, 256, 512, 1024 and 2048 Hz. Since the original sampling rates were 128 Hz for paroxysmal AF and 1024 Hz for persistent AF, when interpolation was necessary, a method based on cubic splines was applied, as previously commented.

The results obtained through the described tests, which will be presented in section 4.1, showed that several combinations of N and f_s provided good discrimination between terminating and non-terminating AF episodes and between patients who resulted in NSR and relapsed to AF after ECV. Thereby, for several N and f_s values, within the 30 seconds-length segment selected for each analyzed patient, the overlapping effect between adjacent N -length windows was tested on SampEn computation. Overlapping factors of 20%, 40%, 60% and 80% between adjacent windows were studied.

3.4 Selection of m and r

Once the optimal combinations of N and f_s were obtained, the most adequate selection of m and r was investigated. The accuracy and confidence of SampEn estimate improves as the number of length m matches increases. The number of matches can be increased by choosing small m (short templates) and large r (wide tolerance). However, penalties appear when too relaxed criteria are used [28]. For smaller r values, poor conditional probability estimates are achieved, while for larger r values, too much detailed system information is lost and SampEn tends to 0 for all the processes. To avoid a significant noise contribution on SampEn computation, one must choose r larger than

most of the noise [28]. Overall, to get optimal m and r values, an approach similar to the developed by Lake et al [29, 30] was used. SampEn was computed using a 10×20 matrix of combinations of $m = 1, 2, \dots, 10$ and $r = 0.05, 0.1, 0.15, \dots, 1$ times the SD of the analyzed segment. Normalizing r to the SD of the analyzed segment provides a translation and scale invariance, in the sense that SampEn remains unchanged under uniform process magnification, reduction, or constant shift to higher or lower values [9]. For each combination, the SampEn mean and SD values corresponding to all the AF episodes were obtained. Sensitivity, specificity, accuracy and statistical significance were also computed for each prediction. Moreover, a leave-one-out cross-validation approach was applied to assess statistically the robustness of the diagnostic accuracy obtained for each combination of m and r .

4 Results

4.1 Selection of N and f_s

Tables 1 and 2 show the mean and SD of SampEn values obtained for terminating and non-terminating AF episodes when N and f_s were selected as described in section 3.3. Similarly, Tables 3 and 4 present the corresponding information for persistent AF patients resulting in NSR and relapsing to AF after ECV. As can be appreciated for a given N , when f_s increases, SampEn decreases. On the other hand, for a given f_s of 256 Hz and above, Tables 1 through 4 show that SampEn increases when N also increases up to a reduced variation point. Thus, for higher N , very similar SampEn values were obtained. In fact, only variations lower than $\pm 5\%$ were observed. On the contrary, for f_s of 128 Hz and below, variable behavior was noticed for each Table and no clear increasing tendency was observed. In addition, the value of N from which the limited increasing behavior was noticed was dependent on f_s . Thus, for rates of 256, 512, 1024 and 2048 Hz, the reduced variation N values for SampEn were 240, 480, 960 and 1920, respectively.

The discriminative differences between subsets both for paroxysmal and persistent AF were mainly maintained for f_s of 256 Hz and above. Indeed, a high predictive accuracy was provided,

such as Fig. 2 shows. Precisely, for $f_s \geq 256$ Hz the lowest accuracy was 86% for paroxysmal AF and 73.02% for persistent AF, being 90% and 74.60% of cross-validated grouped cases correctly classified, respectively. In addition, statistical significant differences between patient groups were observed, given that a statistical significance lower than 0.01 was obtained for all the cases. On the contrary, for $f_s \leq 128$ Hz a limited predictive ability was obtained for the most part of combinations of N and f_s . Moreover, in most of the cases, significant differences between groups were not noticed ($p > 0.01$).

It is also noteworthy that for N equal or higher than the reduced variation values for SampEn, the diagnostic accuracy presented a constant value of 92% for paroxysmal AF and 79.37% for persistent AF, which was higher than the obtained with lower N values for a given f_s (see Fig. 2). In addition, a higher statistical significance ($p < 0.0001$) together with an identification accuracy of 96% and 82.54% of cross-validated grouped cases, respectively, were also provided.

These results show that several combinations of N and f_s can provide good classification for both AF databases. Thereby, the overlapping effect between adjacent N -length windows on SampEn computation was tested only for two different combinations. Precisely, the test was developed for f_s of 256 and 1024 Hz and N of 240 and 960 samples, respectively. The obtained results with overlapping factors of 20%, 40%, 60% and 80% provided very limited variations in SampEn, lower than $\pm 1\%$, both for paroxysmal and persistent AF episodes. In addition, classification improvement was unnoticed.

4.2 Selection of m and r

For the two pairs of N and f_s selected in the previous subsection, SampEn was computed using the 200 combinations of m and r described in section 3.4. Obtained results reported that SampEn decreased when m and r increased. However, SampEn variation with m and r was out of clinical interest, because useful medical information could not be retrieved. Thereby, only the outcomes related to the diagnostic accuracy achieved by each combination of m and r are presented next. Figs. 3 and 4 show the diagnostic accuracy color maps provided by the two tested combinations of

N and f_s for paroxysmal and persistent AF episodes, respectively. Both figures show a remarkable region in which a high accuracy for the two AF databases was reached. In addition, a very good statistical significance ($p < 0.00001$) was also obtained. As can be seen, the region was slightly larger for the pair $N = 960$ samples and $f_s = 1024$ Hz. This area can be approximately delimited, for both AF types, by $m = 1$ to $m = 5$ and r between 0.1 and 0.6 times the SD. Additionally, in all the cases, the optimal combination of m and r providing the highest diagnostic accuracy values was found within the indicated area. Thus, for paroxysmal AF, the maximum accuracy was 96%, with 96% of cross-validated grouped cases correctly identified, which was obtained with $m = 3$ and $r = 0.4$ ($p = 3.4567 \times 10^{-13}$) for $f_s = 256$ Hz and $m = 2$ and $r = 0.35$ ($p = 4.2134 \times 10^{-13}$) for $f_s = 1024$ Hz, respectively. Regarding persistent AF, maximum accuracy of 82.54%, with 87.30% of cross-validated grouped cases accurately identified, was achieved for $f_s = 256$ Hz, $m = 3$ and $r = 0.3$ ($p = 5.2483 \times 10^{-3}$) and, for $f_s = 1024$ Hz, with the pairs $m = 2$ and $r = 0.4$ ($p = 9.3892 \times 10^{-4}$) and $m = 3$ and $r = 0.25$ ($p = 8.7756 \times 10^{-4}$), respectively.

5 Discussion and Conclusions

The motivation of this study was based on the fact that no generalized guidelines exist for the selection of m , r and N when SampEn has to be applied to biomedical data. In this respect, only some values of m and r have been recommended and widely used for heart rate regularity analysis [1, 9, 10] and hormone secretion data [11]. On the other hand, Lewis et al [30] analyzed the optimal parameters when SampEn was applied to estimate the regularity of QT time series during rest and exercise. Additionally, similar studies were developed to characterize surface electromyographic signals [31] and the variability of respiratory patterns [32]. Anyway, a survey on the optimal parameters selection is recommended for the application of SampEn to not previously analyzed biomedical signals.

The results presented in Tables 1, 2, 3 and 4 showed that, for a given N , when f_s was double and higher than 128 Hz, SampEn was approximately reduced by a half. To elucidate a possible explanation of this fact, it is obvious that the higher the f_s , the lower the temporal distance between consecutive samples acquired from the original signal. As a consequence, the probability of finding

equal patterns (i.e. with a distance lower than r) will increase, because the signal variation from sample to sample will be reduced. In addition, for f_s of 256 Hz and above, the differences between each subset, both for paroxysmal and persistent AF, were maintained independently of N . On the other hand, for a specific f_s higher than 128 Hz, when N increased, SampEn also increased up to a reduced variation point. However, this behavior was not observed for $f_s = 128$ Hz and below. In this case, SampEn showed random variations with N and a reduced diagnostic accuracy was also noticed. Overall, a first recommendation is that the MAW sampling rate should be equal or higher than 256 Hz for an appropriate AF organization evaluation with SampEn.

On the other hand, considering the reduced variation point of N (SampEn variation $< \pm 5\%$) where a constant accuracy was reached for both AF types (92% for paroxysmal and 79.37% for persistent), a clear relationship between N and f_s can be elucidated for all of these situations. As the time period analyzed with SampEn, T , is the proportion between N and f_s , for all these cases we have that $T = \frac{240}{256} = \frac{480}{512} = \frac{960}{1024} = \frac{1920}{2048} = 0.9375$ s. As a consequence, any combination such that the analyzed time period is longer than one second would be appropriate to evaluate AF organization with SampEn. Nevertheless, considering a real time implementation of the proposed method, the most adequate combination of N and f_s would be 240 samples and 256 Hz, respectively, since the lowest computational burden would be reached.

Regarding overlapping between adjacent N -length windows, SampEn only presented slight variations with all the tested factors. Thereby, other finding that merits consideration for the application of SampEn is that no accuracy improvement was provided by overlapping. As a consequence, it can be avoided to reduce computational cost.

With regard to optimal values for m and r , developed experiments showed that a high diagnostic accuracy (92% for paroxysmal and 79.37% for persistent AF) can be reached through several combinations. Therefore, not only one, but several combinations of m and r can classify AF events with very good accuracy. However, the region outlined by these combinations was notably independent on the values of N and f_s and on the type of analyzed AF episodes, since slight differences in m and r were only noticed for all the cases. In addition, it is noteworthy that this region was considerably larger than the one typically recommended in the literature [9], because it was approximately delimit-

ited by $1 \leq m \leq 5$ and $0.1 \leq r \leq 0.6$. Nevertheless, these results prove that previous works used adequate values for the SampEn parameters in the estimation of AF organization [3–5]. However, a better diagnostic precision could be reached for each prediction (96% for paroxysmal and 82.54% for persistent AF) with a combination of m and r placed within the wider indicated region. An interesting observation is that optimal values reporting the highest diagnostic accuracy were dissimilar depending on the combination of N and f_s and on each type of AF. As a consequence, it can be suggested that the optimal combination of m and r has to be searched, within the proposed range, before the first application of SampEn to a non-previously analyzed database. This should only be made with very few recordings of that new database. In this way, the highest predictive ability will be reached. Nevertheless, as global conclusion of the developed tests, it could be suggested that the selection of m between 2 and 3 and r between 0.3 and 0.4 would probably yield optimal classification results in most of the situations related to AF.

As indicated in Section 1, a previous work has recently shown that the selection of a low r , between 0.1 and 0.2 times the SD, can lead to an incorrect regularity estimation with ApEn on fast dynamic signals [12]. To circumvent this limitation, Lu et al [12] proposed to choose the r that maximizes ApEn for each analyzed signal. In addition, the authors provided the equations needed to estimate automatically the recommended r for $m = 2$, $m = 3$ and $m = 4$. With the objective of analyzing if a better classification could be reached for both predictions, SampEn was computed making use of the r value provided by the proposed equations and $m = 2, 3$ and 4 . This experiment was developed for the two combinations of N and f_s used in previous tests. In both combinations, accuracy was 88% for spontaneous AF termination prediction and 76.19% for ECV result, independently of the used m value. In this case, cross-validation provided a correct classification of 86% and 76.19% of grouped cases, respectively. This observation leads to the conclusion that selecting the r that maximizes SampEn is inappropriate for AF organization evaluation, because other values of r will provide better classification outcomes.

A possible limitation of this study could be the use of paroxysmal and persistent AF recordings sampled at a fixed rate of 128 Hz and 1024 Hz, respectively, since upsampling involves interpolation and could provoke an overestimation of the analyzed time series regularity with SampEn. Initially,

it would be preferable to digitize the recordings directly at the different sampling rates analyzed in the study. Other alternative could be to use the highest studied sampling rate and only apply downsampling operations to both databases. In any case, the effect of interpolation was tested by comparing SampEn between downsampled and interpolated ECV signals for each sampling rate. Results yielded a relative difference in SampEn between both sets lower than 0.085% in the worst case. As a consequence, it can be considered that the interpolation operation does not alter the result of SampEn. Moreover, the coherent outcomes obtained for f_s values between 128 and 1024 Hz with persistent (original f_s of 1024 Hz) and paroxysmal (original f_s of 128 Hz) AF recordings suggest that signal interpolation did not provoke remarkable effects in the results. Finally, the sampling rate modification use to be a very limited operation, when available, in conventional ECG recording systems.

A final reflexion that has to be addressed is the kind of optimization achieved through the proposed methodology. In fact, the study presented herein optimizes SampEn parameters to maximize classification performance in different scenarios and databases of AF. However, the specific events which are classified are directly associated to measurable differences in AF organization, as discussed in [3–5, 21, 25]. Hence, it can be considered that the achieved results also maximize the estimation of AF organization via SampEn. In other words, the consequence of a method able to estimate robustly AF organization will be the ability to distinguish successfully between terminating and non-terminating paroxysmal AF patients and between successful and unsuccessful AF ECV patients.

To summarize, the present study has demonstrated that the selection of m , r and N plays a critical role in determining the outcome of SampEn when applied to physiological time series. Overall, this work has studied optimal parameters to apply SampEn as an AF organization estimator. The authors consider that both the methodology and the outcomes provided by this study could serve as a startup framework in the application of SampEn to other biomedical signals. In this sense, the fact that the range of values typically recommended and used in the literature may not include optimal values for a concrete biomedical application has to be remarked. On the other hand, the methodology itself, as presented in the manuscript, where the effects of m , r , N and f_s are jointly studied, can be taken as a reference for future studies dealing with the fine-tuning of SampEn in the context of other biomedical

signals.

Acknowledgements

This work was supported by the projects TEC2007–64884 and TEC2008–02241 from the Spanish Ministry of Science and Innovation, PII2C09-0224-5983 and PII1C09-0036-3237 from Junta de Comunidades de Castilla La Mancha, and SAN673/VA02/08 from Junta de Castilla y León.

References

- [1] S. M. Pincus, A. L. Goldberger, Physiological time-series analysis: what does regularity quantify?, *Am J Physiol* 266 (4 Pt 2) (1994) H1643–H1656.
- [2] J. S. Richman, J. R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy., *Am J Physiol Heart Circ Physiol* 278 (6) (2000) H2039–H2049.
- [3] R. Alcaraz, J. J. Rieta, F. Hornero, Atrial activity non-invasive characterization in previous instants before paroxysmal atrial fibrillation termination, *Rev Esp Cardiol* 61 (2) (2008) 154–160.
- [4] R. Alcaraz, J. J. Rieta, Wavelet bidomain sample entropy analysis to predict spontaneous termination of atrial fibrillation., *Physiol Meas* 29 (1) (2008) 65–80.
- [5] R. Alcaraz, J. J. Rieta, A non-invasive method to predict electrical cardioversion outcome of persistent atrial fibrillation., *Med Biol Eng Comput* 46 (7) (2008) 625–635.
- [6] V. Fuster, L. E. Rydén, D. S. Cannom, H. J. Crijns, A. B. Curtis, et al, ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation: A report of the American College of Cardiology/American Heart Association Task Force on practice guidelines and the European Society of Cardiology Committee for practice guidelines (writing committee to revise the 2001 guidelines for the management of patients with atrial fibrillation): developed in col-

- laboration with the european heart rhythm association and the heart rhythm society., *Circulation* 114 (7) (2006) e257–e354.
- [7] A. S. Go, E. M. Hylek, K. A. Phillips, Y. Chang, L. E. Henault, J. V. Selby, D. E. Singer, Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study., *JAMA* 285 (18) (2001) 2370–2375.
 - [8] M. W. Rich, Epidemiology of atrial fibrillation., *J Interv Card Electrophysiol* 25 (1) (2009) 3–8.
 - [9] S. M. Pincus, Assessing serial irregularity and its implications for health., *Ann N Y Acad Sci* 954 (2001) 245–267.
 - [10] D. T. Kaplan, M. I. Furman, S. M. Pincus, S. M. Ryan, L. A. Lipsitz, A. L. Goldberger, Aging and the complexity of cardiovascular dynamics., *Biophys J* 59 (4) (1991) 945–949.
 - [11] S. M. Pincus, D. L. Keefe, Quantification of hormone pulsatility via an approximate entropy algorithm., *Am J Physiol* 262 (5 Pt 1) (1992) E741–E754.
 - [12] S. Lu, X. Chen, J. K. Kanters, I. C. Solomon, K. H. Chon, Automatic selection of the threshold value r for approximate entropy., *IEEE Trans Biomed Eng* 55 (8) (2008) 1966–1972.
 - [13] A. Bollmann, D. Husser, L. Mainardi, F. Lombardi, P. Langley, A. Murray, J. J. Rieta, J. Millet, S. B. Olsson, M. Stridh, L. Sörnmo, Analysis of surface electrocardiograms in atrial fibrillation: Techniques, research, and clinical applications., *Europace* 8 (11) (2006) 911–926.
 - [14] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C. K. Peng, H. E. Stanley, Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals., *Circulation* 101 (23) (2000) E215–E220.
 - [15] I. Daskalov, I. Christov, Improvement of resolution in measurement of electrocardiogram RR intervals by interpolation., *Med Eng Phys* 19 (4) (1997) 375–379.

- [16] S. Petrutiu, J. Ng, G. M. Nijm, H. Al-Angari, S. Swiryn, A. V. Sahakian, Atrial fibrillation and waveform characterization. A time domain perspective in the surface ECG., *IEEE Eng Med Biol Mag* 25 (6) (2006) 24–30.
- [17] I. Dotsinsky, T. Stoyanov, Optimization of bi-directional digital filtering for drift suppression in electrocardiogram signals., *J Med Eng Technol* 28 (4) (2004) 178–180.
- [18] Y. Sun, K. Chan, S. M. Krishnan, ECG signal conditioning by morphological filtering., *Comput Biol Med* 32 (6) (2002) 465–479.
- [19] S. M. M. Martens, M. Mischi, S. G. Oei, J. W. M. Bergmans, An improved adaptive power line interference canceller for electrocardiography., *IEEE Trans Biomed Eng* 53 (11) (2006) 2220–2231.
- [20] R. Alcaraz, J. J. Rieta, Adaptive singular value cancelation of ventricular activity in single-lead atrial fibrillation electrocardiograms., *Physiol Meas* 29 (12) (2008) 1351–1369.
- [21] R. Alcaraz, J. J. Rieta, Sample entropy of the main atrial wave predicts spontaneous termination of paroxysmal atrial fibrillation, *Med Eng Phys* 31 (2009) 917–922.
- [22] M. Holm, S. Pehrson, M. Ingemansson, L. Sörnmo, R. Johansson, L. Sandhall, M. Sunemark, B. Smideberg, C. Olsson, S. B. Olsson, Non-invasive assessment of the atrial cycle length during atrial fibrillation in man: Introducing, validating and illustrating a new ECG method., *Cardio-vasc Res* 38 (1) (1998) 69–81.
- [23] J. J. Rieta, F. Castells, C. Sánchez, V. Zarzoso, J. Millet, Atrial activity extraction for atrial fibrillation analysis using blind source separation., *IEEE Trans Biomed Eng* 51 (7) (2004) 1176–1186.
- [24] L. Sörnmo, P. Laguna, *Bioelectrical Signal Processing in Cardiac and Neurological Applications*, Elsevier Academic Press, 2005.

- [25] R. Alcaraz, J. J. Rieta, Non-invasive organization variation assessment in the onset and termination of paroxysmal atrial fibrillation., *Comput Methods Programs Biomed* 93 (2) (2009) 148–154.
- [26] S. Ramdani, F. Bouchara, J. Lagarde, Influence of noise on the sample entropy algorithm., *Chaos* 19 (1) (2009) 013123.
- [27] R. Govindan, J. Wilsona, H. Eswaranb, C. Loweryb, H. Preialb, Revisiting sample entropy analysis, *Physica A: Statistical Mechanics and its Applications* 376 (2007) 158–164.
- [28] S. M. Pincus, Approximate entropy as a measure of system complexity., *Proc Natl Acad Sci U S A* 88 (6) (1991) 2297–2301.
- [29] D. E. Lake, J. S. Richman, M. P. Griffin, J. R. Moorman, Sample entropy analysis of neonatal heart rate variability., *Am J Physiol Regul Integr Comp Physiol* 283 (3) (2002) R789–R797.
- [30] M. J. Lewis, A. L. Short, Sample entropy of electrocardiographic RR and QT time-series data during rest and exercise., *Physiol Meas* 28 (6) (2007) 731–744.
- [31] W. Chen, Z. Wang, H. Xie, W. Yu, Characterization of surface EMG signal based on fuzzy entropy., *IEEE Trans Neural Syst Rehabil Eng* 15 (2) (2007) 266–272.
- [32] P. Casaseca, M. Martín, C. Alberola, Weaning from mechanical ventilation: a retrospective analysis leading to a multimodal perspective., *IEEE Trans Biomed Eng* 53 (7) (2006) 1330–1345.

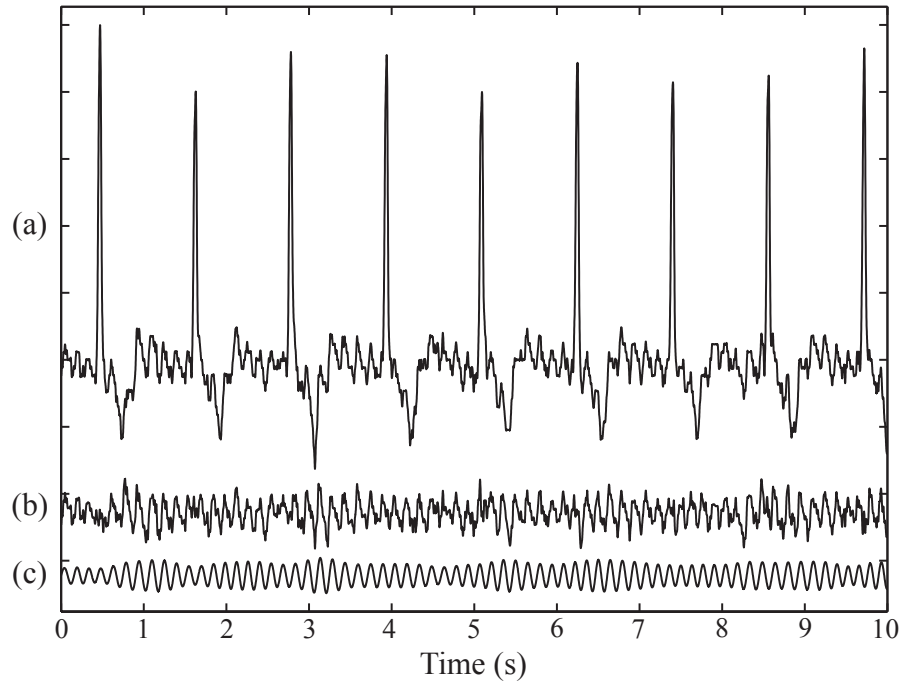


Figure 1: (a) Example of a 10 seconds-length ECG segment with paroxysmal AF. (b) Extracted AA signal. (c) MAW derived from the AA signal whose organization was estimated via SampEn.

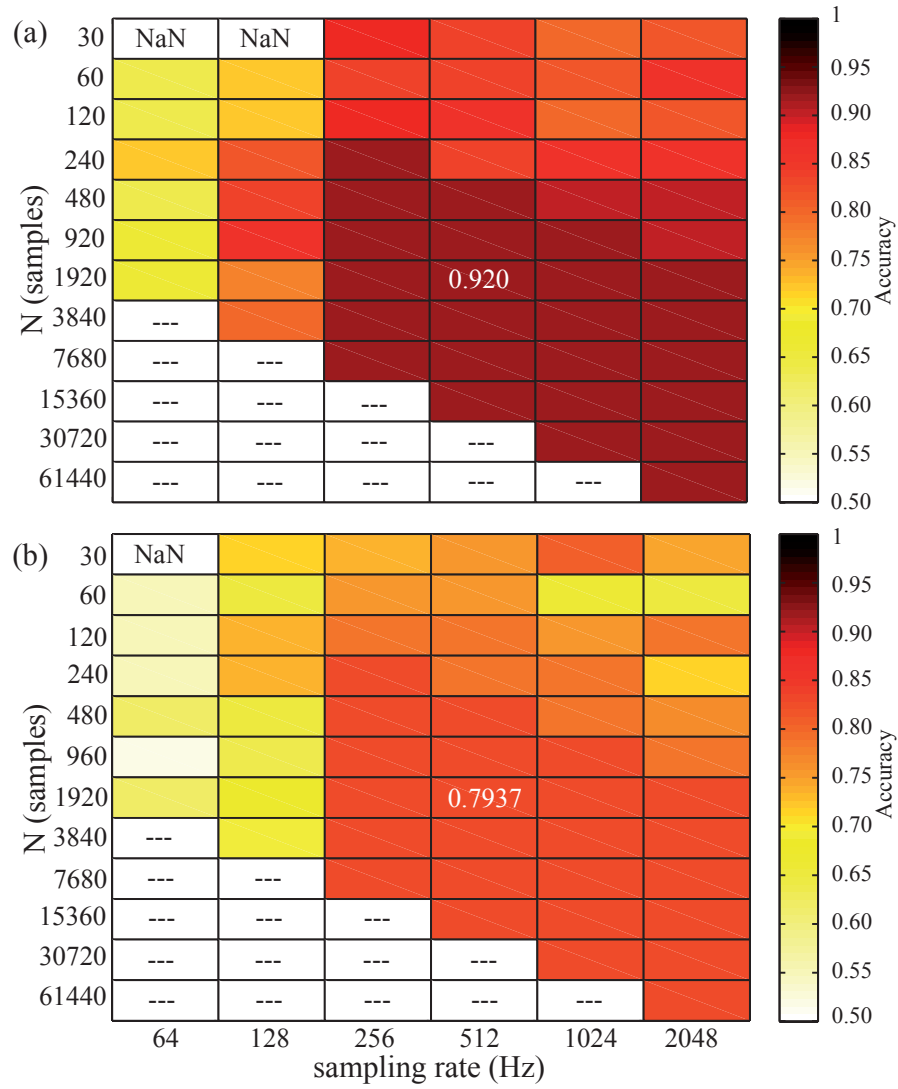


Figure 2: Diagnostic accuracy color maps obtained with different combinations of N and f_s for (a) paroxysmal AF termination and (b) cardioversion of persistent AF. NaN in some cells stands for Not a Number because SampEn did not report similarity in such a short time series.

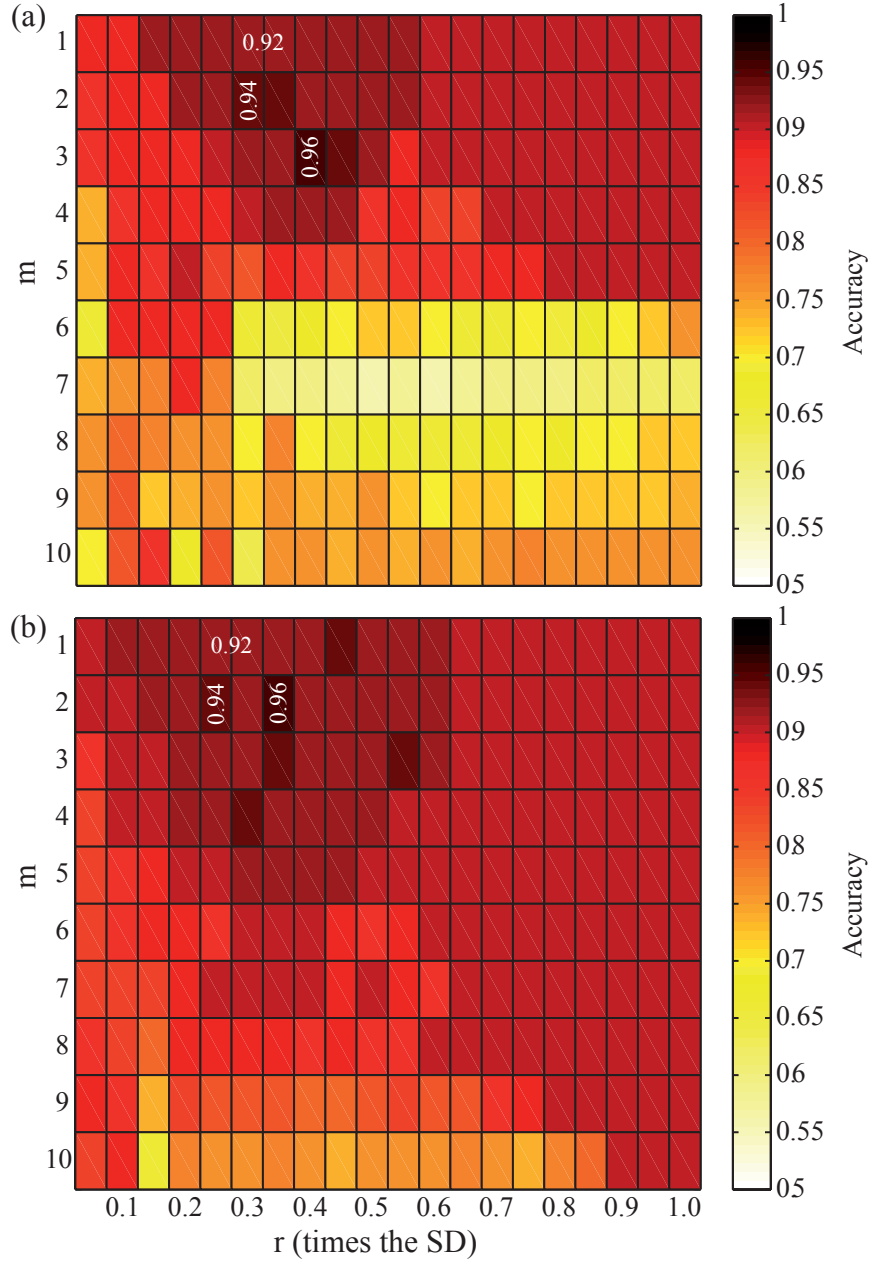


Figure 3: Diagnostic accuracy color maps for paroxysmal AF termination prediction obtained as a function of $m = 1, 2, \dots, 10$ and $r = 0.05, 0.1, 0.15, \dots, 1$ times the SD of the data, being N and f_s (a) 240 samples—256 Hz and (b) 960 samples—1024 Hz, respectively.

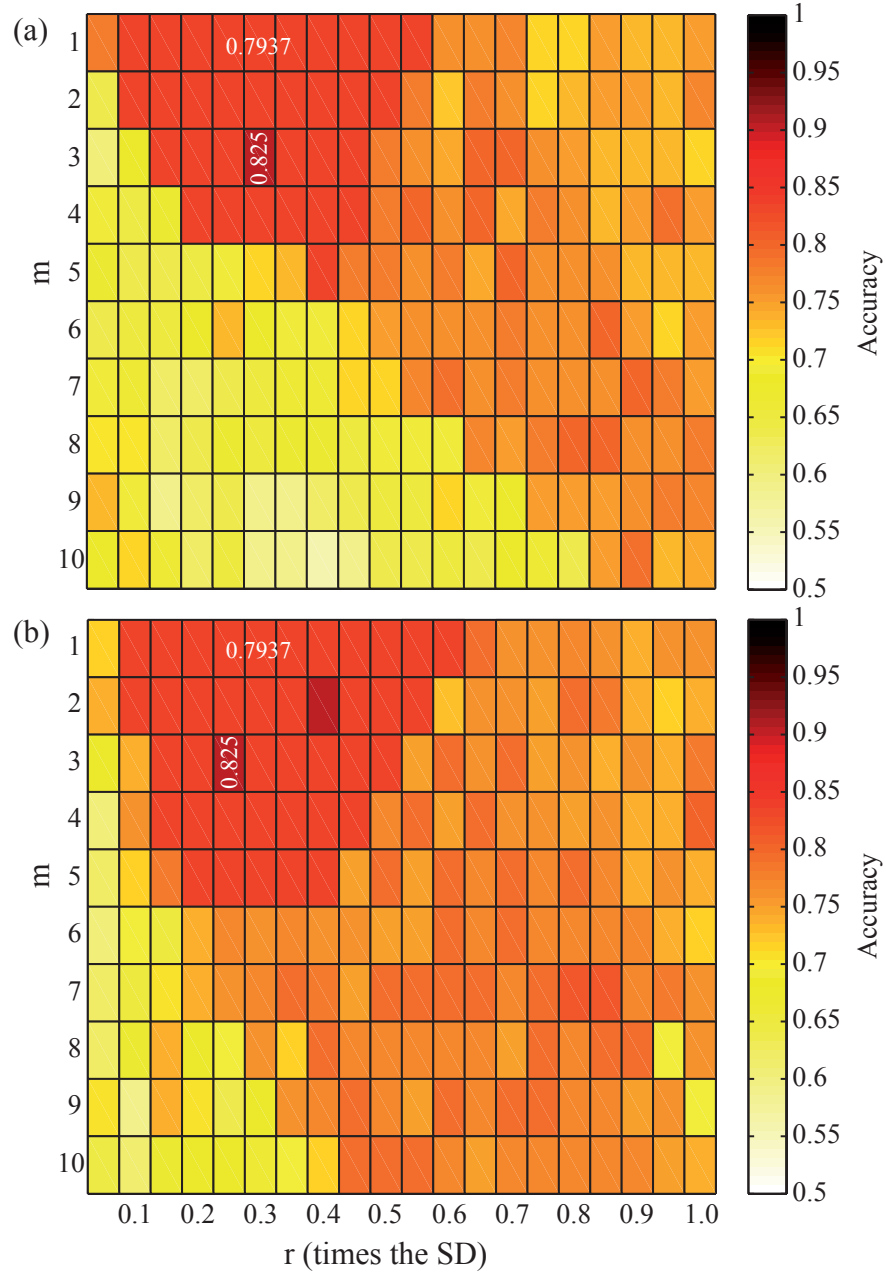


Figure 4: Diagnostic accuracy color maps for ECV result prediction obtained as a function of $m = 1, 2, \dots, 10$ and $r = 0.05, 0.1, 0.15, \dots, 1$ times the SD of the data, being N and f_s (a) 240 samples—256 Hz and (b) 960 samples—1024 Hz, respectively.

Table 1: Mean and SD of SampEn values obtained with different combinations of N and f_s for terminating paroxysmal AF episodes. The N value from which limited variations of SampEn ($< \pm 5\%$) are observed is highlighted for f_s of 256 Hz and above. NaN in some cells stands for Not a Number because SampEn did not report similarity in such a short time series.

N (samples) / f_s (Hz)	64	128	256	512	1024	2048
30	NaN	NaN	0.2022 ± 0.0245	0.1092 ± 0.0124	0.0590 ± 0.0066	0.0349 ± 0.0031
60	0.6056 ± 0.0323	0.4826 ± 0.0438	0.2725 ± 0.0270	0.1191 ± 0.0204	0.0556 ± 0.0070	0.0305 ± 0.0031
120	0.5547 ± 0.0246	0.4702 ± 0.0346	0.3102 ± 0.0334	0.1339 ± 0.0247	0.0534 ± 0.0082	0.0266 ± 0.0030
240	0.5487 ± 0.0269	0.4832 ± 0.0350	0.3352 ± 0.0389	0.1476 ± 0.0268	0.0589 ± 0.0089	0.0256 ± 0.0035
480	0.5461 ± 0.0275	0.4928 ± 0.0463	0.3454 ± 0.0459	0.1570 ± 0.0280	0.0650 ± 0.0105	0.0280 ± 0.0045
960	0.5430 ± 0.0287	0.5001 ± 0.0513	0.3506 ± 0.0569	0.1613 ± 0.0302	0.0691 ± 0.0111	0.0309 ± 0.0051
1920	0.5437 ± 0.0286	0.5012 ± 0.0576	0.3528 ± 0.0624	0.1633 ± 0.0330	0.0712 ± 0.0119	0.0328 ± 0.0054
3840	—	0.5005 ± 0.0656	0.3512 ± 0.0676	0.1628 ± 0.0343	0.0719 ± 0.0126	0.0337 ± 0.0056
7680	—	—	0.3484 ± 0.0766	0.1623 ± 0.0370	0.0724 ± 0.0137	0.0341 ± 0.0061
15360	—	—	—	0.1613 ± 0.0393	0.0720 ± 0.0148	0.0338 ± 0.0059
30720	—	—	—	—	0.0719 ± 0.0129	0.0336 ± 0.0066
61440	—	—	—	—	—	0.0340 ± 0.0059

Table 2: Mean and SD of SampEn values obtained with different combinations of N and f_s for non-terminating paroxysmal AF episodes. The N value from which limited variations of SampEn ($< \pm 5\%$) are observed is highlighted for f_s of 256 Hz and above. NaN in some cells stands for Not a Number because SampEn did not report similarity in such a short time series.

N (samples) / f_s (Hz)	64	128	256	512	1024	2048
30	NaN	NaN	0.2686 ± 0.0289	0.1412 ± 0.0129	0.0759 ± 0.0062	0.0434 ± 0.0033
60	0.5903 ± 0.0300	0.5591 ± 0.0365	0.3282 ± 0.0221	0.1787 ± 0.0262	0.0746 ± 0.0079	0.0390 ± 0.0033
120	0.5444 ± 0.0244	0.5463 ± 0.0231	0.3839 ± 0.0233	0.2020 ± 0.0289	0.0770 ± 0.0102	0.0350 ± 0.0034
240	0.5313 ± 0.0225	0.5502 ± 0.0158	0.4221 ± 0.0260	0.2231 ± 0.0302	0.0858 ± 0.0114	0.0358 ± 0.0043
480	0.5308 ± 0.0218	0.5574 ± 0.0149	0.4219 ± 0.0292	0.2355 ± 0.0314	0.0945 ± 0.0120	0.0400 ± 0.0050
960	0.5307 ± 0.0205	0.5649 ± 0.0126	0.4326 ± 0.0306	0.2414 ± 0.0319	0.1000 ± 0.0125	0.0440 ± 0.0053
1920	0.5310 ± 0.0204	0.5681 ± 0.0120	0.4387 ± 0.0327	0.2433 ± 0.0316	0.1030 ± 0.0129	0.0464 ± 0.0054
3840	—	0.5698 ± 0.0119	0.4395 ± 0.0346	0.2434 ± 0.0311	0.1040 ± 0.0128	0.0475 ± 0.0055
7680	—	—	0.4419 ± 0.0345	0.2425 ± 0.0325	0.1050 ± 0.0132	0.0482 ± 0.0056
15360	—	—	—	0.2424 ± 0.0335	0.1046 ± 0.0131	0.0478 ± 0.0064
30720	—	—	—	—	0.1051 ± 0.0124	0.0480 ± 0.0059
61440	—	—	—	—	—	0.0476 ± 0.0063

Table 3: Mean and SD of SampEn values obtained with different combinations of N and f_s for patients resulting in NSR after ECV. The N value from which limited variations of SampEn ($< \pm 5\%$) are observed is highlighted for f_s of 256 Hz and above. NaN in some cells stands for Not a Number because SampEn did not report similarity in such a short time series.

N (samples) / f_s (Hz)	64	128	256	512	1024	2048
30	NaN	0.5381 ± 0.1427	0.2894 ± 0.0778	0.1339 ± 0.0379	0.0709 ± 0.0172	0.0414 ± 0.0090
60	0.6421 ± 0.1020	0.5109 ± 0.1001	0.3211 ± 0.0801	0.1480 ± 0.0472	0.0682 ± 0.0189	0.0371 ± 0.0099
120	0.6238 ± 0.0976	0.5235 ± 0.0989	0.3438 ± 0.0830	0.1599 ± 0.0501	0.0657 ± 0.0201	0.0327 ± 0.0086
240	0.6095 ± 0.0985	0.5404 ± 0.0976	0.3802 ± 0.0920	0.1751 ± 0.0523	0.0715 ± 0.0233	0.0310 ± 0.0095
480	0.6244 ± 0.0868	0.5432 ± 0.1101	0.3874 ± 0.0986	0.1854 ± 0.0585	0.0771 ± 0.0246	0.0341 ± 0.0103
960	0.6002 ± 0.0851	0.5526 ± 0.1139	0.3981 ± 0.0990	0.1872 ± 0.0601	0.0829 ± 0.0261	0.0359 ± 0.0111
1920	0.6351 ± 0.0888	0.5642 ± 0.1158	0.3862 ± 0.0965	0.1861 ± 0.0672	0.0851 ± 0.0278	0.0389 ± 0.0166
3840	—	0.5512 ± 0.1201	0.3899 ± 0.1018	0.1809 ± 0.0712	0.0844 ± 0.0296	0.0394 ± 0.0121
7680	—	—	0.3954 ± 0.1061	0.1795 ± 0.0689	0.0831 ± 0.0285	0.0389 ± 0.0127
15360	—	—	—	0.1834 ± 0.0707	0.0836 ± 0.0303	0.0401 ± 0.0132
30720	—	—	—	—	0.0861 ± 0.0310	0.0392 ± 0.0145
61440	—	—	—	—	—	0.0393 ± 0.0140

Table 4: Mean and SD of SampEn values obtained with different combinations of N and f_s for patients relapsing to AF after ECV. The N value from which limited variations of SampEn ($< \pm 5\%$) are observed is highlighted for f_s of 256 Hz and above. NaN in some cells stands for Not a Number because SampEn did not report similarity in such a short time series.

N (samples) / f_s (Hz)	64	128	256	512	1024	2048
30	NaN	0.6432 ± 0.0888	0.3094 ± 0.0462	0.1628 ± 0.0314	0.0856 ± 0.0192	0.0449 ± 0.0082
60	0.6911 ± 0.0936	0.5921 ± 0.0619	0.3685 ± 0.0385	0.1872 ± 0.0358	0.0822 ± 0.0169	0.0399 ± 0.0084
120	0.6436 ± 0.0869	0.5839 ± 0.0603	0.4186 ± 0.0451	0.2061 ± 0.0351	0.0839 ± 0.0185	0.0383 ± 0.0080
240	0.6327 ± 0.0681	0.5938 ± 0.0587	0.4511 ± 0.0486	0.2159 ± 0.0387	0.0913 ± 0.0196	0.0372 ± 0.0086
480	0.6354 ± 0.0672	0.6098 ± 0.0563	0.4681 ± 0.0499	0.2376 ± 0.0401	0.0982 ± 0.0201	0.0405 ± 0.0091
960	0.6219 ± 0.0648	0.6132 ± 0.0505	0.4654 ± 0.0541	0.2442 ± 0.0409	0.1046 ± 0.0209	0.0442 ± 0.0096
1920	0.6187 ± 0.0632	0.5991 ± 0.0566	0.4693 ± 0.0536	0.2473 ± 0.0412	0.1072 ± 0.0211	0.0491 ± 0.0101
3840	—	0.6315 ± 0.0438	0.4701 ± 0.0504	0.2461 ± 0.0399	0.1086 ± 0.0218	0.0498 ± 0.0108
7680	—	—	0.4709 ± 0.0529	0.2492 ± 0.0422	0.1072 ± 0.0221	0.0503 ± 0.0112
15360	—	—	—	0.2487 ± 0.0436	0.1081 ± 0.0215	0.0510 ± 0.0114
30720	—	—	—	—	0.1065 ± 0.0234	0.0506 ± 0.0120
61440	—	—	—	—	—	0.0499 ± 0.0118