Comparison of Compression Methods for Impedance and Field Potential Signals of Cardiomyocytes

Pauline Guyot[1,2,4], Levy Batista[1,2,4], El-Hadi Djermoune[1,2], Jean-Marie Moureaux[1,2], Thierry Bastogne[1,2,3,4], Leo Doerr[5], Matthias Beckler[5]

[1]Université de Lorraine, CRAN, UMR 7039, Vandœuvre-lès-Nancy, France
[2]CNRS, CRAN, UMR 7039, Vandœuvre-lès-Nancy, France
[3]INRIA BIGS, Vandœuvre-lès-Nancy, France
[4]CYBERnano, Villers-lès-Nancy, France
[5]NANION, Munich, Germany

Abstract

We present in this paper an extensive comparison of compression methods adapted to impedance and field potential signals of cardiomyocytes. Different combinations of the traditional scheme of lossy compression have been tested and other original methods such as compressed sensing were implemented as well. All algorithms are assessed on several criteria such as compression ratio, distortion of the data, etc. We show that the selected method presents the ability and reliability to compress sensitive data with a compression ratio greater than 5:1 while preserving the relevant information content of the recorded data.

1. Introduction

To accurately identify potentially torsadogenic compounds in an earlier stage of drug development, innovative preclinical strategies including label-free impedance and extracellular field potential recordings of stem cell-derived cardiomyocytes have been recently proposed. Unfortunately, they produce high-content signals and the size of their data files may exceed 10GB, which prevents any web data transfer for remote data analysis. Clearly, those signals have a different origin and morphology than classical electrocardiograms. Therefore, the state-of-the-art compression algorithms for ECG [1,2] may not be optimal for those new signals. Our objective is to compare the performances of several compression algorithms applied to those data.

2. Methods

There exist two types of compression algorithm. Lossless compression has the ability to perfectly reconstruct the original data from the compressed ones (without any error) but their compression ratio ($R = \frac{size\ of\ original\ data}{size\ of\ compressed\ data}$) are low. In contrast, lossy compression introduces loss to obtain a bigger compression ratio. Generally, lossless data compression is a component of lossy algorithms.

2.1. Traditional scheme

Lossy compression schemes are composed of three main stages.

1) A mathematical transformation is applied to the original data. One can use Fourier transform (FFT), Wavelet transform (DWT), Cosinus transform (DCT), etc. The sparsest the signal will be in a domain, the more efficient the compression method will be.
2) A loss of information is achieved through quantization and thresholding. The idea is to further sparsify the new representation of the data by removing the coefficients with low magnitude which have small impact on the reconstruction.
3) Finally, a lossless compression algorithm is used to exploit the structure of the new (binary) signal.

The most critical step is the second one. By sparsifying the transformed data, the signal entropy is expected to drop. The signal entropy corresponds to the average number of bits required to write one coefficient. The explicit formula of entropy is:

$$entropy = -\sum_i P(x_i)\ log_2(P(x_i))$$

 ISSN: 2325-887X DOI:10.22489/CinC.2017.133-142

with P, the probability of one coefficient x_i to appear. It means that if a signal of 100 samples has an entropy of 3, the signal will take *at least* 3*100 = 300 bits in memory. The smaller the entropy is, the greater you can compress your data. It is important to mention that there is a trade-off between this loss of information and the correct reconstruction of the signal. The goal is to compress the signal without perceptible distortion of it. In order to check this error, the percentage root-mean square difference is computed. It measures the percent of energy difference between the original and the reconstructed data:

$$PRD = 100 * \sqrt{\frac{\sum_i (y_i - x_i)^2}{\sum_i x_i^2}} \quad (1)$$

with x the original data and y the reconstructed data. The greater the PRD is, the most compressed the data can be but the more you lost information.

2.2. Selected methods

In each step of the lossy compression algorithm, we selected different possibilities. Every method works on one-dimensional signals.

For the mathematical transformation stage, Wavelet and Cosinus transform were retained. Both returns signals of the same size than input data, contrary to Fourier transform (which generates complex coefficient, twice the size of the input). In general, impedance and field potential signals are sparse in both domains. After different tests, we selected Daubechies 4 as the most efficient family to represent our signals in the wavelet domain.

In the second step, a simple quantization of 10^{-3} was selected in order to reduce the signal entropy without increasing the resulting PRD. Then, a global and hard thresholding is dynamically computed.

Finally, we selected four lossless algorithms: Huffman coding, Run-Length encoding (RLE), Deflate and adaptive Huffman coding.

2.3. Compressed Sensing

Compressed Sensing (CS) does not follow the classical scheme. It is a recent technique based on sparse data representation. The idea is to bypass the Nyquist-Shannon theorem by reconstructing a signal from fewer samples of data than required [3-5].

Consider a real-valued vector $x \in R^N$ and the matrix Ψ which represents an orthonormal basis:

$$x = \Psi . \theta \quad (2)$$

where θ is the coefficient vector of x. If most of the coefficients of θ have negligible amplitude, the signal is said to have a sparse representation in the domain. The model of CS is:

$$y = \phi\, x = \phi\, \Psi . \theta \quad (3)$$

with $\phi \in R^{M*N}$ $(M \ll N)$ the sensing matrix and y is the compressed signal.

There are two main issues for CS:

1) The design of the sensing matrix is crucial. On the one hand, the less coefficients it selects the smaller the compressed signal will be; on the other hand, the small amount of selected coefficient has to fully represent the information of x. The matrix is constant throughout the process and must respect the RIP property.
2) The reconstruction of the original signal x from the compressed signal y is a complex problem, which can be solved through the relaxation of the l_0-norm with the l_1-norm and with greedy algorithms.

We selected the cosinus basis and tested different sensing matrix such as Gaussian, Binary and Binary Block Diagonal matrices [6]. Then, we selected Basis Pursuit (BP) for the reconstruction algorithm.

3. Results

3.1. Dataset

The databases used for the tests come from Nanion technologies (CardioExcyte 96) and are composed of impedance ($F_s = 77.35\,Hz$) and field potential ($F_s = 10kHz$) signal recording of 20 seconds each.

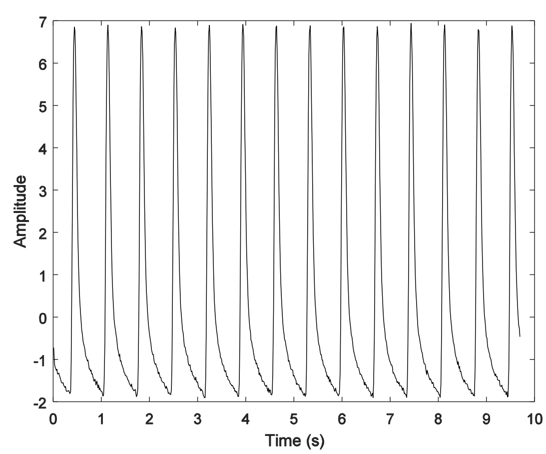

Figure 1. 10-second extract of an impedance signal, which represents the contractility of the cardiac tissue developed in the well.

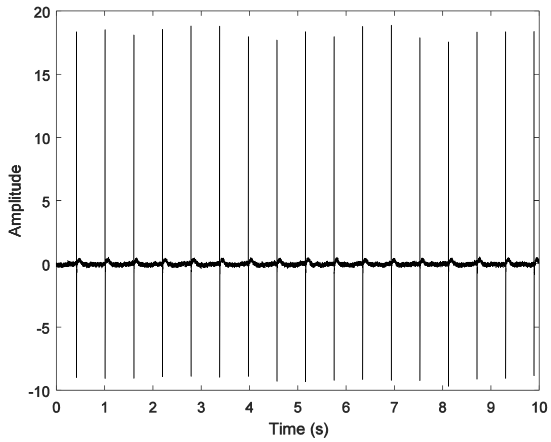

Figure 2. 10-second extract of a field potential signal, which represents the ionic exchange (K^+, Ca^{2+}, Na^+) occurring in the well.

3.2. Results with CS

To wisely choose the sensing matrix, we computed its incoherence with the dictionary at a given PRD.

Table 1. Coherence between the sensing matrix and the cosinus dictionary for an impedance signal.

PRD (%)	Gaussian	Binary	Binary Block Diagonal
5	5.54	5.05	4.47
4.5	5.62	5.05	3.99
3	5.44	5.74	3.16

The more incoherent the sensing matrix is with the dictionary, the greater the reconstruction of the signal will be. The deterministic Binary Block Diagonal matrix [6] presents the best incoherence.

However, the main drawback with the BBD matrix is its size. The quotient $\frac{size\ of\ original\ signal}{size\ of\ compressed\ signal}$ has to be an integer and for an easy use a multiple of 2 or 5. This matrix cannot work on all size of signals and zero padding leads to artefact in the reconstruction. Even when compressed with other dictionaries, edge effects and artefacts were damaging the reconstructed signals.

3.3. Traditional scheme algorithms

Regarding the mathematical transformation, DCT was chosen over DWT for impedance signal and DWT was selected for field potential because each domain lowered the signal entropy best for each type of signal after quantization (see Table 2).

Table 2. Signal entropy for impedance signals after quantization (on a database of 1537 signals of 20 seconds each)

Signals	DWT	DCT
Healthy	4.25	2.8
Slight arrhythmia	3.5	3.4
Arrhythmia	4.54	3.8
Noisy	5.8	5.72
Mean	**4.52**	**4.18**

After a dynamic thresholding, four lossless algorithms were tested on the databases. The compression ratio was the criteria of selection. Table 3 gathers the results for a database of impedance signals focusing on the compression ratio (one coefficient is stored on 16 bits).

Table 2. Compression ratio for impedance signals (on a database of 1537 signals of 20 seconds each)

Signals	Huffman coding	RLE	Deflate	Adaptive Huffman
Healthy	2.33	1.82	1.47	5.81
Slight arrhythmia	2.28	1.67	1.38	6.5
Arrhythmia	3.46	2.34	1.99	11.98
Noisy	11.72	6.91	6.09	15.62
Mean	**4.94**	**3.18**	**2.73**	**9.97**

Clearly, Adaptive Huffman shows the best results. It has the good compression of Huffman coding without the drawback of the storage of the dictionary created by the algorithm. The dictionary is dynamically and identically created in the compression and the decompression allowing the transfer of only the compressed signal.

The selected compression starts with a mathematical transform (DCT for impedance and DWT for field potential signals), followed by a quantization and dynamic thresholding and then Adaptive Huffman is applied. For the decompression, inverse Adaptive Huffman is computed on the compressed signal and then the inverse mathematical transform.

4. Conclusion

Results on all databases clearly present the ability and reliability of the different methods to compress sensitive data coming from cardiomyocytes and with our selected method, we achieved a compression ratio greater than 5:1. It enables biologists to use web-based remote analysis tools by reducing the size of their files without distortion of their data.

References

[1] Adamo A, Grossi G, Lanzarotti R, Lin J. ECG compression retaining the best natural basis k-coefficients via sparse decomposition. Signal processing and control 2015;15:11-17

[2] Lu Z, Youn Kim D, Pearlman W.A. Wavelet compression of ECG signals by the set partitioning in hierarchical trees algorithm. IEEE Transactions on Biomedical Engineering 2015;47:849-856.

[3] Balouchestani M, Raahemifar K, Krishnan S. Low sampling rate algorithm for wireless ECG systems based on compressed sensing theory. Signal Image and Video Processing 2013;9:527-533.

[4] Cambareri V, Mangia M, Pareschi F, Rovatti R, Setti G. A case study in low-complexity ECG signal encoding: how compressing if compressed sensing? IEEE Signal processing letters 2015;22:1743-1747.

[5] Craven D, McGinley B, Kilmartin L, Glavin M, Jones E. Compressed sensing for bioelectric signals: a review. IEEE Journal of Biomedical and Health Informatics 2015;19:529-540.

[6] Ravelomanantsoa A, Rabah H, Rouane A. Compressed sensing: a simple deterministic measurement matrix and a fast recovery algorithm. IEEE Transactions on Instrumentation and Measurement 2015;64:3405-3413.

Address for correspondence.

Pauline Guyot
CRAN – Campus Science
Boulevard des Aiguillettes
54506 Vandoeuvre-lès-Nancy
mail: pauline.guyot@univ-lorraine.fr