Biomedical Signal Processing and Control 49 (2019) 87–95

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Investigation and comparison of ECG signal sparsity and features variations due to pre-processing steps

S. Monem Khorasani, G.A. Hodtani*, M. Molavi Kakhki

Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran

ARTICLE INFO

Article history:
Received 5 August 2017
Received in revised form 30 October 2018
Accepted 17 November 2018
Available online 27 November 2018

Keywords:
Sparsity
Gini index
Entropy
Compressed sensing (CS)
Pre-processing
Savitzky Golay filtering
Signal derivatives

ABSTRACT

The pre-processing steps such as filtering, derivatives, and wavelet transform are necessary for many applications before data transmission, especially in telemedicine; however, the pre-processing makes variations on the signal sparsity, entropy, and compression metrics. In this paper, aiming at an information-theoretical study, we exemplify pre-processing by Savitzky Golay filtering because of its special properties, and then, show that (i) adding noise to an ECG signal decreases its sparsity and increases the diversity index named Gini-Sympson as a special case of Tsallis entropy; (ii) the sparsity of filtered, and wavelet transformed ECG is increased; (iii) Gini index of the modified signal is not more than that of the main one, but the non-zero elements are decreased, (iv) the compression metrics such as PRD and CR are improved if the compressed sensing method is performed on the filtered signal. And, finally, theoretical claims are validated by numerical results.

1. Introduction

The biomedical data acquisition methods usually result in unreal or missing values of the desired signal. Thus the recorded data should be presented in more proper form by applying pre-processing levels including some algorithms such as denoising, wavelet transform, feature extraction, and compression for efficient storage or transmission, especially in telemedicine plans. These processing algorithms can be implemented on different signals such as ECG [1–3], EEG [4,5] and image [6,7]. Signal compression is one of the main steps of signal preparation which is applied in all different types of data in bioelectrics and [8] and wireless sensor networks [9]. Compressive sensing (CS) is a method which senses the data in a compressive form directly at a much lower sampling rate in comparison to other compression techniques. The recovery stage then can be done by nonlinear optimization based on the few measurements [10]. All different kinds of compressed sensing methods have been used vastly in many applications from biomedical signals and images, telehealth, radar, communications, image and video processing and sensor networks [11–14].

Signal recovery with minimum error is one of the key aspects of a compression method, which the main aim is finding the sparsest signal while ensuring that the final signal is consistent with the measurements. The sparsity of each signal which can be measured by different indexes like l_0 or l_1 norms, and Gini Index (GI), is a key constraint to recover the original signal from the compressed one [15–17]. Also, the sparsity content of a signal can be evaluated via different concepts of entropy and diversity [18].

There are some ways to represent a signal in different domains via different coefficients. Signal filtering is an approach to prepare the original data in a suitable form and reject its undesired parts. One of the most important properties of different filter algorithms is their frequency response in the pass band and stop band, which determines whether the filter makes distortion on the original signal or not. Signal derivatives are another appropriate tool for data sparse representation and enhancement the processing levels which has been used in different signals such as ECG [3,19]. One of the applications of signal derivatives is signal denoising, which is called as Total Variation (TV) denoising. The TV denoising problem has received large attention in the communities of signal processing [20,21].

1.1. Our motivation

The main motivation for this work is finding a more detailed relation between a signal sparsity content and some preprocess-

* Corresponding author.
E-mail addresses: Sa_monem@yahoo.com, sa_mo805@um.ac.ir (S.M. Khorasani), hodtani@um.ac.ir (G.A. Hodtani), molavi@um.ac.ir (M.M. Kakhki).

https://doi.org/10.1016/j.bspc.2018.11.004
1746-8094/

ing tasks such as SG filtering, deviation, and wavelet transform. More precisely, in this paper, we try to increase the signal sparsity using some preprocessing levels before compressing the signal, which could lead to more efficient compression performance. We do the preprocessing stages on a noisy ECG signal and calculate the processed signal sparsity using Gini and l_0 norm indices, then compare the sparsity indices of the processed and not processed signals together. We also apply the entropy concept (specially Tsallis entropy) to compare the sparsity of pure and noisy signal.

To the best of our knowledge, despite expanding different algorithms of compression, or different approaches of solving sparse representation problems, the direct relation between the signal sparsity variations and processing algorithms have not been investigated in the literature.

1.2. Paper organization

In section II, some important concepts of our work are explained in brief, such as sparsity, entropy and diversity, SG filters, wavelet transform, compressed sensing and compression performance metrics. In section III, we prove that a) Sparsity of noisy ECG signal is less than that of the clean signal, b) Sparsity of filtered signal by SG filter is more than the sparsity of primary signal, c) Compression ratio (CR) of the compressed filtered signal is more than that of not filtered one, d) Percentage root-mean-squared difference (PRD) of the compressed signal which has been filtered first, is less than that of the compressed signal without any pre-processing stage, and e) SG filters does not change the main time and frequency features of the original signals. Finally we compute Gini index for some ECG signals and express the numerical results in section IV. We conclude the paper in section V.

2. Preliminaries

In this section, we describe the mathematical basis of the sparsity and entropy concepts, Savitzky-Golay filtering, compressed sensing and some compression performance metrics.

2.1. Sparsity and entropy

Each signal x has elements with some probability distribution, which are represented via numerical coefficients. The elements with equal probability lead to similar coefficients, therefore the randomness of signal elements can be defined by the repetition of coefficients or their compressibility. This relation in a signal can imply the relation between entropy, diversity, and sparsity with different kinds of formulations [18].

The concept of sparsity has originally an economically basic and it is used in different areas such as wireless communications, image processing, signal recovery, denoising, compressed sensing, and signal representation [16]. In the sparse representation models, most of the signal information is concentrated in few big coefficients and other signal elements are neglected. Many different sparsity-based algorithms have been applied in noise cancellation and signal recovery routines, such as LASSO or Basis pursuit Denoising (BPD) [22–24].

Entropy is a measure of the unpredictability of information content which its relation with sparsity can be explained as relations (1) and (2) [18]:

$$\text{sparsity(P)} \uparrow \equiv \text{entropy(X)} \downarrow \tag{1}$$

$$\text{sparsity(P)} \downarrow \equiv \text{entropy(X)} \uparrow \tag{2}$$

In general, a sparse measure S is defined as a function which maps complex vectors to a real number. There are many measures of sparsity which are used to calculate a number which determines the sparsity of a set of coefficients $\vec{X} = [x_1\, x_2 \ldots x_j \ldots x_N]$. Some of these metrics are l_p norms, pq-mean, Hoyer, and Gini index (special Tsallis entropy). Gini index is the most applicable one because of having all the ideal features of a sparsity measure. Given a vector, $[x_1\, x_2 \ldots x_j \ldots x_N]$, we order from smallest to largest, $x_{(1)} < x_{(2)} < \ldots < x_{(N)}$ where (1), (2), . . ., (N) are the new indices after the sorting operation. The Gini Index is given by (3) [16]:

$$S\left(\vec{X}\right) = 1 - 2\sum_{k=1}^{N} \frac{x_{(k)}}{\left\|\vec{x}\right\|_1} \left(\frac{N - k + \frac{1}{2}}{N}\right) \tag{3}$$

Diversity is an important notion in a variety of scientific subjects such as physics, ecology and information theory which can be related to the entropy and sparsity. A diversity index is a quantitative measure that reflects how many different types there are in a dataset. For a given number of types, the value of a diversity index is maximized when all types are equally abundant. Having random variable X with probability distribution $p(x)$, some diversity indexes are as follow (relations (4) and (5)) [25]:

$$Simpson \equiv \sum_{i=1}^{N} p_i^2 \tag{4}$$

$$Gini - Simpson \equiv 1 - \sum_{i=1}^{N} p_i^2 \tag{5}$$

Among several extensions of entropy, Shannon, Renyi, and Tsallis are more stated in the literature about sparsity and diversity. Shannon and Renyi both have the additivity property for two independent random variables X and Y, means $H(X, Y) = H(X) + H(Y)$. Tsallis entropy is referred to a nonadditive one because it does not have an additivity for two independent random variables. Given a random variable X with probability distribution p(x), Tsallis entropy [26] is defined as (6):

$$S_q(X) \equiv -\sum_{x} p(x)^q \ln_q p(x), \; q \neq 1 \tag{6}$$

where q is a parameter as an extension of Shannon entropy, and q-logarithm function $\ln_q x$ is defined as the relation (7) for any nonnegative real number x and q.

$$\ln_q x = \frac{x^{1-q} - 1}{1 - q} \tag{7}$$

2.2. Total variation (TV) and signal derivatives

The total variation of a signal or a function can evaluate the signal changes between its different values. Total variation definition for an N-point signal x(n) is expressed as (8):

$$TV(x) = \sum_{n=2}^{N} \left|x(n) - x(n-1)\right| \tag{8}$$

It is indeed first order difference of a discrete signal and can also be written in this form (relation (9)):

$$TV(x) = \left\|Dx\right\|_1 \tag{9}$$

Where D is an $(N-1) \times N$ matrix as relation (10):

$$D = \begin{bmatrix} -1\,1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & -1\,1 \end{bmatrix} \tag{10}$$

The concept of TV is vastly applied as a denoising method called TV denoising or sparse derivative denoising [20]. This method refers to estimate a signal (x) from its noisy observation ($y = x + z$) while having sparse derivatives of the signal. The problem is equivalent to this optimization problem, $arg\ min_x \{\frac{1}{2}\|y-x\|_2^2 + \lambda \left\|Dx\right\|_1\}$, Where λ is a regularization parameter chosen to be positive. The problem can be rewritten as (11):

$$arg\ min_x \frac{1}{2}\sum_{n=0}^{N-1}\left|y(n)-x(n)\right|^2 + \lambda\sum_{n=1}^{N-1}\left|x(n)-x(n-1)\right| \tag{11}$$

Other approximations of higher order differences for the sparse-derivative denoising problem [20] can be applied. Higher order derivatives can decrease the staircase characteristic of the first order derivative, therefore leads to more flexible recovered signal. This extension of TV has been applied for MRI reconstruction, texture extraction and image processing [27].

2.3. Savitzky-Golay (SG) filter

Savitzky and Golay proposed a method of data smoothing based on local least-squares polynomial approximation. The low pass filters obtained by this method are widely known as Savitzky-Golay filters. They demonstrated that least squares smoothing reduces noise while maintaining the shape and height of waveform. Subsequently, this property of the SG filters has been found to be attractive in many applications such as ECG processing [28]. To design an SG filter (M,N), it should be chosen a frame size (FS) or window length (2M + 1) where M is half the width of window and N determines the degree of employed polynomial or polynomial degree (PN). Fig. 1 shows the frequency response of three SG filters varying the PN parameter while the FS is fixed [29].

The impulse response of SG filter is in form of (12):

$$\sim h(n) = \sum_{k=0}^{N}\sim a_k n^k, M \le n \le M \tag{12}$$

Because the odd-indexed coefficients of the impulse response design polynomial are all zero so that we can express (12) as:

$$\sim h(n) = \sum_{k=0}^{\left\lfloor \frac{N}{2}\right\rfloor}\sim a_{2k} n^{2k} \tag{13}$$

Now the output of the filter (relation (14)), can be explored via the convolution of the impulse response and the input:

$$y[n] = \sum_{m=-M}^{M} h[m]x[n-m] = \sum_{m=n-M}^{n+M} h[n-m]x[m] \tag{14}$$

replacing h[n] by (13) in (14), and assuming the input x[n] as the original signal adding noise we get the filter output as (15) [29]:

$$y[n] = \sum_{m=-M}^{M}\sum_{k=0}^{\frac{N}{2}}\sim a_{2k} m^{2k} x[n-m] \tag{15}$$

2.4. Compressed sensing (CS)

The main aim of CS is to reconstruct completely a signal based on its reduced measurements considering that the signal is sparse or approximately sparse as a priori information. If we suppose an n dimensional x signal, we can take the measurements via a measurement or sensing matrix A, as Eq. (16):

$$y = Ax + z \tag{16}$$

Where x is n×1 signal vector, y is m × 1 measurement vector or the samples, and A is m × n measurement matrix and z is measurement noise which can be interpreted as a stochastic unknown error term and can be modeled as an additive white Gaussian noise (AWGN). Depending on the applications, the main signal is often sparse itself, but sparsification of not sparse signals is required [22]. This sparsification stage is done by multiplying a transformation matrix Ψ in the original signal which can be written as (17):

$$\theta = \Psi x \tag{17}$$

And the CS observations can be written as Eq. (18):

$$y = A\Psi^{-1}\theta + z \tag{18}$$

here θ is the resulting sparse representative vector for x. It means that most of the θ elements are equal to zero [10].

In the compressed sensing recovery stage, the undetermined case of problem is interesting for us, in which we have fewer equations than unknowns, (m < n), therefore the problem is of course ill-posed case. But if we suppose that x is sparse or approximately sparse, the problem will be changed and it has been shown that the solution x* (relation (19)) recovers x precisely:

$$\min \|\sim x\|_1\ s.t. A\sim x = y \tag{19}$$

The relation between CS and approximately sparse signals can be stated via isometry constant and the restricted isometry property (RIP). The RIP concept measures how the sensing matrix A distorts the l_2-norm of sparse vectors. In general sensing matrices with the RIP are easy to guarantee the reconstruction process with minimum error. For each integer s = 1,2,3,. . ., isometry constant (δs) of a matrix A is the smallest number which satisfies the inequality (20) for all s-sparse vectors x [30]:

$$\left(1-\delta_s\right)\|x\|_2^2 \le \left\|Ax\right\|_2^2 \le \left(1+\delta_s\right)\|x\|_2^2 \tag{20}$$

In general, the purpose of a compression system is to ignore redundancy and irrelevant information efficiently from the original signal. Consequently, the error criterion measures the ability of the compression system to reconstruction the primary signal exactly from the compressed samples. There are different metrics for evaluation the compression method performance. Compression Ratio (CR) is one of the main measures which is defined as is the ratio between the size of the compressed file (m) and the size of the source file (n) and is stated as relation (21):

$$CR = \frac{n-m}{n} * 100 \tag{21}$$

Another compression criterion is Percentage root-mean-squared difference (PRD), which many compression algorithms have employed it to compare the results. PRD is defined as (22):

$$PRD = \sqrt{\frac{\sum_{i=1}^{n}[\tilde{x}-x]^2}{\sum_{i=1}^{n}x^2}} * 100 = \frac{x-\tilde{x}_{l_2}}{x_{l_2}} * 100 \tag{22}$$

here x is the original signal and $\sim x$ is the reconstructed one [31,32]. The link between the measured PRD and the signal diagnostic distortion is determined based on [33], which classifies the different values of PRD based on the signal quality controlled by a specialist. Table 1 reports the resulting different quality classes and corresponding PRD. According to this table, it can be understood that lower PRD leads to better signal reconstruction quality.

Table 1
PRD and corresponding quality class [33].

PRD	Reconstructed Signal Quality
0-2 %	V̈ery goodq̈uality
2-9 %	v̈ery goodör g̈oodq̈uality
≥ 9 %	Not possible to determine the quality group

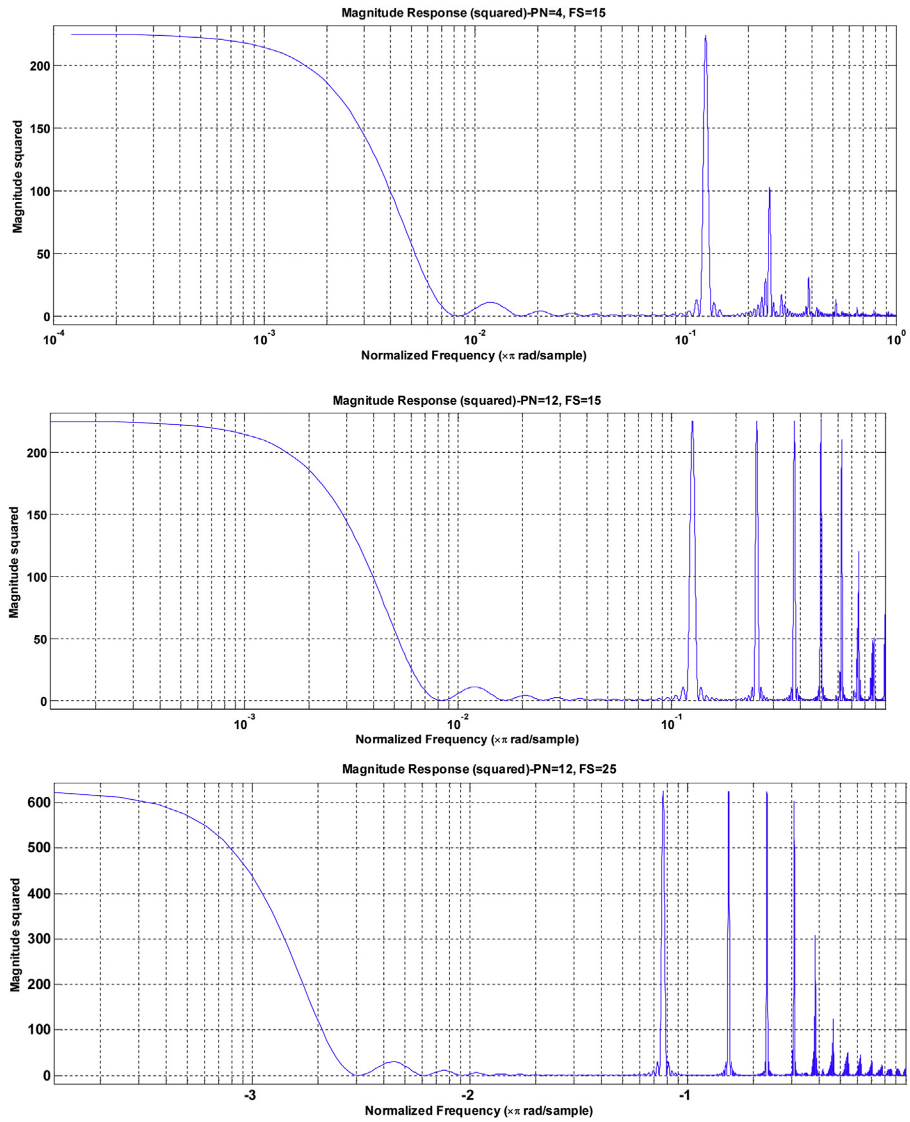

Fig. 1. Frequency response of different SG filters.

2.4.1. *Main results*

In this section, we investigate the sparsity and compression metrics variations due to the processing stages Specifically, it is proved that:

- Adding noise to a signal (filtering the noisy signal) decreases (increases) its sparsity.
- Performing derivation on a signal, SG filtering and wavelet transform sparsify the ECG signal.
- Compression ratio (CR) of the compressed filtered signal is more than that of non- filtered one.
- PRD of the compressed signal which had been filtered first will be less than that of the compressed signal without any pre-processing stage.
- The SG filtering does not alter main time and frequency features of original signals because of its special frequency response.

2.5. *The effect of noise on the sparsity*

First, we explain that adding noise to a signal reduces the sparsity of the primary signal, by using Shannon entropy and also by Tsallis entropy and related indices.

(i) It is easily shown that when X and Y are two independent random variables, we have $H(X+Y) \geq \max\{H(X), H(Y)\}$.

In a communication system or during a processing level where the input signal (X) has been polluted by additive white Gaussian noise (Z) in the output, the signal and noise often satisfy the independent assumption. Therefore it can be stated that:

$$H(X + Z) \geq \max\{H(X), H(Z)\} \tag{23}$$

The inequality in relation (23) means that the entropy of noisy signal is more than that of the original one, so based on relations (1) and (2) it can be concluded that the noisy signal sparsity is decreased in comparison to the pure signal.

- We can also use diversity and entropy relation to illustrate that a noisy signal sparsity is less than that of the pure one. As mentioned in section II, diversity of a distribution is measured by different numerical indexes like Gini-Sympson index. Applying Tsallis entropy relation (6) and substituting q parameter with 2, we then have relations (24) and (25):

$$S_{q=2}(X) = -\sum_{i=1}^{N} p_i^{q=2} ln_{q=2} p_i \tag{24}$$

$$S_2(X) = -\sum_{i=1}^{N} p_i^2 ln_2 p_i = -\sum_{i=1}^{N} p_i^2 (1 - \frac{1}{p_i}) \tag{25}$$

Knowing that $\sum p_i = 1$, (25) can be written as relation (26):

$$S_2(X) = 1 - \sum_{i=1}^{N} p_i^2 \tag{26}$$

It means that Tsallis entropy with q = 2 is Gini-Sympson index of diversity, relation (5). When in a noisy signal the entropy is more than that of the clean signal (due to relation (23)), we can say that the Gini-Sympson index for diversity of that signal also is more than that of the pure one. The greater diversity means the smaller sparsity.

2.6. The effect of filtering on the sparsity

Now we investigate the effect of filtering and wavelet transform (as a pre-processing stage) on the sparsity level of a signal. In the pre-processing level, the noisy signal (x + z) is filtered by SG filter and the added noise is filtered too. Here we imply that $S(y) > S(x+z)$, where y is the filter output. Knowing that the joint entropy of two random variable X and Y can be stated as (27), (widely used in information theory and easily proved from the definitions of $H(X)$, $H(X, Y)$, and conditional entropy):

$$H(X, Y) = H(XY) + H(Y) = H(YX) + H(X) \tag{27}$$

if Y = f(X) where f is a deterministic function, then $H(f(X)|X) = 0$, and obviously the inequality (28) can be concluded:

$$H(f(X)) \leq H(X) \tag{28}$$

We can apply the relation (24) to the SG filter input and output, because the filter output is a deterministic function of the input elements. The function is SG filter impulse response (13). Therefore performing filtering on the signal x + z will decrease its entropy, $H(f(x+z)) \leq H(x+z)$. Due to relations (1) and (2) it can be explained that that why the sparsity of filtered signal is more than the sparsity of the original signal, this means the following inequality, (29):

$$S(f(x+z)) \geq S(x) \tag{29}$$

2.7. Comparing CR and PRD of filtered signal and the original one

Here we show that increasing sparsity by pre-processing, improves the compression performance. Due to the sparsity assumption, we can suppose that $\boldsymbol{x_s}$ is the filtered signal which is k-sparse, $\boldsymbol{x}$ is the original signal which is p-sparse. Using the relation (21) we have:

$$CR_1 = \frac{k - m_1}{k}, \; CR_2 = \frac{p - m_2}{p}$$

Where m_1 and m_2 are the size of compressed nonfiltered signal and compressed filtered one respectively. Noting to $k < p$ and possibly assuming $m_2 < m_1$ for filtering, we have (30):

$$k < p, \; m_2 < m_1 \tag{30}$$

Then, it might be $m_2 p < m_1$ k, or relation (31):

$$\frac{k - m_2}{k} > \frac{p - m_1}{p} \tag{31}$$

And,

$$CR2 > CR1 \tag{32}$$

Considering $m_2 < m_1$, the inequality (32) can be true in some cases.

As it was mentioned before, lower PRD measures mean that reconstruction has been done more completely. In order to compare two compression algorithms with and without pre-processing levels, the PRDs are formulated and proved that PRD of the pre-processed compressed signal is less than that of the not processed compressed signal. First, we state some of the norm important inequalities which have been used in our analysis.

$$\|x + y\|_2 \leq \|x\|_2 + \|y\|_2 \tag{33}$$

$$\|x - y\|_2 \geq \|x\|_2 - \|y\|_2 \tag{34}$$

$$\left\|Mx\right\|_2 \leq \left\|M\right\|_2 \|x\|_2, (M\, is\, a\, matrix) \tag{35}$$

Now we show that $PRD_2 < PRD_1$, while PRD_2 stands for SG filtered compressed signal. For this purpose we begin with PRD definition according to the relation (22):

$$\frac{\|x_{SG} - \sim x_{SG}\|_2}{\|x_{SG}\|_2} < \frac{\|x - \sim x\|_2}{\|x\|_2} \tag{36}$$

where x_{SG} is the ECG filtered signal by SG algorithm, and $\sim x_{SG}$ is the reconstructed signal based on the x_{SG} compression. Also, x and $\sim x$ are the original and its decompressed signals, respectively. We can write the compression as the relation (16), with compression matrix A and the compressed signal y. It could be supposed that $\sim \boldsymbol{x}$ is approximately equal to $\boldsymbol{y}$ ($\sim x \cong Ax$), where the approximation can be modeled as relation (37):

$$\sim x = (A + \sim A)x \tag{37}$$

Also the filtered signal x_{SG} could be modeled as (38):

$$x_{SG} = E\,x \tag{38}$$

Where E is the coefficient matrix of the SG filter and the decompressed signal after compressing filtered one can be rewritten as relation (39):

$$\sim x_{SG} = (E + \sim E)x \tag{39}$$

Where $\sim A$ and $\sim E$ are small, by smallness we mean that $\left\|\sim A\right\|_2 < \left\|A\right\|_2$ and $\left\|\sim E\right\|_2 < \left\|E\right\|_2$. Now begin with the relation (36) and replace the component by (33–35) relations. The left side of (36) could be changed as following:

$$\frac{\|x_{SG} - \sim x_{SG}\|_2}{\|x_{SG}\|_2} = \frac{\left\|(E + \sim E)x - Ex\right\|_2}{\left\|Ex\right\|_2} = \frac{\left\|\sim Ex\right\|_2}{\left\|Ex\right\|_2} \tag{40}$$

The right side of (36) is rewritten as:

$$\frac{\|x - \sim x\|_2}{\|x\|_2} = \frac{\left\|(A + \sim A)x - x\right\|_2}{\|x\|_2} = \frac{\left\|(A - I)x + \sim Ax\right\|_2}{\|x\|_2} \tag{41}$$

Now the inequality (36) is changed to the following, using (40) and (41):

$$\frac{\left\|\sim Ex\right\|_2}{\left\|Ex\right\|_2} < \frac{\left\|(A - I)x + \sim Ax\right\|_2}{\|x\|_2} \tag{42}$$

Using (33) the right side of (42) could be written in the following form, relation (43):

$$\frac{\left\|\sim Ex\right\|_2}{\left\|Ex\right\|_2} < \frac{\left\|(A - I)x\right\|_2 + \left\|\sim Ax\right\|_2}{\|x\|_2} \tag{43}$$

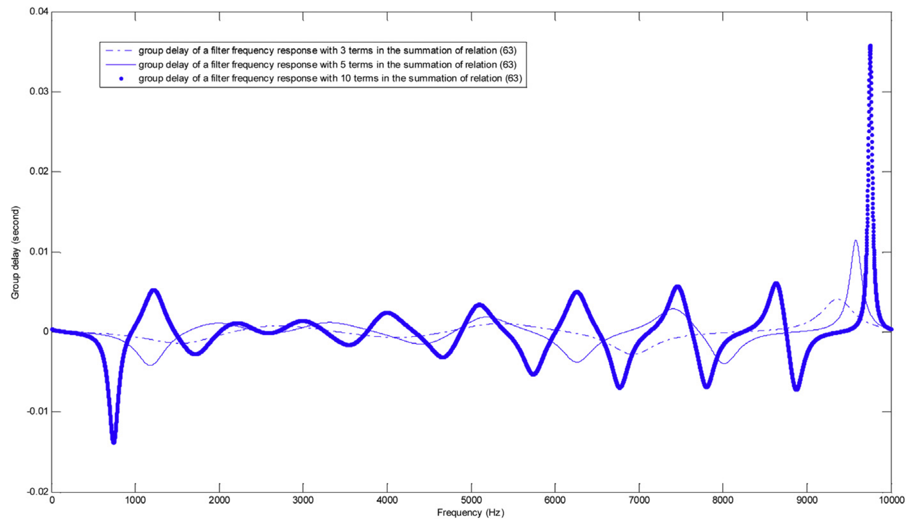

Fig. 2. Group delay of Savitzky Golay filters based on their frequency response.

Applying (35) in the left side of (42), we have (44):

$$\frac{\left\|\sim Ex\right\|_2}{\left\|E\right\|_2\|x\|_2} < \frac{\left\|(A-I)x\right\|_2 + \left\|\sim Ax\right\|_2}{\|x\|_2} \tag{44}$$

Then:

$$\frac{\left\|\sim Ex\right\|_2}{\left\|E\right\|_2} < \left\|(A-I)x\right\|_2 + \left\|\sim Ax\right\|_2 \tag{45}$$

Supposing that A, $\sim A$ *and* $\sim E$ have restricted isometry property and so they satisfy inequality (20) and there is some $0 < \delta_s < 1$ which:

$$\left\|(A-I)x\right\|_2 < \sqrt{(1+\delta_{s1})}\|x\|_2 \tag{46}$$

$$\left\|\sim Ax\right\|_2 < \sqrt{(1+\delta_{s2})}\|x\|_2 \tag{47}$$

$$\left\|\sim Ex\right\|_2 > \sqrt{(1-\delta_{s3})}\|x\|_2 \tag{48}$$

Then we can write (45) using inequalities (46–48) as (49):

$$\sqrt{(1-\delta_{s3})}\|x\|_2 < \left(\sqrt{(1+\delta_{s1}}\right) + \sqrt{(1+\delta_{s2})})\|x\|_2\left\|E\right\|_2 \tag{49}$$

Squaring all components we have:

$$1-\delta_{s3} < [\,1+\delta_{s1}+1+\delta_{s2}+2\sqrt{(1+\delta_{s1})}\sqrt{(1+\delta_{s2})}]\left\|E\right\|_2^{\;2} \tag{50}$$

In the inequality (50), the left side is a number smaller than one, and the right side is obviously greater than one, because all terms are positive. Therefore it is proven that PRD_2 is smaller than PRD_1.

2.8. Preserving main signal features after SG filtering

The amplitude of each part of the original signal will not be changed if the filter frequency response is flat in its pass band which SG filter satisfies.

To preserve frequency components of the original signal we should consider designing an appropriate filter with cut off frequency greater than the highest frequency in the original signal. The nominal normalized cutoff frequency depends on both the implicit polynomial order N and the length of the impulse response, (2M + 1). The passband of the filter gets wider increasing N [29].

To explain how filtering effects on the time features of the signal, it is better to find the group delay of the filter response. To reach this target we begin from the impulse response of the SG filter (13), performing z-transform on it, the relation (51) can be written:

$$H\,(z) = \sum_{n=0}^{+\infty}\sum_{k=0}^{N}\sim a_k n^k\, z^{-n} \tag{51}$$

It can be written as:

$$H\,(z) = \sum_{n=0}^{+\infty}\sim a_0\, z^{-n} + \sim a_1 n^1\, z^{-n} + \sim a_2 n^2\, z^{-n} + \ldots + \sim a_N n^N\, z^{-n} \tag{52}$$

Expanding relation (52) based on n, we have:

$$H(z) = \sim a_0 + (\sim a_0 + \sim a_1 + \sim a_2 + \ldots + \sim a_N)z^{-1} + (\sim a_0 + 2\sim a_1 + 4\sim a_2 + \ldots + 2^N\sim a_N)z^{-2} + \ldots \tag{53}$$

In relation (53) we can write all the coefficients of z^{-1} as some constant values and rewrite the relation as (54):

$$H(z) = \sim a_0 + \sim b_1 z^{-1} + \sim b_2\, z^{-2} + \ldots \tag{54}$$

Now substituting $z^{-1} = e^{-jw}$, it can be written as the relation (55):

$$H\left(e^{jw}\right) = \sim a_0 + \sim b_1 e^{-jw} + \sim b_2\, e^{-2jw} + \ldots \tag{55}$$

Which $e^{jw} = \cos(w) + j\sin(w)$ can be applied:

$$H\left(e^{jw}\right) = \sim a_0 + \sim b_1\,[\cos(w) - j\sin(w)] + \sim b_2\,[\cos(2w) - j\sin(2w)] + \ldots \tag{56}$$

The argument of relation (56) is:

$$\varphi(w) = \tan^{-1}\frac{\sim b_1\,[\sin(w)] + \sim b_2\,[\sin(2w)] + \sim b_3\,[\sin(3w)] + \ldots}{\sim a_0 + \sim b_1\,[\cos(w)] + \sim b_2\,[\cos(2w)] + \sim b_3\,[\cos(3w)] + \ldots} \tag{57}$$

The group delay of a frequency response is defined as $-\frac{d\varphi(w)}{dw}$, performing on the relation (57) we have:

$$T_g\,(w) = -\frac{d}{dw}\tan^{-1}\frac{\sum_{i=1}^{+\infty}\sim b_i\,[\sin(iw)]}{\sim a_0 + \sum_{i=1}^{+\infty}\sim b_i\,[\cos(iw)]} \tag{58}$$

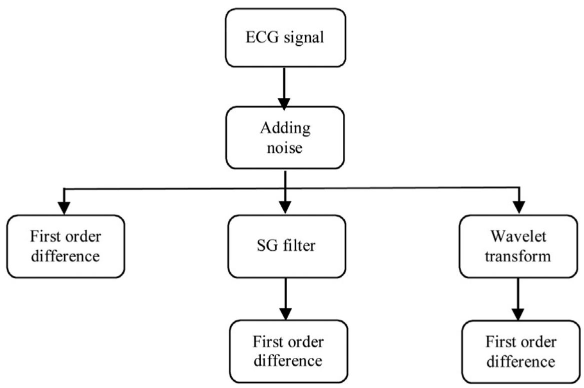

Fig. 3. the processing steps to compare signal sparsity.

Table 2
Numerical results of Gini index calculation.

	Main	1^{st} diff	Noisy	Noisy 1^{st} diff	SG filter	SG filter and 1^{st} diff
00	0.3401	0.1119	0.2657	0.1888	0.3346	0.1904
01	0.3382	0.1122	0.2678	0.1889	0.3191	0.1911
02	0.3320	0.1084	0.2439	0.1888	0.2836	0.1842
03	0.3076	0.1020	0.2753	0.1880	0.3209	0.1860
04	0.3336	0.1149	0.2440	0.1893	0.2983	0.1847

To investigate the variations of the group delay based on frequency *(w)*, we compute the derivatives considering some limited number of terms in relation (58), such as relation (59):

$$T_g(w) = -\frac{d}{dw}\tan^{-1}\frac{\sim b_1\ [sin(w)] + \sim b_2\ [sin(2w)]}{\sim a_0 + \sim b_1\ [cos(w)] + \sim b_2\ [cos(2w)]} \quad (59)$$

In Fig. 2 we plot the group delay of some Savitzky Golay filters based on relation (58). We can understand that the time delay of the different frequency components of a signal is approximately ignorable in low frequencies (passband of the filter). This means that different parts of the original signals which have different frequency ranges, pass through the filter with similar time delay. Having this property leads to a filtering level without any distortion, and this is the reason for applying SG filters in this study.

3. Simulation results

We make some processing levels on a pure ECG signal and a noisy one and rate their sparsity via computing Gini index and the number of non-zero elements (l_0 norm). Our goal is comparing the sparsity value of the main signal and the processed one to find a method which increases signal sparsity more than others. Aiming that we follow the block diagram in Fig. 3 and compute the sparsity of signal in every step.

To do that we apply the MIT-BIH database of cardiac signals. This databank consists of 48 signal records of 30 min from 48 different patients with 360 Hz sampling frequency. Approximately 60 percent of these records are for patients with arrhythmia [34].

We consider SG filter in two cases which are different in their polynomial degree (PN) and frame size (FS) to prevent the ability of comparing the sparsity of filtered signals. The polynomial degrees are 3 and 4 and the frame sizes are 41 and 21. Fig. 4 shows the main ECG signal and its noisy and differenced versions. Fig. 5 also shows the main, noisy ECG signal and its filtered one.

Table 2 illustrates the results of Gini index values of signals after different processing levels. In this table, we consider the SG filter with PN = 4 and FS = 21. The noisy signal is the main ECG signal adding to a Gaussian noise with SNR = 0.2 dB.

To have another index of sparsity, we use l_0 norm index which calculates the non-zero elements of signal. Table 3 shows the non-zero elements of main ECG signal and its processed ones via derivation, SG filtering. It can be concluded that 1^{st} derivative of a signal has less non-zero elements than the main signal. Also, the filtered signal is sparser than the noisy one but less than the original one. Performing derivation on the noisy and filtered signal is not effective way to increase the sparsity content of the signal.

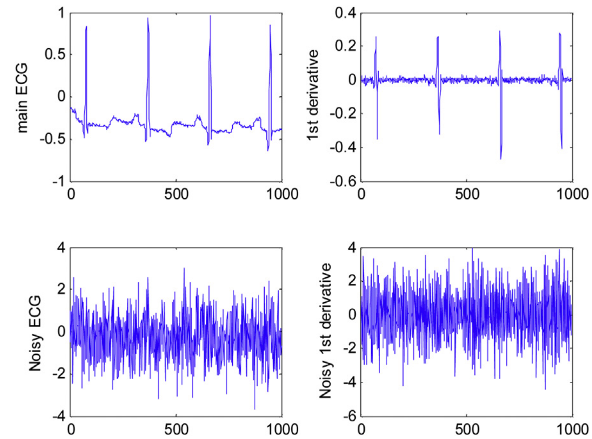

Fig. 4. Main ECG signal and its noisy and differenced version.

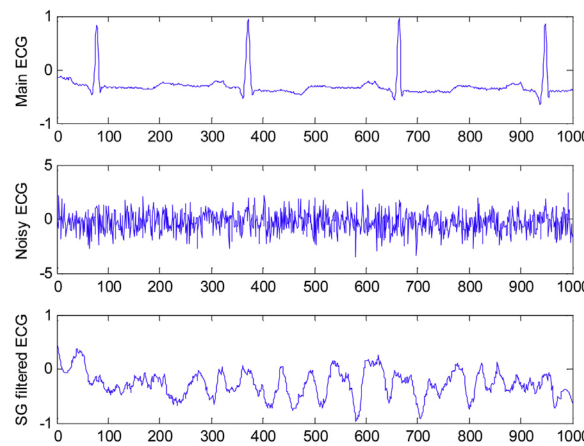

Fig. 5. Main ECG signal and its noisy and filtered version.

Table 3
Numerical results of l_0 norm index calculation (non-zero elements).

	Main	1^{st} diff	Noisy	Noisy 1^{st} diff	SG filter	SG filter and 1^{st} diff
00	33	20	367	476	223	382
01	27	21	380	484	224	391
02	93	19	409	488	294	391
03	82	35	336	491	204	405
04	65	34	405	496	302	371

Table 4 illustrate the difference between Gini and l_0 norm of the filtered signal applying two different SG filters with various polynomial degrees and frame sizes. In this comparison it is obvious that increasing the frame size of SG filter increases the Gini index and decreases the none-zero elements of signal. This means that changing the filter characteristics effectively can improve the sparsity of a signal.

Table 5 is about the results of wavelet transforming (daubechis wavelet in this research) main and noisy and derived ECG signal and computing none-zero elements. It is observed from this table that performing wavelet transform clearly improve the sparsity of signal, as the number of none-zero elements of the wavelet transformed noisy signal is less than the noisy one.

Table 4
Numerical results of Gini and l_0 norm index calculation of two SG filters.

	PN = 4, FS = 21				PN = 3, FS = 41			
	Gini		L_0		Gini		L_0	
	Filtered	Filtered 1st drivation	Filtered	Filtered 1st drivation	Filtered	Filtered 1st drivation	Filtered	Filtered 1st drivation
00	0.3346	0.1904	223	382	0.3646	0.1934	97	159
01	0.3191	0.1911	224	391	0.3592	0.1946	82	180
02	0.2836	0.1842	294	391	0.3271	0.1924	185	190
03	0.3209	0.1860	204	405	0.3509	0.1906	117	197
04	0.2983	0.1847	302	371	0.3276	0.1883	227	187

Table 5
Numerical results of l_0 after wavelet transform.

	Before wavelet		After wavelet	
	Noisy	Noisy 1st diff	Noisy	Noisy 1st diff
00	367	476	154	252
01	380	484	165	242
02	409	488	189	239
03	336	491	157	247
04	405	496	195	244

The results of simulations show that the Gini index of the ECG signal after adding noise decreases significantly, and SG filtering improves the Gini index of noisy signals. The numerical results are prepared at the Table 2 for five samples of the ECG database. It can obviously be concluded that SG filtering impact on the sparsity level of the noisy signals. Moreover, comparing the first and the last columns show that filtering increases the sparsity to more than the sparsity of the main signal.

4. Discussion

There are many cases which work on denoising different signals such as the image or biomedical data. Considering to our work which is concentrated on ECG signal, here we point more to some works in this field and compare our idea to them. In all of them it is only considered to perform some algorithms on a signal and evaluate their proposed method. [2] uses daubechies wavelets in ECG signal denoising and [3] studies on ECG enhancement and QRS detection based on signal derivatives, and SG filtering is one of the methods which has been used in ECG signal applications due to its time and frequency properties [28]. Many works have been done in the field of ECG compression [4,5], and some about other signals filtering and denoising [20,21], and compression [8,9,10,11,12,13,14]. On the other hand, a signal sparsity and sparse representation is one of the important subjects in the compression studies [15,22,23,24].

The most important idea of this study is to combine the mentioned methods and perform them on an ECG signal to explain some relation between signal processing and its sparsity, in order to have more efficient compression and recovery stages. To the best of our knowledge this point of view is not considered in any work before. As the application of pre-processing levels have been studied in [6] and [7] in order to get better results, in this work we compose three paths of processing which are proposed to have been performed before signal compression. We apply first derivatives, SG filtering and wavelet transform on pure and noisy ECG signal and compute the quality of the output signals in the form of sparsity measure, Gini index and l_0. All the simulation results show that the signal sparsity increases after pre-processing and therefore it could be better compressed.

To have a better view of the simulation results, we show the mean of Gini index and non-zero elements of the main and processed signals in Table 6. In this table, it is obvious that the mean value of Gini index and non-zero elements of a noisy filtered signal is improved compared to not filtered one.

Table 6
Mean of the results in different processing cases.

	Main	Noisy	SG filtered	Wavelet transformed
Mean of Gini index	0.3303	0.2593	0.3113	0.3424
Mean of l_0 (non-zero elements)	60	379.4	249.4	172

In addition to simulation the claims, we study the mentioned effect and the relation of sparsity with some entropy concepts (especially Tsallis entropy which is non additive entropy) in a mathematical form in section III. Also, in this section we investigate the effect of increasing the signal sparsity on the compression metrics such as CR and PRD.

5. Conclusion

In this work, the effect of the pre-processing stage such as filtering, signal derivation and wavelet transform have been studied on the sparsity level and recovery metrics of the ECG signal, by exploiting the concepts of Shannon entropy, Tsallis entropy and related indices, diversity and group delay of the SG filters frequency response. The claims were explained mathematically and validated by simulation. It has been concluded that the sparsity of a noisy signal is less than that of the main signal and filtering increases the signal sparsity. Evaluating the signal sparsity has been done via Gini index and non-zero elements of the signal in different cases. Also, we have concluded that compressing the filtered or wavelet transformed signal with more sparsity content leads to better reconstruction and lower compression error by using the measures of CR and PRD. We have validated the results numerically by simulating the proposed preprocessing levels on the ECG signals of MIT-BIH database. The simulation results show that we can increase the signal sparsity before its compression specially using filtering and wavelet transforming. This brings us advantages in the signal recovery level and leads to more exact signal reconstruction.

References

[1] N.H.S. Mukherjee, V. kumar, Application of wavelet techniques in ECG signal processing: an overview, Int. J. Eng. Sci.Technol. (IJEST) 3 (October (10)) (2011) 7432–7443.
[2] H. Murthy, M. Meenakshi, ECG signal denoising and ischemic event feature extraction using daubechies wavelets, Int. J. Comput. Appl. 67 (April (2)) (2013) 29–33.
[3] X. Ning, I.W. Selesnik, ECG enhancement and QRS detection based on sparse derivatives, Biomed. Signal Process. Control 8 (2013) 713–723.
[4] K. Srinivasan, J. Dauwels, R. Reddy, A two-dimensional approach for lossless EEG compression, Biomed. Signal Process. Control 6 (October (4)) (2011) 387–394.
[5] L. Lin, Y. Meng, J. Chen, Z. Li, Multichannel EEG compression based on ICA and SPIHT, Biomed. Signal Process. Control 20 (July) (2015) 45–51.
[6] A. Das, A. Kar, D. Bhattacharyya, Pre-processing for automating early detection of cervical cancer, in: 15Th International Conference on Information Visualization, IEEE Computer Society, 2011, pp. 597–600.
[7] C.N. Manisha, E.S. Reddy, Y.S. Krishna, A hierarchical pre-processing model for offline handwritten document images, Int. J. Res. Stud. Comput. Sci. Eng. (IJRSCSE) 2 (March (3)) (2015) 41–45.

[8] D. Craven, B. McGinley, L. Kilmartin, Glavin M, E. Jones, Compressed sensing for biomedical signals: a review, IEEE J. Biomed. Health Inform. 19 (March (2)) (2015).
[9] N. Kimura, S. Latifi, A survey on data compression in wireless sensor networks, IEEE International Conference on Information Technology: Coding and Computing (ITCC) (2005).
[10] E.J. Cands, M.B. Wakin, An introduction to compressive sampling, IEEE Signal Process. Mag. (March) (2008) 21–30.
[11] Z.H. Peng Zhang, R.C. Qiu, B.M. Sadler, A compressed sensing based ultra-wideband communication system, IEEE ICC 2009 Proceedings (2009) 1–5.
[12] J. Shi, W. Yin, A. Sankaranarayanan, R. Baraniuk, Video compressive sensing for dynamic MRI, BMC Neurosci. (2012).
[13] L. Weizman, Y. Eldar, D.B. Bashat, Compressed sensing for longitudinal MRI: An adaptive-weighted approach, Med. Phys. 42 (September) (2015) 5195–5208.
[14] K. Pawar, G. Egan, J. Zhang, Multichannel compressive sensing MRI using noiselet encoding, PLoS One (May) (2015).
[15] Y. Liu, I. Gligorijevic, V. Matic, M.D. Vos, S.V. Huffel, Multi-sparse signal recovery for compressive sensing, 34Th Annual International Conference of the IEEE EMBS Proceedings (2012) 1053–1056.
[16] N. Hurley, S. Rickard, Comparing measures of sparsity, IEEE Trans. Inform. Theory 55 (October (10)) (2009) 4723–4741.
[17] D. Zonoobi, A.A. Kassim, Y.V. Venkatesh, Gini index as sparsity measure for signal reconstruction from compressive samples, IEEE J. Selected Topics Signal Process. 5 (September)5) (2011).
[18] G. Pastor, I. Mora-Jimenez, R.J. antti abd, A.J. Caamano, Mathematics of sparsity and entropy: Axioms, core functions and sparse recovery, IEEE Trans. Inform. Theory, arxiv (2015).
[19] N.M. Arzeno, A.D. Deng, C.S. Poon, Analysis of first-derivative based QRS detection algorithms, IEEE Trans. Biomed. Eng. 55 (February (2)) (2008) 478–484.
[20] I. Selesnick, H. Graber, D.S. pfeil, R.L. Barbour, Simultaneous low-pass filtering and total variation denoising, IEEE Trans. Signal Process. 62 (March (5)) (2014).
[21] L. Condat, A direct algorithm for 1D total variation denoising, IEEE Signal Proc. Lett. 20 (11) (2013) 1054–1057.
[22] S. Ravishankar, Y. Bresler, Learning sparsifying transforms, IEEE Trans. Signal Process. 61 (March (5)) (2013) 1072–1086.
[23] S. Vujovic, Stancovic I, M. Dakovic, L. Stancovic, Comparison of a gradient-based and LASSO (ISTA) algorithm for sparse signal reconstruction, 5Th Mediterranean Conference on Embedded Computing (MECO) (2016).
[24] M. Tan, I.W. Tsang, L. Wang, Matching pursuit LASSO part 2: Appplications and sparse recovery over batch signals, IEEE Trans Signal Process. 63 (February (3)) (2015).
[25] Z. Zhang, M. Grabchak, Entropic representation and estimation of diversity indices, J. Nonparametr. Stat. (March) (2014).
[26] S. Furuichi, On uniqueness theorems for Tsallis entropy and Tsallis relative entropy, IEEE Trans. Inform. Theory 51 (October (10)) (2005) 3638–3645.
[27] Y. Hu, M. Jacob, Higher degree total variation (HDTV) regularization for image recovery, IEEE Trans. Image Process. 21 (May (5)) (2012).
[28] S. Hargittai, Savitzky–Golay least-squares polynomial filters in ECG signal processing, Comput. Cardiol. (September) (2005) 763–766.
[29] R.W. Schafer, What is a savitzky-golay filter? IEEE Signal Process. Mag. (July) (2011) 111–117.
[30] E.J. Cands, The restricted isometry property and its implications for compressed sensing, Comptes Rendus Math. 346 (March (9-10)) (2008) 589–592.
[31] S. Lee, J. Kim, M. Lee, A real-time ECG data compression and transmission algorithm for an e-health device, IEEE Trans. Biomed. Eng. 58 (September (9)) (2011) 2448–2455.
[32] Y. Zigel, A. Cohen, A. Katz, The weighted diagnostic distortion (WDD) measure for ECG signal compression, IEEE Trans. Biomed. Eng. 47 (November (11)) (2000) 1422–1430.
[33] H. Mamaghanian, N. Khaled, D. Atienza, P. Vandergheynst, Compressed sensing for real-time energy-efficient ECG compression on wireless body sensor nodes, IEEE Trans. Biomed. Eng. 58 (September (9)) (2011) 2456–2466.
[34] MIT-BIH Arrythmia Database Directory (2005). [Online]. Available: https://www.google.com/search?q=http%3A%2F%2Fwww.+physionet.org%2Fphysiobank%2Fdatabase%2Fmitdb&ie=utf-8&oe=utf-8&client=firefox-b-ab.