Classical low-pass filter and real-time wavelet-based denoising technique implemented on a DSP: a comparison study[*]

Ch. Dolabdjian[a], J. Fadili, and E. Huertas Leyva

Groupe de Recherche en Informatique, Image et Instrumentation de Caen[b], ISMRA et Université de Caen, 6 Bd du Maréchal Juin, 14050 Caen Cedex, France

Abstract. We have implemented a real-time numerical denoising algorithm, using the Discrete Wavelet Transform (DWT), on a $TMS320C3x$ Digital Signal Processor (DSP). We also compared from a theoretical and practical viewpoints this post-processing approach to a more classical low-pass filter. This comparison was carried out using an ECG-type signal (ElectroCardiogram). The denoising approach is an elegant and extremely fast alternative to the classical linear filters class. It is particularly adapted to non-stationary signals such as those encountered in biological applications. The denoising allows to substantially improve detection of such signals over Fourier-based techniques. This processing step is a vital element in our acquisition chain using high sensitivity magnetic sensors. It should enhance detection of cardiac-type magnetic signals or magnetic particles in movement.

1 Introduction

Discrete wavelet-based smoothing techniques are becoming a popular and an elegant alternative to classical linear filters (digital or analog). It is particularly adapted for analysis and non-parametric estimation of non-stationary signals such as those we aim to detect [1,2]. In our context, we define non-stationary signals as those with phenomena well localised in time (*e.g.* with isolated singularities).

It is now agreed that although linear Fourier-based approaches have long predominated because of their simplicity, they are of limited performance and cannot handle properly non-stationary signals [4]. Indeed, due to the infinite support nature of the Fourier basis waveforms (*i.e.* sines and cosines), classical linear filters present many deficiencies for analysing signals with high temporal variations and/or well localised in time. Because of the Heisenberg uncertainty [3,4], the Fourier transform spreads the energy of such temporally well localised signals on a high (possibly infinite) number of coefficients. This lack of sparsity makes discrimination of the signal from a noisy background a hard task.

From these preliminary considerations has grown the idea of expanding a signal over a family of elementary functions (waveforms called also *time-frequency atoms*) that are well concentrated both in time and frequency. Such bases are of interest because they can efficiently approximate many types of signals with just few atoms. Windowed Fourier transforms and Wavelet transforms are two important classes of local time-frequency decompositions. However, the former is very redundant in $L^2(\mathbb{R})$ and has a fixed window (mother function) length. This window length constancy poses the problem of compromise between time and frequency resolutions: a long analysis window yields a good frequency resolution but deteriorates the temporal resolution and *vice versa*. This compromise can be negotiated by considering a scale parameter in the window expression in order to analyse the signal in a multiscale framework. The wavelet transform is a much more flexible and an inherent multi-scale approach whose atoms possess good localisation in both time and frequency. It is designed to represent efficiently (*i.e.* sparse representation) signals with both smooth and transient features. Owing to these properties, in the last 10 years, the wavelet transform has become a very popular technique in digital signal processing for analysing and estimating signals in noisy backgrounds. Throughout this paper, we restrict our attention to discrete, finite-energy, real-valued and N-dimensional signals (without loss of generality $N = 2^J$) and to orthonormal bases of $L^2(\mathbb{R})$.

The organization of this paper is as follows. We first review some of the most important aspects of the wavelet transform. More comprehensive accounts can be found elsewhere [3–5]. Then, we will focus our attention on the Discrete Wavelet Transform (DWT) and the denoising problem. Results of the denoising procedure and implementation aspects are presented. This is followed by a theoretical comparison study between the nonparametric wavelet-based denoising and low-pass Fourier-based approaches. We finally conclude on this work and give some perspectives.

2 The wavelet transform

The dyadic orthonormal wavelet transform of a finite energy signal is defined as the inner product [10]:

$$d_{j,i} = \langle f, \psi_{j,i} \rangle = 2^{-j/2} \int_{\mathbb{R}} f(t) \psi\left(2^{-j}t - i\right) \mathrm{d}t, \ (j,i) \in \mathbb{Z}^2.$$
$$(1)$$

The coefficient $d_{j,i}$ is the detail coefficient (or the wavelet coefficient) at scale j and position i. The function ψ is the wavelet function (mother wavelet), whose collection of dilations j and translations i form an orthonormal basis of the Hilbertian space $L^2(\mathbb{R})$. Any continuous function qualifies if the admissibility conditions are satisfied: ψ is well-localised around zero (compact support) and oscillates ($\int \psi(t) \mathrm{d}t = 0$). These conditions can be strengthened to include more vanishing moments (up to an order R) and/or higher order continuous derivatives [5]. From a filter-bank point of view, the wavelet $\psi_{j,i}$ can be viewed as an octave bandpass filter in $[-\pi/2^j, -\pi/2^{j+1}] \cup [\pi/2^{j+1}, \pi/2^j]$.

Multiresolution approximations compute the approximation of the signal (the smooth component) at scale j as the inner product [10]:

$$s_{j,i} = \langle f, \phi_{j,i} \rangle = 2^{-j/2} \int_{\mathbb{R}} f(t) \phi\left(2^{-j}t - i\right) \mathrm{d}t, (j,i) \in \mathbb{Z}^2.$$
$$(2)$$

The coefficient $s_{j,i}$ is the smooth coefficient at scale j and position i. The function ϕ is the scaling function (father wavelet), whose collection of translations i form an orthonormal basis of $L^2(\mathbb{R})$. From a filter-bank point of view, $\phi_{j,i}$ can be viewed as a low-pass filter in the octave $]-\pi/2^{j+1}, \pi/2^{j+1}[$. The multiresolution decomposition of signal can be entirely described by the function $\phi(t)$ [10].

In digital signal processing, the pyramidal algorithm [10] is routinely used to calculate the discrete wavelet transform (DWT) of a sampled signal. The computational complexity of both the analysis and synthesis processes is in $O(N)$[1]. Mallat [10] has demonstrated the existence of a sequence of quadrature mirror filters (QMF) associated to the functions ϕ and ψ: a low-pass

---

filter h used to calculate the smooth component and a high-pass filter g for the detail component with the well-known relation

$$g(n) = (-1)^n h(1 - n).$$

Different choices for h and g are available in the literature [3,5] and to each mother wavelet is associated a pair of QMF filters. These filters are at the heart of the multi-resolution wavelet analysis. They allow an easy and fast implementation of the analysis-synthesis algorithms. Thus, the decomposition in equation (1) (resp. Eq. (2)) is equivalent to a high-pass (resp. low-pass) filtering operation followed by a decimation by 2. This is further detailed in the next section where the implementation steps are described.

3 Practical aspects of the denoising procedure

3.1 The analysis step

Here, we prefer the Daubechies wavelet family [5] given the maximum compactness of their support for a given number of vanishing moments. However, the question remains as how many vanishing moments R the wavelet should possess when used to estimate the signal. Its compactness must mitigate the extent of inter-coefficient correlations introduced by periodic boundary correction in computing wavelet coefficients at the limits (0 and $N - 1$) of a finite times series [4]. Furthermore, using a Matched Filter Theory argument, one has to chose the wavelet whose shape matches that of the signal of interest (e.g. ECG signal). From our simulations study, the Daubechies D4 wavelet has the necessary properties to satisfy these requirements and is therefore used in our context. This wavelet is associated to an approximation filter with 4 coefficients (c_0, c_1, c_2, c_3). For this filter, the DWT written in matrix form is illustrated in Figure 1. It transforms the smooth $N/2^{j-1}$-dimensional vector into smooth $(s_{j,i})$ and detail $(d_{j,i})$ $N/2^j$-dimensional vectors at scale j Figure 2.

3.2 The denoising step

Once the detail coefficients are calculated up to scale J, we can now apply our denoising algorithm. The main idea is that the DWT allows to represent effectively a large class of signals using only few coefficients, the remaining coefficients are essentially predominated by the noise. Then, a simple thresholding procedure is applied to the detail coefficients in each scale according to a kill-or-keep principle

$$d_{j,i}^* = \rho_\lambda(d_{j,i}) = d_{j,i} I(|d_{j,i}| > \lambda) \qquad (3)$$

where λ is the threshold and I the indicator function. This is the so-called hard thresholding technique in nonparametric curve estimation widely used in the statistics community (see e.g. [6,7]). An alternative shrink-or-keep procedure can also be used (soft thresholding). Soft

---

[1] Read order of N which means that there exists a constant $A > 0$ such as the complexity function behaves as AN.

$$\mathbf{W}_1.Y = \begin{bmatrix}
c_0 & c_1 & c_2 & c_3 & 0 & 0 & 0 & & & 0 & 0 & 0 \\
c_3 & -c_2 & c_1 & -c_0 & 0 & 0 & 0 & 0 & 0 & & 0 & 0 \\
0 & 0 & c_0 & c_1 & c_2 & c_3 & 0 & 0 & 0 & & & 0 \\
0 & 0 & c_3 & -c_2 & c_1 & -c_0 & 0 & 0 & 0 \\
& & 0 & 0 & 0 & c_0 & c_1 & c_2 & c_3 & 0\,0\,0 \\
& & 0 & 0 & 0 & c_3 & -c_2 & c_1 & -c_0 & 0\,0\,0 \\
& & & & & \cdot & \cdot & \cdot \\
& & & & & \cdot & \cdot & \cdot \\
& & & & & & \cdot & \cdot & \cdot \\
& & & & & & \cdot & \cdot & \cdot \\
0 & 0 & 0 & & & & 0 & 0\,0\,c_0 & c_1 & c_2 & c_3 \\
0 & 0 & 0 & & & & 0 & 0\,0\,c_3 & -c_2 & c_1 & -c_0 \\
c_2 & c_3 & 0 & 0 & 0 & & & 0\,0\,0 & 0 & c_0 & c_1 \\
c_1 & -c_0 & 0 & 0 & 0 & & & 0\,0 & 0 & c_3 & -c_2
\end{bmatrix}
\begin{bmatrix}
y_1 \\ y_2 \\ y_3 \\ \\ \\ \\ \\ \\ \\ \\ y_{N-3} \\ y_{N-2} \\ y_{N-1} \\ y_N
\end{bmatrix}
=
\begin{bmatrix}
s_1 \\ d_1 \\ s_2 \\ d_2 \\ . \\ . \\ . \\ . \\ . \\ . \\ s_{(N/2)-1} \\ d_{(N/2)-1} \\ s_{N/2} \\ d_{N/2}
\end{bmatrix}$$

Fig. 1. Example of the DWT (analysis step) of an N-dimensional signal vector Y written in matrix form. Here we used a Daubechies wavelet with 4 coefficients (D4). The matrix $\mathbf{W}_1$ includes both the convolution and the decimation operators.

Fig. 2. DWT analysis (left to right) and synthesis (right to left) steps of a discrete N-dimensional signal.

thresholding yields smaller variance but much higher bias than hard thresholding [8] and will therefore be omitted.

For each set of coefficients $d_{j,i}$, at scale j, the corresponding threshold λ_j is calculated [2]. Then, the comparison in equation (3) is applied. It is well known that, in the white noise case, the choice of threshold $\lambda = \sigma\sqrt{2\log N}$ is optimal in minimax sense [6]. The standard deviation of the noise σ can be estimated by $\hat{\sigma} = \frac{MAD(d_{1,i})}{0.6745}$, where MAD is the median absolute deviation. To avoid

calculating the $square-root$ operation, the detail coefficients are squared and directly compared to λ^2.

3.3 The reconstruction step

The completeness and uniqueness of the wavelet transform implies that a signal analysed up to a scale J can be perfectly reconstructed from its detail coefficients $d_{j,i}$ ($j \in \{1,\ldots,J\}$) and the remaining approximation coefficient(s) at the scale J [4]. Then, the thresholded DWT vector is reconstructed according to the synthesis steps in Figure 2, yielding a denoised version of the observed

---

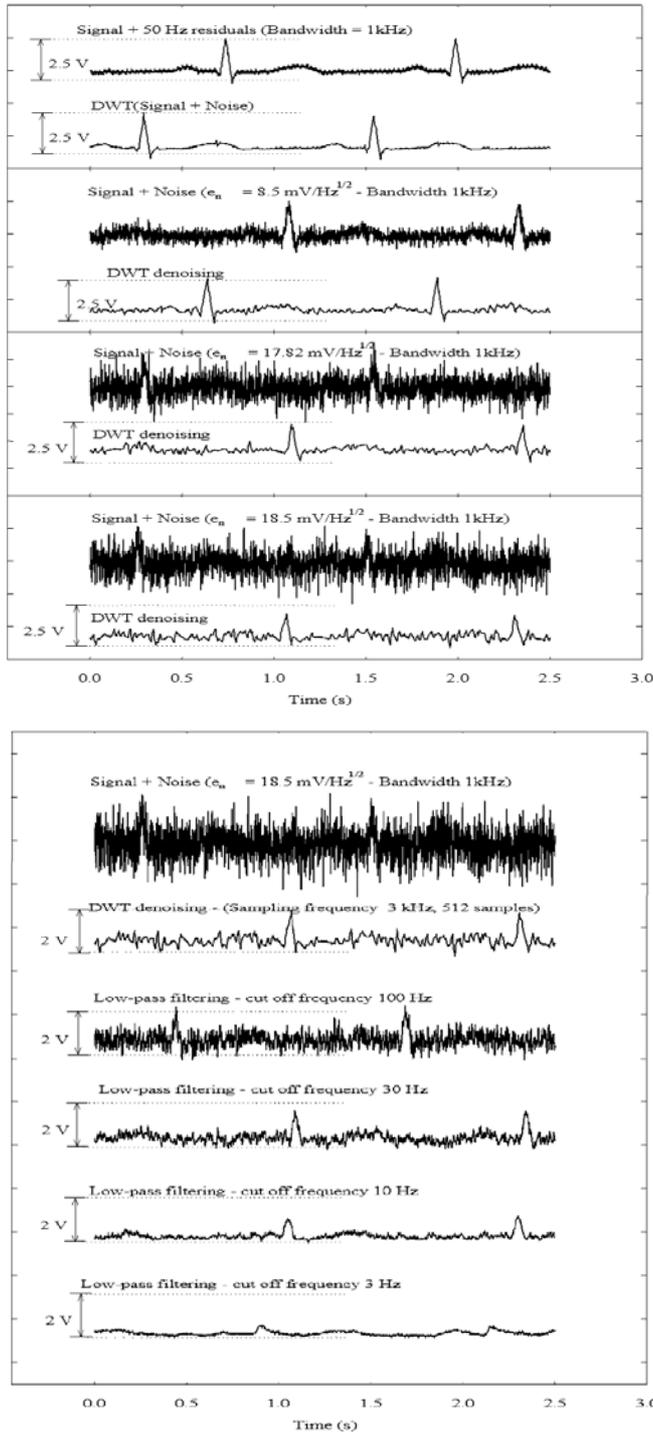[2] For the white noise case the thresholds λ_j at all scales are the same.

Fig. 3. Top: Results of the real-time implementation on a *DSP* of the *DWT*-based denoising approach (with hard thresholding), applied to a reference 512-samples *ECG*-type signal for different noise variances. Bottom: Comparison between the implemented *DWT*-based non-linear denoising and classical linear low-pass filter applied to the same reference signal for a low input *SNR* of −2.24 dB. Four successive cut off frequencies (100 Hz, 30 Hz, 10 Hz and 3 Hz) were tried for the low-pass filter. As we can clearly see the denoising approach outperforms the low-pass filtering technique at all frequencies. This will be further investigated in the theoretical comparison study.

signal. Note that, as the wavelet basis is orthonormal, the reconstruction matrix is the transpose of the analysis matrix.

4 Real-time implementation on the TMS320C3x DSP

Our processing system consists of a TMS320C31 digital signal processor (*DSP*) of Texas Instrument associated to a 14 bit analog-to-digital and digital-analog converters [1]. The correct operation of the program, first of all, was checked using the emulator provided with the *DSP* Starter-Kit in order to test the various stages of the algorithm calculation during decomposition, synthesis and thresholding. The real time processing of the denoising was carried out with two tables of 512 data which are alternatively managed in the mode "*in place*". Thus, on the same data table, the signal previously treated is transfered and the acquired signal is simultaneously recorded. During this time of acquisition, the digital processing is carried out on the other table. Hence, the signal is filtered in continuous way as in real time (see Fig. 3). The optimization of the program is performed directly by programming in assembly language code.

5 Signal-to-Noise ratio comparison

We suppose a degradation model where a signal $X(n)$ of size N (*e.g. ECG*-type signal of Fig. 3) is contaminated by the addition of a noise. This noise is modelled as the realisation of a stationary random process $\varepsilon(n)$ of known distribution. The measured data are:

$$Y(n) = X(n) + \varepsilon(n) \quad n \in \{0, \dots, N-1\} \cdot \quad (4)$$

In estimation theory, the goal is to minimize the estimation error, which is a loss function. A mean squared distance is certainly not perfect but it is sufficiently precise and mathematically simple in most applications. The estimation risk $\Phi\left(X, \tilde{X} = DY\right)$ is the average loss between the true signal X and its estimate $\tilde{X}$ obtained from the noisy observations transformed with a decision operator $\tilde{X} = DY$. In a Bayesian framework, the optimal estimator is the one maximising the posterior expectation $E(X(n)/Y(1,\dots,N))$. However, this conditional expectation is generally a complicated non-linear function and is difficult to estimate or to implement efficiently. This can be considerably simplified by limiting ourselves to a large but finite set of stochastic signals. Further simplification can be obtained under normality by invoking the central limit theorem. Thus our loss function is chosen to be the square Euclidean norm:

$$\Phi\left(X, \tilde{X}\right) = E\left(\left\|X - \tilde{X}\right\|^2\right). \quad (5)$$
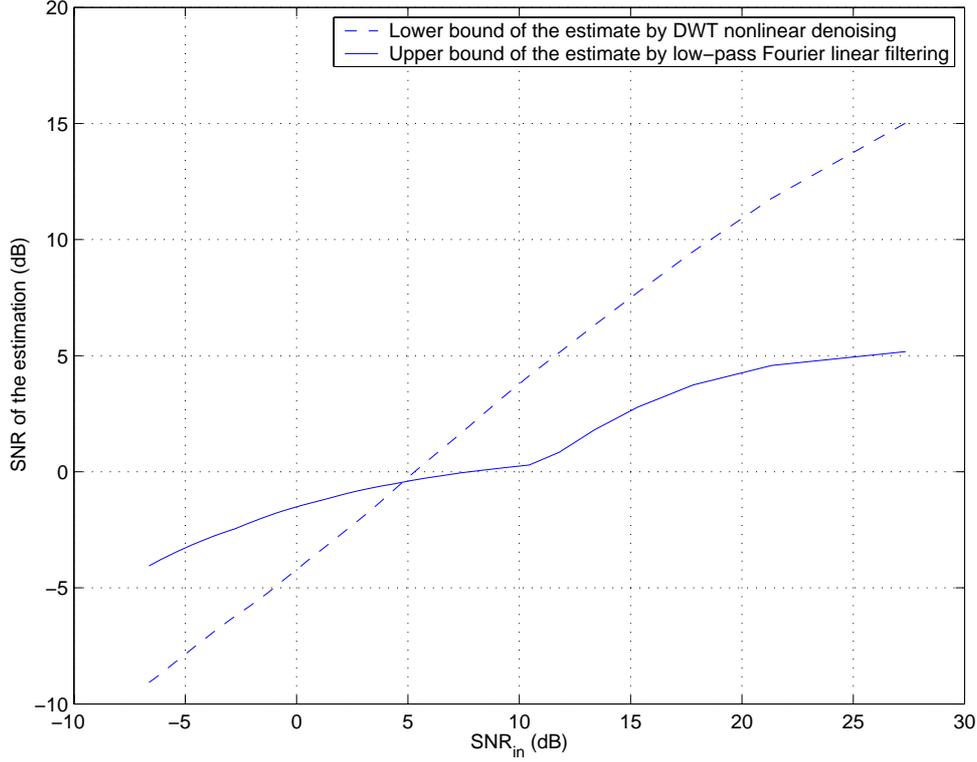
Fig. 4. SNR comparison between DWT non-linear denoising and classical low-pass linear filtering for a an ECG-type reference signal, as a function of the input SNR_{in}. The cut off frequency of the low-pass filter has been optimised for each simulated value of the noise variance (*i.e.* SNR_{in}). The quantization noise is supposed to be negligible.

5.1 Linear estimation

For a linear decision operator D (*e.g.* low-pass filter), the estimate can be written as the convolution:

$$\tilde{X}(n) = DY = \sum_{l=0}^{N-1} h(n-l)Y(l) = h * Y(n) \qquad (6)$$

where h is the impulse response of the linear filter. For the non-linear case, D can take much complicated expressions (*e.g.* Eq. (3)). In the following, we shall suppose that the noise process is white Gaussian with variance σ^2. The Fourier basis is a Karhunen-Loève basis for stationary signals. Taking the projection of equation (4) and using the Plancherel formula, we can show that the quadratic risk can be expressed

$$\Phi\left(X, \tilde{X}\right) = \sum_{n=0}^{N-1}\left[\left|\hat{X}(n)\right|^2\left(1-\hat{h}(n)\right)^2 + \hat{h}(n)^2\sigma^2\right] \quad (7)$$

where $\hat{X}$ and $\hat{h}$ are the discrete Fourier transforms of X and h. The first and second parts in equation (7) are interpreted as the bias (distortion) and the variance of the estimator. This quantity is often specified by the output SNR measured in decibels. For the linear low-pass filter g with a cut off frequency f_c and a noise specified by its power spectral density (PSD) e_n, the output SNR is:

$$SNR_{LPF}(dB) = 10\log_{10}$$

$$\times \frac{\sum_n \left|\hat{X}(n)\right|^2 |\hat{g}(n)|^2}{\sum_n \left|\hat{X}(n)\right|^2 |1-\hat{g}(n)|^2 + Ne_n^2\frac{\pi}{2}f_c}. \quad (8)$$

Mimicking this expression, the input SNR takes the classical form:

$$SNR_{in}(dB) = 10\log_{10}\frac{\sum_n \left|\hat{X}(n)\right|^2}{Ne_n^2\frac{\pi}{2}\text{Bandwidth}}. \quad (9)$$

5.2 Thresholding estimation

The main result of this paper is due to a pioneering work in nonparametric curve estimation by Donoho *et al.* [6,7]. This is stated in the following theorem:

Theorem 1 *For soft or hard thresholding and for the appropriate choice of threshold $\lambda = \sigma\sqrt{2\log N}$, the estimation risk in the basis $\mathcal{W}$ has an upper bound for all $N \geq 4$*

$$\Phi_{\mathcal{W}}(X) \leq (2\log N + 1)\left(\sigma^2 + \sum_{n=0}^{N-1}\min\left(X_{\mathcal{W}}(n)^2, \sigma^2\right)\right).$$
$$(10)$$

The lower bound on the output SNR for the DWT-based hard thresholding estimator is:

$$SNR_{\mathcal{W}}(dB) \geq 10 \log_{10}$$

$$\times \frac{\sum_{n=0}^{N-1}\left[X_{\mathcal{W}}(n)^2 I\left(|X_{\mathcal{W}}| > \lambda\right)\right]}{(2\log N + 1)\left(\sigma^2 + \sum_{n=0}^{N-1} \min\left(X_{\mathcal{W}}(n)^2, \sigma^2\right)\right)} \quad (11)$$

where $X_{\mathcal{W}}$ is the projection of X in the basis $\mathcal{W}$ (*e.g.* its DWT). The factor $\sqrt{2\log N}$ is optimal among all diagonal estimators in $\mathcal{W}$. Proof of equation (11) is a direct calculation from equation (10). The latter has been proved by Donoho *et al.* (see *e.g.* [6]).

5.3 SNR comparison

Using the SNR criterion, we can compare the two estimators (DWT denoising *versus* low-pass filter) in an objective way. This is portrayed in Figure 4 for the 512 samples ECG reference signal as a function of the input SNR_{in} (Bandwidth= 1kHz). In this figure, the lower bound of $SNR_{\mathcal{W}}$ in equation (11) has been calculated. The upper bound in equation (8) of the low-pass filter is also plotted. The cut off frequency was optimised for each variance value to maximise SNR_{LPF}. We can see that in addition to its adaptivity, the denoising is clearly superior to the simple low-pass filtering. This effect is accentuated as the noise variance increases where a lower cut off frequency is necessary to filter out the noise components but with the side effect of distorting the original signal. For real life ECG signals, the lower bound of the denoising approach is higher that the upper bound of the low-pass filter. In fact, it turns out that in practical situations, DWT denoising clearly outperforms the simple linear low pass filtering.

6 Conclusion

The implemented program uses a data size of 512 samples which is limited by the RAM size available on the DSP. In order to optimise the memory management, the calculations are executed *in place*. It is worth noting that increasing the regularity of the wavelet will impose a limit on the maximum decomposition scale. Indeed, the computing time must be less than the acquisition time of the 512 samples.

As far as the denoising is concerned, an improved Translation Invariant (TI) thresholding estimator can be obtained by denoising K translated versions of the signal and averaging them after reverse translations [9]. This has advantage of reducing the Gibbs-like effect [4]. Alternative denoising procedures can be applied in the Wavelet-Packet (WP) or Cosine-Packet (CP) bases [4]. Simulation results under Matlab should the superiority of the translation invariance, WP and CP denoising which are very encouraging leading to substantial gain in SNR. All these alternatives require additional calculations (*e.g.* $O(KN)$ for TI or $O(N \log N)$ for WP). The real-time implemented program as well as the hardware must be adapted to satisfy these requirements and allow efficient real-time processing.

References

1. C. Dolabdjian, S. Saez, A. Rayes Toledo, Colloque Interdiscplinaire en Instrumentation C2I'98 **1**, 683 (1998).
2. A. Aldroubi, M. Unser, *Wavelets in Medicine and Biology* (CRC Press, Boca Raton FL USA, 1996).
3. M.V. Wickerhauser, *Adapted wavelet analysis: from theory to software* (Wellesley IEEE Press, New York, 1994).
4. S.G. Mallat, *A Wavelet tour of signal processing second edition* (Academic Press, New York, 1999).
5. I. Daubechies, *Ten lectures on wavelets* (PA SIAM, Philadelphia, 1992).
6. D.L. Donoho, I.M. Johnstone, Biometrika **81**, 425 (1999).
7. D.L. Donoho, I.M. Johnstone, Minimax estimation via Wavelet Shrinkage, Annals of statistics (1994).
8. A.G. Bruce, H.-Y. Gao, Understanding WaveShrink: Variance and Bias Estimation, Technical report StatSci Division MathSoft Inc, 1995.
9. R. Coifman, D.L. Donoho, Translation-invariant denoising, Technical Report 475, Dept of Statistics Stanford University 1999.
10. S.G. Mallat, IEEE Trans. PAMI **11**, 674 (1989).