DETECTION OF NON RANDOM PHASE SIGNAL IN ADDITIVE NOISE WITH SURROGATE ANALYSIS

Manouane Caza-Szoka and Daniel Massicotte

Université du Québec à Trois-Rivières, Department of Electrical and Computer Engineering,
3351, Boul. des Forges, Trois-Rivières, Québec, Canada
Laboratoire des Signaux et Systèmes Intégrés
{manouane.caza-szoka, daniel.massicotte}@uqtr.ca

ABSTRACT

The Surrogate Analysis (SA) is known to detect nonlinear signals, non-stationary signals and ARMA systems driven by non-Gaussian processes. This paper adds to address the detection of non-random phase signal, of which the linear phase signal is the best-known example. This is a new interpretation of the SA. In order to highlights the benefits of the interpretation, a new theoretical signals is constructed. The signal has a perfect Gaussian distribution and is not affected by periodic extension and is a linear phase signal. The SA will be shown able to detect this signal in a noise with exactly the same power spectrum. It will be clear that the SA is able to detect phase linearity even when the data is normally distributed. An application of the detection by SA is given regarding very noisy and short time electrocardiogram (ECG) signal and compared to higher order statistics and normality tests for this purpose.

Index Terms**—** Nonlinear Analysis, Hypothesis Testing, Detection, Bootstrap Method, Biomedical Signal, Fractal Dimension.

1. INTRODUCTION

The Surrogate Analysis (SA) [1] is a widely used method of nonlinear analysis applied to a large spectrum of domain. The first applications were found in physics to detect low order chaos [2]. Since a large literature has been developed for the SA in the study of bio signals, notably in brain imaging techniques such as EEG and MEG [3][4]. Also, financial time series have been explored through SA [5].

The SA is a statistical test that has for null hypothesis that the signal can be obtained from a linearly filtered stationary white Gaussian noise (ARMA process). Any departure from this hypothesis yields a positive result [6]. Hence, although the assessment low order chaos was first aim of the SA, other underlying phenomena have been targeted. The effect of the time varying aspect has been noticed in [1][7], more as a bias of the SA. It was, however, later developed as a useful tool to assess non-stationarity in [8]. The non-Gaussian driving process deviation aspect has been considered in the context of financial time series [9]. It seems that the whole spectrum of possible departure from the null hypothesis has been treated. From another point of view, let is considered specifically the Fourier based SA [1], which is the original and most commonly used version of the SA. This approach indirectly addresses the ARMA null hypothesis. More directly stated, the null hypothesis is that the signal is such as its phase in the Fourier Spectrum is random. From this hypothesis, a new possible departure from the null hypothesis is revealed. But there is a link between the two points of view. From the ARMA hypothesis, one could ask: "Can an ARMA system produce a signal in which there is a non-random Phase Spectrum?" A simple illustration of a non-randomness is the difference between a Dirac function and a white Gaussian noise: both have a constant power spectrum, but while the Dirac's phase spectrum is linear, or even constant, the noise's one is uniformly distributed. Obviously, it is usually not necessary to use nonlinear analysis to detect a Dirac signal.

This paper shows that the SA can detect some non-randomness in the phase spectrum of a signal. In the simplest form, this non-randomness will be a null or linear one. The demonstration is a new theoretical signal that has Gaussian distribution by construction, which is not affected by periodic extension and is a linear phase signal. A less theoretical simulation where a real electrocardiogram (ECG) signal is detected within a frame of approximately 2 seconds in high levels of noise will show that the method has potential practical applications. In that case, there is no question of linear phase signal but clearly, the ECG is not a random signal and its Phase Spectrum should neither be random. The Fractal Dimension (FD) will be used in the SA.

The detection of ECG signal is an important aspect in Fetal-ECG [10]. It is necessary for channel selection after application of Principal Component Analysis (PCA) and Independent Component Analysis (ICA). In [11], only the Kurtosis is used to perform the channel selection. This paper shows that the SA can be more sensitive for detecting the ECG component. Also, the proposed method could be used also in a preprocessing step prior to the PCA and ICA by selecting the "most interesting" channels. This is important since the position of the sensors of Fetal-ECG systems are not necessarily placed at constant locations in relation to the fetus and the number of sensors are generally high. Let's note briefly that the ECG detection could in a similar way be useful in wearable ECG devices [12], especially when a good contact of the electrodes with the skin is not guaranteed. It must be emphasized that, the detection problem has never been tackled by the use of SA, e.g. [13][14]. The SA was only used to assess the signal characteristic in order to validate a model or the validity of a nonlinear feature, with the exception of [15] where a SA scores were used as a signal feature for a classification problem.

The paper is organized as follows: Section 2 presents the SA and FD algorithms. Section 3 details the test signals and experimentations and performance results are discussed in Section 4. Finally, the conclusions are drawn in Section 5.

2. ALGORITHMS

The Surrogate Analysis (SA) and the Fractal Dimension (FD) methods are detailed in this section.

2.1. Surrogate Analysis

The SA is presented in its most standard version as in the Theiler's paper [1] and is shown in Fig. 1, where x defined the original data signal. It consists of comparing a nonlinear feature of the original series to the same features calculated on surrogate series. The surrogate series are generated by phase randomization in the Fourier domain, more precisely by sequentially applying the Fast Fourier Transform (FFT) to the original data to transform the data in the frequency domain, randomize the phase with conjugate symmetry to keep the signal real, and using the Inverse FFT (IFFT) to obtain the time domain surrogate.

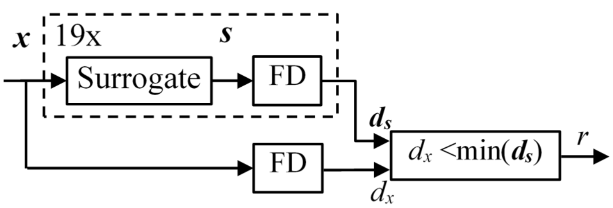


Fig. 1. Surrogate analysis for a left-sided unilateral test. Nineteen surrogate signals are generated. The FD of the original data $\boldsymbol{x}$ is compared to the minimum FD of the surrogates.

2.2. Fractal Dimensions

Two methods for calculating the FD are considered, namely the Higuchi's [16] and Katz's [17] methods. These come from slightly different theory and yield different results, especially when series are not pure fractals, which is almost always the case. Also, the methods have parameters which can highly modify their behavior [18][19]. It is desirable that the two methods show different aspects of the signals.

The Higuchi's method consists of the slope L Spectrum, i.e. the logarithm of the signal absolute length with respect to the logarithm of the subsampling factor k:

$$L_m(k) = \sum_{i=1}^{\frac{N}{k}-1} \left| x(ik+m) - x\big((i-1)\cdot k + m\big)\right| / k \qquad (1)$$

This spectrum is averaged for m, the starting point:

$$L(k) = \sum_{m=1}^{k} L_m(k)\,/k \qquad (2)$$

It is possible to only take some subsampling factor k. When the maximum of k is high, the values are generally logarithmically spaced. On the other side, the Katz's FD, D_K, is obtained by:

$$D_K = \log(L)/\log(d) \qquad (3)$$

where L is the total length of the time series:

$$L = \sum_n sqrt((u_{n+1} - u_n)^2 + \alpha^2) \qquad (4)$$

and d is the maximum distance between any two points.

$$d = \sqrt{\max_n\big(((u_n - u_1)^2 + \alpha^2(n-1)^2)\big)} \qquad (5)$$

The signal at time step n is denoted by u_n. The parameter α is the homogeneity factor between the time and signal units. The method was made for waveforms, i.e. for multidimensional data where the dimensions are homogeneous. For time series, an appropriate ratio between the time step and signal units must be chosen [19]. However, by choosing this ratio high enough, it can be shown that the Katz's FD will correspond to a regularized version of the offset of the Higuchi's method L Spectrum, likely giving different information. Finally, to ensure that the Katz's method is amplitude scale invariant, a normalization of the signal is done before calculation of the FDs. Let's note that Higuchi's method is intrinsically amplitude scale invariant.

3. TEST SIGNALS AND NUMERICAL EXPERIMENTATION

Two experiments are proposed to highlights the detection capacity of the SA, one purely numerical and the other semi-empirical.

3.1. Perfectly Gaussian Impulse in Colored Noise

The goal of the first experiment is to detect a signal that has a perfectly Gaussian distribution and which is in noise which has exactly the same power spectrum. The signal is composed built around the Inverse Cumulative Distribution Function (ICDF) of the Gaussian distribution. If the ICDF was to be used as the final signal, it would be extremely easy to detect by the SA because of the discontinuity induced by the periodic extension of the FFT. Therefore, by using twice the ICDF in reverse order, the periodic extension discontinuity can be completely removed. To make the signal more similar to other common pulse, it is rearranged such that it starts and finished at zero. This is simply equivalent to applying a circular delay. It is noteworthy that the peak values are determined by the number of data points used in the ICDF. The additive noise is constructed to have the same power spectrum as the signal itself by using the same process as for the generation of the surrogate series. Such approach to create a colored noise was not found in the literature. The signal and an example of the noisy signal are shown in Fig. 2 (a) and (b).

The detection is considered unsupervised, meaning that no information about the signal will be used, apart the fact that it is not stochastic. The question asked by applying the detection algorithms is "Is there something that is non-stochastic?". Although the SA has never been used for a

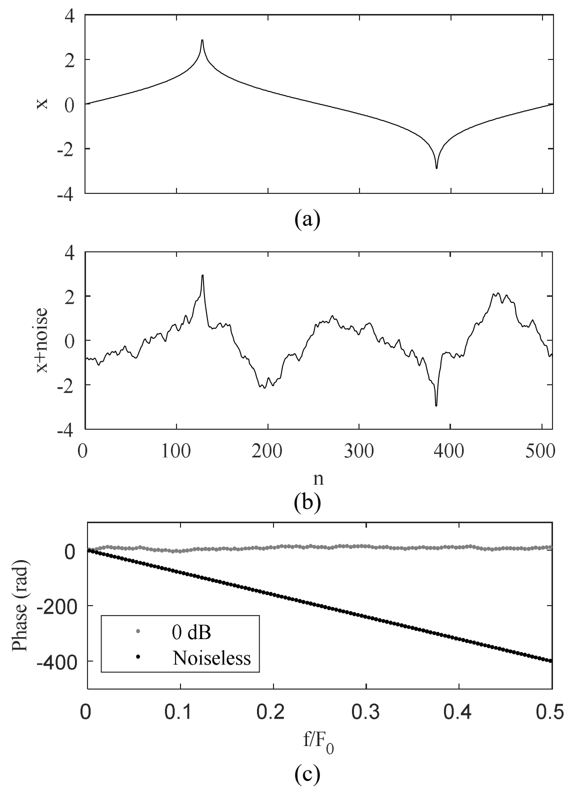


Fig. 2. Proposed signal with perfect normal distribution (a) and noisy version (b) at an SNR of 0 dB. The signal is created by the inverse cumulative function of the Normal distribution of 256 points, used twice, for a total of 512 points. The signal starts and finishes at 0. The added noise is done by phase randomization of the signal in the Fourier domain. It has the same power spectrum as the original signal. The unwrapped version of the phase spectrum is shown in (c) for the noiseless and noisy signal.

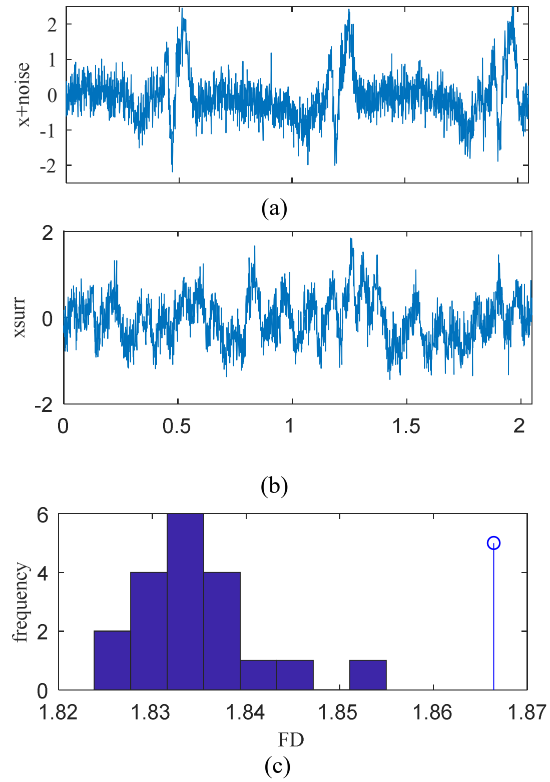


Fig. 3. Example of the different signals involved in the ECG detection experiment for a 2.048 second window and an SNR of 8 dB. In (a), the noisy ECG is shown. It is on the windowed version that the phase randomization is performed. An example of surrogate signal is shown in (b). It is on the windowed version that the FD is calculated. The distribution of the 19 surrogate series FDs is presented for Higuchi's method in (c) along with the FD of the noisy series located by the stem at FD = 1.866.

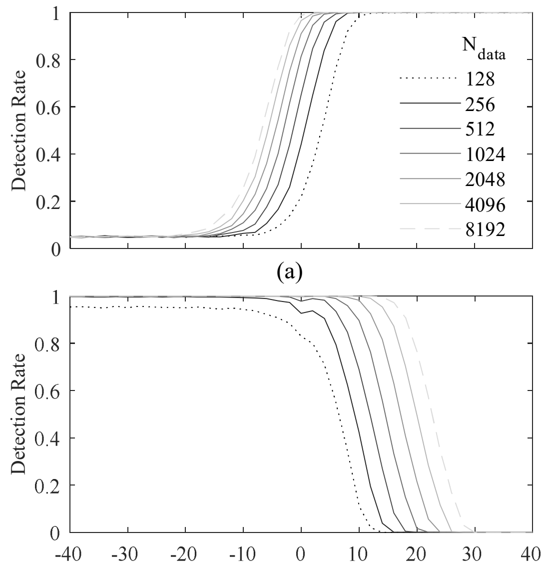
detection problem, SA is made to answer that last question. It will be compared with Kolmogorov-Smirnov (KS) [20] and Shapiro-Wilk (SW) [21] normality test which represents a good overview of the existing methods [22]. The KS and SW tests used $\alpha = 0.05$. The effect of both the signal to noise ratio (SNR) and the number of data points are observed. The Monte-Carlo simulation will have 10 000 trials per point.

3.2. Detection of ECG in Noise

To show the potential of the SA to be used for real-world applications, the detection of ECG within noise was done. A real ECG signal, from Physionet Databank [23] (first subject, fourth channel), is taken as the known, noiseless signal. Obviously, this signal contains noise, but it is of much lower power than the noise that is added synthetically. The additive noise is a random Gaussian white noise. The particular channel was chosen so that simpler detection methods such as thresholding would perform poorly. The time span of the windows was 2.048 seconds. It is a reasonably short period of time to detect presence of ECG for most thinkable applications. Although it is clear that any filtering method would increase the effective SNR drastically (easily over 7 dBs), the interaction between the filtering and nonlinear features is not well defined in the literature, and even less for SA. For this reason, the use of filtering is postponed for future works. In the case of windowing, it is shown in [24] that it can avoid spurious detections. Hence, a Tuckey window with a flat section ratio of 0.5 was used. Examples of the different signals involved in the ECG detection are shown in Fig. 3 (a) and (b). The windowed version of the signal is not shown. The Monte-Carlo simulation will have 30 000 trials per points, with windows selected randomly along the time series.

4. RESULTS AND DISCUSSION

The phase spectrum of the noiseless signal and the noisy version at 0 dB are presented in Fig. 2 (c). For the noiseless version, it is perfectly linear. With the noise, the non-randomness of the phase is unclear from the plot, but it was detected by the SA. Hence, the SA can detect non-random relation that is difficult to see from the phase spectrum. In Fig. 4, the detection rates for Higuchi's SA and SW test are shown. Quite surprisingly, in Fig. 4 (b) the SW test was giving positive results at low SNR. This means that even though the original signal is Gaussian, its Surrogate can become non-Gaussian. It is a fact, in our knowing, not found



(b)

Fig. 4. Detection Rates of the normally distributed pulse in surrogate noise. The Higuchi (a) based SA method and SW (b) normality tests are shown.

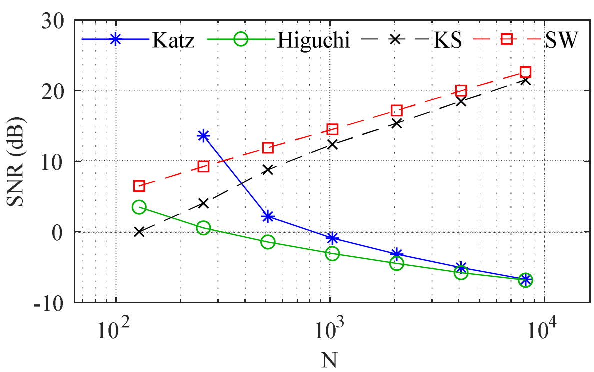


Fig. 5. SNR at 50% detection rate of the Normally Distributed Impulse according to the number of data points N for the four detection methods.

in the literature. Because of this reverse relation, the performances of the different algorithms were compared at a 50% detection rate in Fig. 5. While for the SA method, Katz and Higuchi, more data gives better results, it's the opposite for the KS and SW tests. Hence, the longer the data, the more its Surrogate has a Gaussian distribution.

Although both Katz's and Higuchi's are FD methods, they had fairly different performance in both experiments. This is not necessarily a negative point for Katz's method. It might indicate that the two methods detect uncorrelated aspects and in this case, it may be possible to use both methods together.

Because of the behaviour observed on the theoretical study, it was decided to adjust the method so that a 5% false positive rate would be obtained on pure noise for the ECG detection simulations. The noisy ECG and one surrogate example are given in Fig. 3 (a) and (b) while an example of Surrogate FD

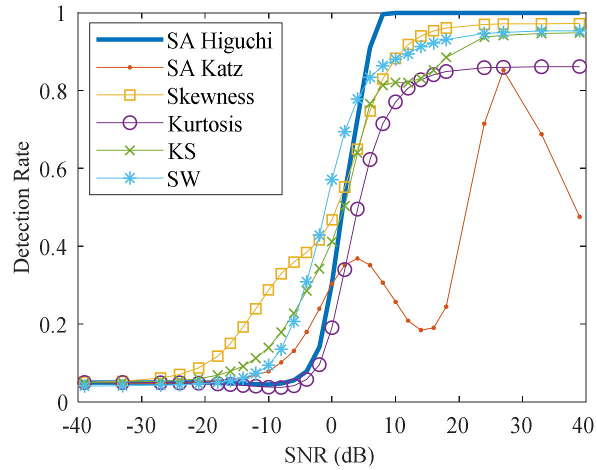


Fig. 6. Results for ECG detection in noise for SA, HOS and normality tests. The threshold were set so that there is a Detection Rate (misclassified) of 5%.

distribution is found in (c). The ECG detection results are shown in Fig. 6. No particular method was dominating the other. However, the Higuchi's SA had the sharpest transition and was the only one to obtain 100% success rate at high SNR. Again, it could be suggested to use multiple nonlinear features to obtain better results.

5. CONCLUSION

In this paper, a new interpretation of the SA, along with a signal with perfect normal distribution and linear phase was presented. This signal allowed to highlight the fact that SA is able to detect non ARMA signals even when the data is normally distributed. Moreover, the use of SA for the detection of ECG (or even any desired signal) in noise was shown for the first time. A theoretical simulation showed that the SA could outperform other nonlinear features that are based on the distribution. The interpretation of the theoretical results was that SA could detect phase linearity. To show the potential of the SA and its new interpretation on real-world applications, the detection of an ECG signal in noise was therefore tried by the method. It was shown that SA could again be used advantageously. Further research needs to be done to apply the filtering with SA and to combine the different nonlinear features in a single detector.

ACKNOWLEDGEMENTS

This work has been funded by the Natural Sciences and Engineering Research Council of Canada and the Research Chair in Signals and Intelligence of High Performance Systems.

REFERENCES

[1] J. Theiler, S. Eubank, A. Longtin, B. Galdrikian, and J. D. Farmer, "Testing for nonlinearity in time series: the method of surrogate data," Physica D: Nonlinear Phenomena, vol. 58, no. 1-4, pp. 77–94, 1992.

[2] E. Bradley and H. Kantz, “Nonlinear time-series analysis revisited,” Chaos: An Interdisciplinary Journal of Nonlinear Science, vol. 25, no. 9, p. 097610, 2015.

[3] C. Stam, “Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field,” Clinical Neurophysiology, vol. 116, no. 10, pp. 2266–2301, 2005.

[4] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, and C. E. Elger, “Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state,” Physical Review E, vol. 64, no. 6, 2001.

[5] S.-H. Chen, T. Lux, and M. Marchesi, “Testing for non-linear structure in an artificial financial market,” Journal of Economic Behavior & Organization, vol. 46, no. 3, pp. 327–342, 2001.

[6] J. Timmer, “What Can Be Inferred from Surrogate Data Testing?,” Physical Review Letters, vol. 85, no. 12, pp. 2647–2647, 2000.

[7] J. Theiler, “Spurious dimension from correlation algorithms applied to limited time-series data,” Physical Review A, vol. 34, no. 3, pp. 2427–2432, Jan. 1986.

[8] J. Timmer, “Power of surrogate data testing with respect to nonstationarity,” Physical Review E, vol. 58, no. 4, pp. 5153–5156, Jan. 1998.

[9] W. Zhou, "The components of empirical multifractality in financial returns", EPL (Europhysics Letters), vol. 88, no. 2, p. 28004, 2009.

[10] Martinek, R. Kahankova, J. Jezewski, R. Jaros, J. Mohylova, M. Fajkus, J. Nedoma, P. Janku and H. Nazeran, "Comparative Effectiveness of ICA and PCA in Extraction of Fetal ECG From Abdominal Signals: Toward Non-invasive Fetal Monitoring", Frontiers in Physiology, vol. 9, 2018.

[11] R. Sameni, C. Jutten and M. B. Shamsollahi, "What ICA provides for ECG processing: Application to noninvasive fetal ECG extraction," In Proc. of the International Symposium on Signal Processing and Information Technology (ISSPIT'06), Vancouver, Canada, pp. 656-661,. 2006.

[12]N. Lokare, L. Gonzalez and E. Lobaton, "Comparing wearable devices with wet and textile electrodes for activity recognition," 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016.

[13] R. Govindan, K. Narayanan and M. Gopinathan, "On the evidence of deterministic chaos in ECG: Surrogate and predictability analysis", Chaos: An Interdisciplinary Journal of Nonlinear Science, vol. 8, no. 2, pp. 495-502, 1998.

[14] P. Tass, M. Rosenblum, J. Weule, J. Kurths, A. Pikovsky, J. Volkmann, A. Schnitzler and H. Freund, "Detection of Phase Locking from Noisy Data: Application to Magnetoencephalography", Physical Review Letters, vol. 81, no. 15, pp. 3291-3294, 1998.

[15] M. Caza-Szoka, D. Massicotte, F. Nougarou, and M. Descarreaux, “Surrogate analysis of fractal dimensions from SEMG sensor array as a predictor of chronic low back pain,” 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2016.

[16] T. Higuchi, “Approach to an irregular time series on the basis of the fractal theory,” Physica D: Nonlinear Phenomena, vol. 31, no. 2, pp. 277–283, 1988.

[17] M. J. Katz, “Fractals and the analysis of waveforms,” *Computers in Biology and Medicine*, vol. 18, no. 3, pp. 145–156, 1988.

[18] R. Esteller, G. Vachtsevanos, J. Echauz, and B. Litt, “A comparison of waveform fractal dimension algorithms,” IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications, vol. 48, no. 2, pp. 177–183, 2001.

[19] P. Castiglioni, “What is wrong in Katz’s method? Comments on: ‘A note on fractal dimensions of biomedical waveforms,’” Computers in Biology and Medicine, vol. 40, no. 11-12, pp. 950–952, 2010.

[20] F. J. Massey, “The Kolmogorov-Smirnov Test for Goodness of Fit,” Journal of the American Statistical Association, vol. 46, no. 253, pp. 68–78, 1951.

[21] S. S. Shapiro and M. B. Wilk, “An analysis of variance test for normality (complete samples),” Biometrika, vol. 52, no. 3-4, pp. 591–611, Jan. 1965.

[22] N. M. Razali and Y. B Wah, “Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling Tests,” Journal of statistical modeling and analytics, vol. 2, no. 1, pp. 21-33, 2011.

[23] Bousseljot, R.; Kreiseler, D.; Schnabel, A. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. Biomedizinische Technik, Band 40, Ergänzungsband 1 (1995) S 317

[24] T. Suzuki, T. Ikeguchi, and M. Suzuki, “Effects of data windows on the methods of surrogate data,” Physical Review E, vol. 71, no. 5, 2005.