International Journal of Bioelectromagnetism
Vol. 14, No. 4, pp. 223 - 228, 2012

www.ijbem.org

Improved Detection of Nonlinearity in Nonstationary Signals by Piecewise Stationary Segmentation

Mahmoud Hassan [a,c], Jon Guðnason[b], Jérémy Terrien [c], Brynjar Karlsson[b] and Catherine Marque [d]

[a]Laboratoire de traitement du signal et d'image (LTSI), INSERM, Université de Rennes 1, Rennes, 35000 France
[b] School of Science and Engineering, Reykjavik University, Reykjavik, 103 Iceland
[c]Université de Technologie de Compiègne – Department of electronics, F60205 Compiègne Cedex France
[d] Université de Technologie de Compiègne – CNRS UMR 6600 BioMécanique et BioIngénierie, F60205 Compiègne Cedex, France.
ccorrespondence: Mahmoud Hassan, Université de Technologie de Compiègne – Department of electronics, F60205 Compiègne Cedex France.
Email : hassan@ru.is; Phone :

Abstract. Recently, much attention has been paid to the use of nonlinear analysis techniques for the characterization of biological signals. These signals are however often strongly non-stationary. This is in contradiction to the assumption of commonly used nonlinearity measures, which assume that the signal is stationary. We propose to use a Piecewise Stationary Segmentation (PSP) of the signals of interest, before the computation of nonlinearity measures. We show on synthetic as well as real signals (speech and uterine EMG) that the proposed piecewise stationary segmentation approach increases the accuracy of the measures by making a good trade-off between the stationary assumption and length of the analyzed segments, when compared to the classical windowing method.
Keywords: Nonstationary signals, nonlinear analysis, speech phonemes sand uterine EMG signals.

1. Introduction

One of the most common ways of obtaining information about physiological systems is to study the features of the signal(s) using time series analysis. Due to the intrinsic nonlinearity of most of biological systems, nonlinear techniques are often very useful in analyzing them as they can detect both linear and nonlinear characteristics of the signals. To analyze the nonlinearity or complexity, different measures have been proposed [1]. Most of them however make some statistical assumptions on the signals of interest. Most nonlinear measures are only reliable in the analysis of long stationary time series. Many biological signals are however highly non-stationary. Nevertheless, using a sliding window in which the signals of interest are supposed to be stationary is usually used in nonlinear analysis of these non-stationary signals.

In the windowing approach, a trade-off between the length of the analysis window and the stationary assumption has to be made. The length of the window also limits the accuracy of the temporal detection of abrupt changes that can reflect biological mechanisms in the underlying systems. The length of the window is thus an important parameter that has to be set according to a prior knowledge of the minimal length of the stationary parts of the signals or by trial and error. Here, we are inspired by a similar approach used for synchronization analysis between signals. The approach has shown a better performance than the windowing approach in detecting relationships between non-stationary signals recorded at different locations [2].
The aim of this paper is to use a pre-processing step that segments automatically the signals of interest into their longest stationary parts. To do this we used an algorithm able to detect non-uniform stationary partitions in a signal. Applying it to both synthetic and real data shows the advantages of this approach in the determination of nonlinear characteristics. We also show an application to two real signals where the use of nonlinear analysis was increased the detection efficiency considerably.

2. Material and Methods

A. Nonlinear measures

Several nonlinear measures exist that can be used to detect the presence of nonlinear behavior inside signals. In this paper, we use the sample entropy:

***Sample Entropy (SamEn)**:* is the negative natural logarithm of the probability that two sequences similar for *m* points remain similar at the next point, where self-matches are not included in calculating the probability. High *SamEn* value indicates the presence of a strong complexity inside the signal while a lower value of *SamEn* indicates more self-similarity in this time series. Formally, given *N* data points from a time series *x*, to define *SamEn*, one should follow these steps:

1. Form N-m+1 vectors $X(1)$, ..., $X(N\text{-}m+1)$ defined by $X(i)=[x(i), x(i+1), ..., x(i+m-1)]$, for $1 \le i \le N\text{-}m+1$. Those vectors represent *m* consecutive values of the signal, commencing with the i^{th} point.

2. Calculate the distance between $X(i)$ and $X(j)$, as the maximum absolute difference between their respective scalar components:

$$d[X(i),X(j)] = \max_{k=1,2,...m}\left(\left|x(i+k)-x(j+k)\right|\right)$$

3. For a given $X(i)$, count the number of j ($1 \le j \le N\text{-}m$, $i \neq j$), such that the distance between $X(i)$ and $X(j)$ is less than or equal to $r*SD$:

$$B_r^m(i) = \frac{1}{N-m-1}\sum_{j=1, j\neq i}^{N-m}\Theta(r*SD-d[X(i),X(j)])$$

4. Calculate B_r^m as

$$B_r^m(i) = \frac{1}{N-m}\sum_{i=1}^{N-m}B_r^m(i)$$

5. We then increase the dimension to m+1 and calculate $A_r^m(i)$:

$$A_r^m(i) = \frac{1}{N-m-1}\sum_{i=1, j\neq i}^{N-m}\Theta(r*SD-d[X(i),X(j)])$$

6. Calculate A_r^m as:

$$A_r^m(i) = \frac{1}{N-m}\sum_{i=1}^{N-m}A_r^m(i)$$

7. Finally, the sample entropy is defined as:

$$SamEn(m,r,N) = -\ln\left(\frac{A_r^m}{B_r^m}\right)$$

*B. Piecewise stationary signal pre-segmentation (*PSP*)*

Several algorithms exist that can be used to segment a piecewise stationary signal. We used an algorithm developed by Carré and Fernandez [3]. This algorithm starts by a decomposition of a signal into successive dyadic partitions. The log-spectrum corresponding to each partition is computed and denoised by undecimated wavelet transform if appropriate. A binary tree of the spectral distance between two adjacent partitions is then formed. The search for the tree, which minimizes the sum of the spectral distances, is realized by the use of the best basis algorithm of Coifman-Wickerhauser. The PSP algorithm has been shown to be useful as pre-treatment of synchronization analysis [2].

C. Signals

*1) **Synthetic signals**:* To study the robustness of the proposed approach, we used chaotic Rössler while we control the 'complexity degree' inside the signal. The model is defined by:

$$dx/dt = -z - y$$
$$dy/dt = x + 0.15y$$
$$dz/dt = 0.2 + \mathrm{CD}(zx - 5)$$

Where **CD** represent the complexity degree. The ***y*** values were recorded at intervals of Δt=0.5. Noise was superimposed on each ***y*** value by the addition of Gaussian random variables, mean = 0, standard deviation = 0.1. The different **CD** values used here are represented in table 1.

***Table 1**: Parameters of Rossler signals*

Time (s)	0-30	30-60	60-90	90-120
CD	1.1	0.89	1.0	0.83

*2) **Speech signals**:*

Much interest has recently been devoted to the use of nonlinear approach for speech characterization and modeling, such as validating the presence of nonlinear behavior inside normal vowels [4] and the possible use of nonlinear features for speaker identification [5]. Most of these studies used methods that assume the stationarity of the signals. Here we use the PSP algorithm to show the effect of the nonstationarity on the nonlinear analysis in phonemes detection and characterization.

Speech signals from the labeled TIMIT database are used in this paper. It was developed jointly by Texas Instruments (TI) and the Massachusetts Institute of Technology (MIT) and is one of the most employed databases in the field of the Automatic Speech Recognition [6]. Signals were sampled at 16 kHz. It contains 10 read sentences from 630 speakers from 8 dialect regions of USA.

*3) **Uterine EMG signals**:*

Another area of interest for nonlinear analysis is the uterine EMG signals. Recent studies have showed the presence of nonlinear behavior inside uterine signals and the possibility of using these nonlinear features for clinical application such as the classification between pregnancy and labor contraction [7]. However, these methods have been applied to the whole signals without taking into account the nonstationary characteristics of the signals.

In this paper, real uterine EMG signals that were recorded on women during pregnancy and labor and were acquired at 200 Hz. A precise description of the experimental setup can be found in [8]. The EMG signals were segmented manually to extract segments containing uterine contractions.

Evaluation of methods

We compared *SamEn* computed on each stationary part of the signals or along a sliding window of finite length. On synthetic signals, we evaluated the robustness of the segmentation algorithm by Monte-Carlo simulations. For different noise levels (Gaussian white noise) the relative root mean square error (**RRMSE**) of the estimators were computed against the parameter values computed in the reference segments and can be defined as

$$\mathrm{RRMSE} = \frac{\mathrm{RMS}(SamEn_{seg} - SamEn_{ref})}{\mathrm{RMS}(SamEn_{ref})}$$

Where **RMS** is the *root mean square* and $SamEn_{ref}$ is the reference calculated on a very long stationary signal for each of the values for CD in table 1.

3. Results

A. Synthetic signals

Figure 1 presents a typical example of the synthetic signals used and the corresponding nonlinearity analyses with *SamEn*. The results obtained by the windowing approach show a nonlinearity pattern that approximately follows the reference function. High variations can be found during periods of low nonlinearity inside the signals. The stationary approach shows marked transitions between the different nonlinear degrees with constant parameter values. The differences between the reference **CD** values and the estimates are due to the intrinsic bias of *SamEn*.

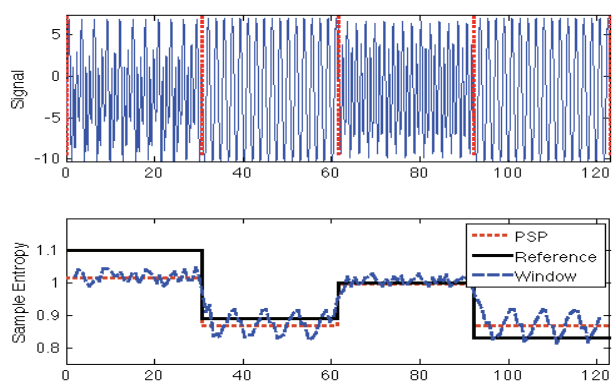

Figure.1. *Example of the output of the Rössler system (top panel) and the corresponding nonlinear analysis using SamEn (bottom panel) obtained by the PSP (dotted line) and the windowing approach for a window length of 8s (dashed line). The nonlinear function **CD** is presented as a continuous line.*

We compute **RRMSE** with different noise level with 100 Monte Carlo iterations. Results are presented in Figure 2. The stationary approach presents a lower error than the windowing approach whatever the noise level.

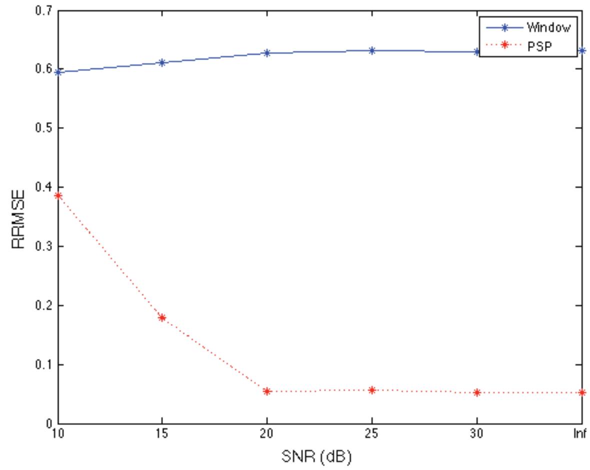

***Figure 2. RRMSE** values obtained with the PSP (dotted line) and the windowing approach (continuous line) for SamEn.*

B. Real signals

1) Speech Phonemes:

Here, we test our algorithm on phonemes detection for non-stationary characterization of phonemes in nonlinear analysis. Figure 3 shows how the PSP can correctly detect different phonemes types such as 'ih', 'sh', and 'n'; while we can notice the presence of high non-stationary characteristics in 'el', which may be a common characteristic of the this phoneme type.

We then compute the evolution of *SamEn* of reference segmentation (using TIMIT database), windowing and PSP segmentation. In Figure 4, the windowing approach gives a pattern similar to the reference although there are some discrepancies. The borders of the different zones are not sharp as they should be, but smoothed due to the windowing. The results of the PSP approach show a similar pattern to the reference except where we have a strong non-stationarity.

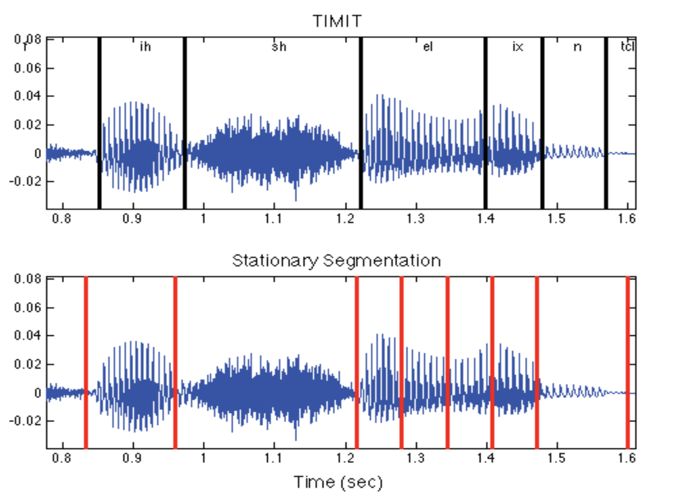

***Figure. 3.** Reference -TIMIT- segmentation (top panel) and segmentation obtained with the PSP algorithm (bottom panel).*

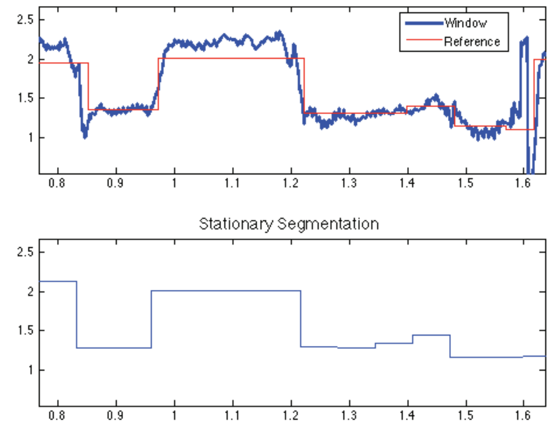

Figure 4. *Nonlinear analyses using SamEn computed by the windowing approach (top panel, reference complexity: continuous line) and by the PSP approach (bottom panel).*

2) ***Uterine EMG signals:***

Another important application of the proposed approach is the analysis of the uterine EMG signals. We expect that respecting the non-stationary characteristic of the EHG signals can improve the obtained results. Here, we use the approach to compare between the classification of pregnancy and labor contractions using the usual approach (applying the method on the entire burst), windowing and the PSP algorithm. An example of the application of PSP on labor burst is shown in figure 5. The figure shows the presence of strong non-stationary behavior inside the uterine burst during labor.

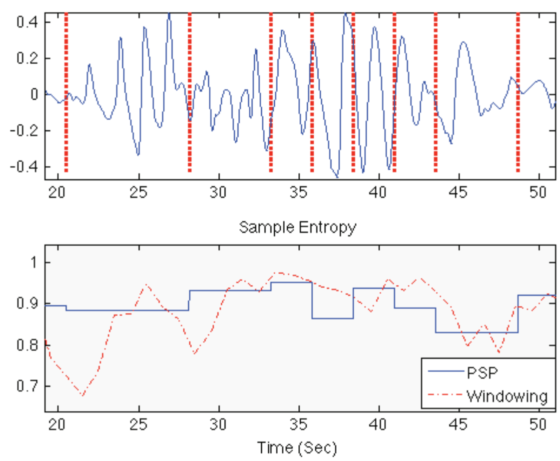

Figure 5 *Typical example of the segmentation process for labor contraction.*

ROC parameters are then computed to test the utility of the proposed algorithm in the classification of pregnancy and labor contractions. The results presented in table 2 indicate the improving of the classification rate when taking into account the no stationary characteristics of the signals. The results indicate also the superiority of the proposed algorithm on the windowing approach with Area Under Curve (AUC)=0.82 and 0.78 respectively.

Table 2.

	Specificity	Sensitivity	AUC
Whole burst	0.38	0.64	0.54
Windowing	0.76	0.82	0.78
PSP	0.83	0.85	**0.82**

4. Discussion

Most real signals are non-stationary. In the context of nonlinearity/complexity analysis, the most commonly used approach is the windowing of the signals of interest before computing the different nonlinear measures. The length of the analysis window is however an important parameter and controls the tradeoff between stationary assumption of the signals in the window and the

accuracy of the analysis. In this paper, we present the advantages of using an automatic segmentation procedure of the signal that searches for the longest locally adapted stationary parts.
Applying the PSP algorithm on the speech phonemes, we observed that the phonemes present different nonstationary characteristics and we think that these observations should be respected in applying nonlinear analysis to these signals. The application of PSP to the uterine EMG signals shows noticeable improving of the classification rate of pregnancy and labor contractions compared to the usual (whole burst) and windowing approach.

5. Conclusion

In this paper, we show the advantages of using the piecewise stationary pre-segmentation of non-stationary signals in the analysis of nonlinearity of systems. We used an algorithm dedicated to piecewise stationary signal segmentation. We believe that the approach can increase the accuracy of the nonlinear analysis of various systems. When applying to speech, this algorithm can open the door for new characterization of different phonemes regarding their nonstationary and nonlinear characteristics. Applying it to uterine EMG, this preprocessing step can highlight the difference in nonlinearity degree between pregnancy and labor contractions. Generally we think that the PSP method can vastly improve the performance of the nonlinear measures and to make them even better suited for different real applications.

References

[1] H. Kantz and T. Schreiber, *Nonlinear time series analysis* vol. 7: Cambridge Univ Pr, 2004.

[2] J. Terrien, M. Hassan, C. Marque, and B. Karlsson, "Use of piecewise stationary segmentation as a pre-treatment for synchronization measures," *Conf Proc IEEE Eng Med Biol Soc,* vol. 2008, pp. 2661-4, 2008.

[3] P. Carré and C. Fernandez, "Research of stationary partitions in nonstationary processes by measurement of spectral distance with the help of nondyadic Malvar's decomposition," in *IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, Pittsburgh, PA, USA, 1998, pp. 429-432.

[4] I. Tokuda, T. Miyano, and K. Aihara, "Surrogate analysis for detecting nonlinear dynamics in normal vowels," *The Journal of the Acoustical Society of America,* vol. 110, p. 3207, 2001.

[5] A. Petry and D. A. C. Barone, "Speaker identification using nonlinear dynamical features," *Chaos, Solitons & Fractals,* vol. 13, pp. 221-231, 2002.

[6] J. S. Garofolo, "Getting started with the DARPA TIMIT CD-ROM: An acoustic phonetic continuous speech database," *National Institute of Standards and Technology (NIST), Gaithersburgh, MD,* vol. 107, 1988.

[7] M. Hassan, J. Terrien, C. Marque, and B. Karlsson, "Comparison between approximate entropy, correntropy and time reversibility: application to uterine electromyogram signals," *Med Eng Phys,* vol. 33, pp. 980-6, Oct 2011.

[8] B. Karlsson, J. Terrien, V. Guðmundsson, T. Steingrímsdóttir, and C. Marque, "Abdominal EHG on a 4 by 4 grid: mapping and presenting the propagation of uterine contractions," in *11th Mediterranean Conference on Medical and Biological Engineering and Computing*, Ljubljana, Slovenia, 2007, pp. 139-143.