# AUDIO SIGNAL DENOISING WITH COMPLEX WAVELETS AND ADAPTIVE BLOCK ATTENUATION

*Guoshen Yu, Emmanuel Bacry, Stéphane Mallat*

CMAP, Ecole Polytechinque, 91128 Palaiseau Cedex, France

## ABSTRACT

We investigate a new audio denoising algorithm. Complex wavelets protect phase of signals and are thus preferred in audio signal processing to real wavelets. The block attenuation eliminates the residual noise artifacts in reconstructed signals and provides a good approximation of the attenuation with oracle. A connection between the block attenuation and the decision-directed *a priori* SNR estimator of Ephraim and Malah is studied. Finally we introduce an adaptive block technique based on the dyadic CART algorithm. The experiments show that not only the proposed method does eliminate the residual noise artifacts, but it also preserves transients of signals better than short-time Fourier based methods do.

***Index Terms*—** Complex Wavelets, Block, Ephraim and Malah, CART

## 1. INTRODUCTION

Major signal denoising techniques are based on attenuation in time-frequency signal representations. Short-time Fourier, whose atoms have a fixed scale, is until today the most popular time-frequency representation for audio signal processing. Complex wavelet representation has a resolution in time and frequency that depends on wavelet scales [12]. The short-time Fourier representation is well suited for analyzing stationary parts of signals, whereas the highly local wavelet atoms in high frequency bands allow to capture transient features.

Most attenuation rules can be applied independently of the representation used, although they were initially investigated in one or another. The classic audio noise reduction methods in the short-time Fourier representation include Wiener estimator and "spectral subtraction" algorithms [11, 14]. They generally lead to a residual noise artifact referred to as "musical noise" [4]. Ephraim and Malah proposed some noise suppression rules, together with a decision-directed recursive estimator of the *a priori* signal-to-noise ratio (SNR), that efficiently reduce the musical noise [9, 10]. Their suppression rules have been reinvestigated through years [4, 16] and a non-causal *a priori* SNR estimator has been proposed [5].

The hard or soft thresholding technique [7] is a powerful non-linear estimator often used in wavelet estimation. However, its direct application in audio signal denoising is problematic, as it creates some non-uniform "liquid noise" artifact, also called "bird noise" [17]. Some improvement may be achieved through more delicate thresholding functions [15]. A more effective way is to do attenuation using the Ephraim and Malah decision-directed *a priori* SNR estimator [1, 6]. As reported in [17], astonishing as it is, all the existing references employ real wavelets in audio signal processing, whereas complex wavelets are more adapted as they protect signal phase and thus help to reduce significantly the liquid noise artifact.

In this paper, we investigate the block attenuation methods that were initially applied in orthogonal wavelet signal representations [3]. We study the block size and the thresholding level in redundant time-frequency signal representations and we see that the block attenuation eliminates the residual noise artifacts through a temporal regularization and it provides a good approximation of the attenuation with oracle. Moreover, we point out that there is a close connection between the block attenuation and the decision-directed *a priori* SNR estimator of Ephraim and Malah. Finally we develop a more flexible adaptive block technique based on the dyadic CART algorithm [2, 8]. The experiments show that not only the proposed method does eliminate the liquid noise, but it also preserves transients of signals better than short-time Fourier based methods with the Ephraim and Malah SNR estimator do.

## 2. TIME-FREQUENCY REPRESENTATIONS

### 2.1. Estimation in Time-Frequency Representations

Let $y$ be the noisy signal that is the sum of the desired signal $f$ and the noise $\epsilon$, i.e., $y[n] = f[n] + \epsilon[n]$, $0 \leq n \leq N-1$. We assume in this paper that $\epsilon$ is stationary and white Gaussian with zero mean, i.e., $\epsilon$ is independently identically distributed (i.i.d.) as $\mathcal{N}(0, \sigma^2)$, and it is independent of $f$.

Denote $\mathcal{B} = \{g_m\}_{m\in\Gamma}$ a frame in either short-time Fourier or complex wavelet representation. Decomposing $y$ in $\mathcal{B}$, we obtain $y_{\mathcal{B}}[m] = f_{\mathcal{B}}[m] + \epsilon_{\mathcal{B}}[m]$, $\forall m \in \Gamma$, where $y_{\mathcal{B}}[m] = \langle y, g_m\rangle$, $f_{\mathcal{B}}[m] = \langle f, g_m\rangle$ and $\epsilon_{\mathcal{B}}[m] = \langle \epsilon, g_m\rangle$. The signal denoising consists in estimating $f_{\mathcal{B}}[m]$ given $y_{\mathcal{B}}[m]$, $\forall m \in \Gamma$, i.e., $\hat{f} = \sum_{m\in\Gamma} \hat{f}_{\mathcal{B}}[m]\tilde{g}_m = \sum_{m\in\Gamma} D_m(y_{\mathcal{B}}[m])\tilde{g}_m$, where $\hat{f}$ is the denoised signal, $\hat{f}_{\mathcal{B}}[m] = D_m(y_{\mathcal{B}}[m]) = a[m]y_{\mathcal{B}}[m]$ is a diagonal estimate of $f_{\mathcal{B}}[m]$ and $\tilde{\mathcal{B}} = \{\tilde{g}_m\}_{m\in\Gamma}$ is the dual frame.

The estimator that minimizes the risk $E[\|f - \hat{f}\|_2^2]$ is the *attenuation with oracle* written as

$$a[m] = \frac{|f_{\mathcal{B}}[m]|^2}{|f_{\mathcal{B}}[m]|^2 + \sigma^2}. \tag{1}$$

Using a *hard thresholding* or a *soft thresholding* operator defined as $D_m(x) = \rho_T^H(x) = xI\{|x| > T\}$ or $D_m(x) =$

$\rho_T^S(x) = \mathrm{sgn}(x)(|x| - T)_+$ with the threshold $T = \sigma\sqrt{2\log_e N}$, we can mimic the performance of the inaccessible attenuation with oracle within a factor of $2\log_e N$ [7].

### 2.2. Short-Time Fourier Frame

A short-time Fourier frame can be written as $\{g_m\}_{m\in\Gamma} = \{w(t - u_n)e^{i\xi_k t}\}_{(n,k)\in\mathbb{Z}}$, for $m = (n, k)$, where $w(t)$ is a short-time window and $(u_n, \xi_k)$ are the sampling positions in the time-frequency plane.

Denoising with short-time Fourier representation may generate a residual noise artifact, referred to as musical noise, that is composed of sinusoidal components with random frequencies [4].

### 2.3. Gabor Wavelet Frame

A wavelet frame is $\{g_m\}_{m\in\Gamma} = \left\{\frac{1}{\sqrt{a^j}}\psi\left(\frac{t-nu_0a^j}{a^j}\right)\right\}_{(j,n)\in\mathbb{Z}}$, for $m = (j, n)$, where $\psi(t)$ is a wavelet function, $u_0$ is the time sampling step and $a > 1$ is the scale dilation factor. A Gabor wavelet $\psi(t) = w(t)\exp(i\eta t)$ is obtained with a Gaussian window $w(t) = \frac{1}{(\sigma_G^2\pi)^{1/4}}\exp\left(\frac{-t^2}{2\sigma_G^2}\right)$. If $\sigma_G^2\eta^2 \gg 1$, then $\hat{\psi}(w) \approx 0$ for $w < 0$. Such Gabor wavelets are considered to be approximately analytic [12]. A wavelet at scale $s = a^j$ is $\psi_s(t) = w_s(t)\exp(i\eta t/a^j)$, where $w_s(t) = \frac{1}{\sqrt{a^j}}w(t/a^j)$, $a = 2^{1/v}$ is the dilation factor and $v$ is number of voices per octave. In contrast to the short-time Fourier representation that has a fixed time-frequency resolution, wavelets have high resolution in time but low resolution in frequency at high frequency bands, and vice versa at low frequency bands. As an example, Fig. 2-a illustrates the Gabor wavelet representation of some clarinet notes.

Denoising with wavelet representation may generate a residual noise artifact, referred to as liquid noise, that is composed of randomly distributed wavelets which can appear at very small scales.

The Gabor wavelets are more adapted for audio signal processing than real wavelets because the former, like the Fourier, separate module and phase components while the latter has module of coefficients necessarily oscillatory. Estimators such as the attenuation with oracle and the hard thresholding deal with the modules only. Hence the Gabor wavelets protect phase of signals and generate less liquid noise.

Numerically we calculate the Gabor wavelet dual frame $\{\tilde{g}_m\}_{m\in\Gamma}$ from the Gabor wavelet frame $\{g_m\}_{m\in\Gamma} = \{\psi_j(n-u)\}_{\substack{j=0,1,\ldots,J\\u=0,1,\ldots,K-1}} + \{\phi_J(n-u)\}_{u=0,1,\ldots,K-1}$ using the Extrapolated Richarson algorithm [12]. Special attention should be paid to avoid that the scaling function $\phi$ of the *complex* wavelets covers all the negative frequencies. One should also beware of the aliasing problem that may destroy the analyticity of the Gabor wavelets. An approach that solves these issues consists in implementing first a $real$ wavelet transform $Wf^R(u,s) = \langle f, \psi_s^R\rangle$, with $\psi_s^R(t) = w_s(t)\cos(\eta t/s)$, and then obtaining the $complex$ wavelet coefficients $Wf(u,s)$ by taking the analytical part of $Wf^R(u,s)$.

## 3. BLOCK ATTENUATION V.S. EPHRAIM-MALAH

### 3.1. Block Attenuation

Audio signal denoising by element-wise thresholding is likely to create some residual noise artifacts, musical noise with short-time Fourier representation [4] or liquid noise with wavelet representation [17]. Some temporal regularization is needed in order to attenuate the artifacts.

The basic motivation of the block attenuation is that, if neighboring coefficients contain some signal, then it is likely that the coefficients of current interest also do, and consequently a lower threshold should be used [3]. The NeighBlock method [3] consists in grouping the wavelet coefficients at each scale into blocks $b_i$ and then using an identical attenuation factor $a[B_i]$ to all the coefficients in the block $b_i$:

$$a[B_i] = \left(1 - \frac{\lambda L\sigma^2}{S_L^2}\right)_+ \tag{2}$$

where $B_i$ includes $b_i$ and is twice large, such overlapping reinforcing the redundancy, $S_L^2 = \sum_{n\in B_i}|y_\mathcal{B}[n]|^2$, $L$ is the size of $B_i$ and $\lambda$ is the thresholding level. Working in orthogonal wavelet representation, the author proposed the optimal block size $L = \log N$ and thresholding level $\lambda \approx 4.5024$.

For audio signal denoising, blocks of larger size are required since $L = \log N$ does not provide enough temporal regularization: it hardly attenuates the liquid noise. In the following propositions, we see that our block attenuation asymptotically converges to the attenuation with oracle; we choose an appropriate block size $L$ and thresholding level $\lambda$; we see that the practical block attenuation eliminates the residual noise artifact and remains a good approximation of the attenuation with oracle. We first develop the properties in the orthogonal context and then extend the results to the redundant time-frequency representations.

**Proposition 1** (Asymptotical Convergence to Oracle)**.** *With the thresholding level* $\lambda = 1$*, if* $\forall B_i$, $\{|f_\mathcal{B}[m]|^2\}_{m\in B_i}$ *are asymptotically consistent, i.e.,* $\forall B_i, \forall m \in B_i$, $|f_\mathcal{B}[m]|^2 = \lim_{L\to+\infty}\sum_{n\in B_i}|f_\mathcal{B}[n]|^2/L$*, then the block attenuation defined in Eq.(2) converges to the attenuation with oracle.*

One can verify Prop. 1 noting that $S_L^2/L \to |f_\mathcal{B}[m]|^2 + \sigma^2$.

**Proposition 2** (Noise Distribution)**.** *Suppose that the underlying signal* $f$ *is zero, i.e.,* $y = \epsilon$*. Assume that the frame* $\mathcal{B}$ *is an orthogonal basis. Then we have*

$$\sqrt{L}S_L^2/\sigma^2 \xrightarrow{\mathcal{L}} \mathcal{N}(1, 1). \tag{3}$$

Prop. 2 is a corollary of the Central Limit Theorem. In practice, when $L \geq 50$ we see that the distribution of $S_L^2/\sigma^2$ is very close to $\mathcal{N}(1, 1/L)$.

**Proposition 3** (Block Size and Thresholding Level)**.** *Assume that the frame* $\mathcal{B}$ *is an orthogonal basis. We have: (i) Given a block* $B_i$ *of size* $L$ *large enough in which* $\{|f_\mathcal{B}[m]|^2\}_{m\in B_i}$ *are consistent, the "optimal" thresholding level is* $\lambda = 1 + 3\sqrt{1/L}$*. (ii) With* $\lambda = 1 + 3\sqrt{1/L}$*, a larger block reduces the risk of* $\hat{f}$*, given that* $\{|f_\mathcal{B}[m]|^2\}_{m\in B_i}$ *are consistent in the block.*

The "optimal" $\lambda$ is a tradeoff between the maximum noise removal that requires $\lambda \geq 1 + 3\sqrt{1/L}$, a corollary of Prop. 2,

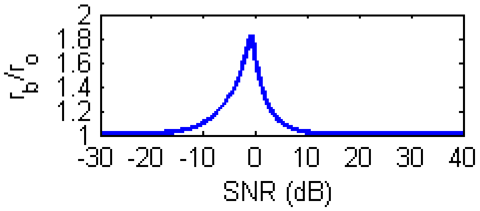

**Fig. 1**. Ratio of the risk of the block attenuation and the risk of the attenuation with oracle as a function of the SNR.

and the minimum signal distortion that demands $\lambda$ as close to 1 as possible, which is deduced from Prop. 1. With a larger $L$, not only the noise distribution in $S_L^2/\sigma^2$ is closer to $\mathcal{N}(1, 1/L)$, but the resulting $\lambda$ is also closer to 1. For example, with $L = 50$ and $\lambda = 1 + 3\sqrt{1/50} \approx 1.42$, we practically eliminate the residual noise artifact and the reconstructed signal is much less degraded than results of conventional thresholding. Smaller $\lambda$, for instance $\lambda = 1 + 2\sqrt{1/L}$, is also an option which distorts less the underlying signal and tolerates more residual noise.

**Proposition 4** (Practical Approximation to Oracle)**.** *The practical block attenuation is a good approximation of the attenuation with oracle. For example, with $L = 50$ and $\lambda = 1.42$, the risk of the block attenuation, denoted as $r_b$, is at most about 1.8 times of the risk of the attenuation with oracle, denoted as $r_o$, given that $\forall B_i$, $\{|f_{\mathcal{B}}[m]|^2\}_{m\in B_i}$ are consistent.*

Let us verify Prop. 4 quickly. Rewrite the block attenuation factor $a[B_i] \approx (1 - \lambda\sigma^2/(|f_{\mathcal{B}}[m]|^2 + k\sigma^2))_+$ with $k$ a random variable distributed as $\mathcal{N}(1, 1/L)$. One can verify that a smaller $k$ makes $a[B_i]$ further from the oracle attenuation factor $a[m]$ defined in Eq.(1). Thus it is enough to check a "bad" case with $k$ small, for example $k = 1-3\sqrt{1/L} \approx 0.57$ that happens with a probability inferior to 0.1%. Fig. 1 plots $r_b/r_o$ as a function of SNR, given $\lambda = 1.42$ and $k = 0.57$. We see that $r_b$ reaches the peak at about $1.8r_o$ when the SNR is about 0 dB. Indeed, with very high or very low SNR it is immediate that $a[B_i] \approx a[m]$. This is a significant improvement in comparison with the hard and soft thresholding.

Minor modification should be made when $\mathcal{B}$ is not orthogonal. If $\mathcal{B}$ is a redundant Gabor wavelet frame, for example, it is enough to use blocks of size $L' = KL$, where $K \approx 6s\sigma_G$ is the support of the Gabor wavelets.

### 3.2. Block Attenuation v.s. Ephraim-Malah

The decision-directed *a priori* SNR estimator of Ephraim and Malah [9] can be written in a more general way as

$$\hat{\xi}[m] = \alpha\frac{|\hat{f}_{\mathcal{B}}[m-1]|^2}{\sigma^2} + (1-\alpha)\left(\frac{|y_{\mathcal{B}}[m]|^2}{\sigma^2} - 1\right)_+ \quad (4)$$

where $\alpha$ is a weighting parameter. The first term is an empirical SNR of the previous coefficient and the second term is a maximum likelihood estimate of the SNR of the current coefficient. This *recursive* estimator imports a temporal regularization on $\hat{\xi}[m]$ with a causal smooth window exponentially decreasing towards the past.

Thus both the block attenuation and the Ephraim and Malah SNR estimator use regularization in time. The former uses a fixed rectangular window for all samples in a block while the latter employs a smooth sliding window.

## 4. ADAPTIVE BLOCKS

The motivation of the adaptive block technique is that we would like to use large blocks where modules of coefficients do not have much variation and otherwise we would prefer small ones.

We apply the dyadic CART [2, 8] algorithm to adaptively select the block size. The CART methodology of tree-structured adaptive non-parametric regression is built around ideas of recursive partitioning and it develops a piecewise constant reconstruction. Let us denote $R_k^p$ the $p$-th dyadic partition at depth $k$ of the whole block $R$ of size $L_{\max}$, i.e., $R = R_k^1 \cup \ldots \cup R_k^{2^{k-1}}$ and $R_k^p$ is of size $2^{-k+1}L_{\max}, \forall p$. The bottom-up dyadic CART algorithm that selects adaptively the block size from $\{2^{-k+1}L_{\max}\}_{k=1,\ldots,K}$ is summarized as follows. The output of the algorithm is $\{FLAG(i)\}_{i=1,\ldots,L_{\max}}$, where we find the partition depth for each sample in $R$.

```
***Initialization.***
Set $FLAG(i) = K, \forall\ i, 1 \le i \le L_{\max}$.
For each partition $R_K^p$, $p = 1, \ldots, 2^{K-1}$, compute $COST(R_K^p)$.
***Main Loop.***
For each partition depth $k = K-1, K-2, \ldots, 1$
    For each partition $R_k^p$, $p = 1, \ldots, 2^{k-1}$
        Compute $COST(R_k^p)$.
        If $COST(R_k^p) \le COST(R_{k+1}^{2p-1}) + COST(R_{k+1}^{2p})$
            Update the flags, i.e., set $FLAG(i) = k$,
            $\forall\ i,\ (k-1)\cdot 2^{-k+1}L_{\max}+1 \le i \le k\cdot 2^{-k+1}L_{\max}$
        end
    end
end
```

The algorithm decides to merge the children partitions into a parent partition if the cost of the parent is inferior to the sum of the costs of its children, or otherwise keep them splited. We define the cost $COST(R)$ of a partition $R$ as

$$COST(R) = \sum_{n\in R}(|y_{\mathcal{B}}[n]| - A_R)^2 + \lambda_C \quad (5)$$

where $A_R = \sum_{n\in R}|y_{\mathcal{B}}[n]|$ and $\lambda_C$ is a weighting parameter that balances between the amount of variation in the partition and the split cost that is set equal to 1 for each partition. The first term can also be interpreted as the $\mathbf{l^2}$ error of approximating the signal in the partition to a constant. With the second term, we penalize small blocks if they have the same variation as their parent. $\lambda_C$ should be chosen proportional to $\sigma^2$: to see this intuitively, one observes that, without the presence of the underlying signal $f$, i.e., $y = \epsilon$, the first term in Eq.(5) is proportional to $\sigma^2$. The dyadic CART is fast and it can achieve the global optimization for $N$ samples with a complexity $\mathcal{O}(N)$.

Fig. 2 shows the performance of the dyadic CART with an example of the adaptive blocks selected by the algorithm: in the region where the noise dominates, the majority of the blocks are of the largest size; in those places where the signal varies a lot, for example at the transitions between the successive harmonics, the algorithm selects correctly the smallest blocks.

## 5. EXPERIMENTS AND RESULTS

The experiments presented below have been performed on speech signals sampled at 11 kHz and corrupted by white

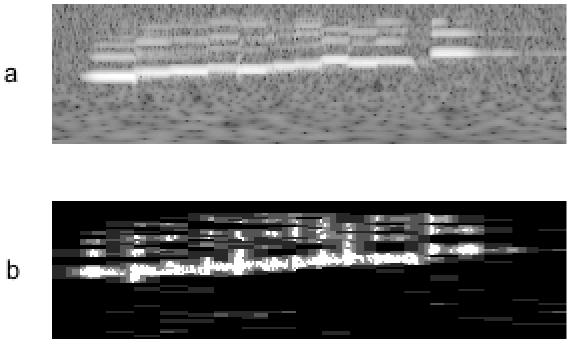

**Fig. 2**. a: log scalogram of some clarinet notes. b: adaptive blocks selected by the dyadic CART algorithm (gray level from light to dark: block size from 32 to 2048).

| | Input SNR (dB) | | | |
|---|---|---|---|---|
| | 0 | 5 | 10 | 20 |
| ABA-CW | 9.56 | 11.97 | 15.15 | 23.14 |
| HDT-CW | 8.78 | 11.00 | 13.93 | 21.86 |
| EMW-CW | 9.03 | 11.45 | 14.48 | 22.42 |
| EMW-STF | 9.17 | 11.43 | 14.37 | 22.16 |

**Table 1**. ABA-CW: Adaptive Block Attenuation with Complex Wavelets. HDT-CW: Hard Thresholding with Complex Wavelets. EMW-CW/STF: Ephraim and Malah decision-directed *a priori* SNR estimator + Wiener with Complex Wavelets / Short-Time Fourier.

Gaussian noise. We use the Gabor wavelets with $\sigma_G = 4$ and $\eta = \pi$ and we set $v = 12$ voices per octave and the maximum wavelet log scale $J = 84$, i.e., the wavelets cover the frequency plane down to 40 Hz. The dyadic CART algorithm selects adaptively the block among 6 sizes from $L' = 1024$ to $L' = 32$, which is equivalent to orthogonal samples from $L \approx 42$ to $L \approx 1.3$. We set the thresholding level $\lambda = 1 + 2\sqrt{1/L}$.

Table 1 presents the performance of different methods. The block attenuation (1st row) gains about 1 dB SNR in almost all cases over the complex wavelet thresholding (2nd row): the former eliminates the non-uniform liquid noise and degrades much less the underlying signal. In comparison with the short-time Fourier based Wiener filtering with the Ephraim and Malah SNR estimator (4th row), the complex wavelets based adaptive block attenuation (1st row) not only results in much less residual noise, but it also preserves better the transient parts of the signal; the stationary parts of the reconstructed signal are of similar quality. Let us note that the Ephraim and Malah SNR estimator using the complex wavelet representation (3rd row) does not perform as well as the block attenuation (1st row).[1]

A number of experiments have been performed on various music signals as well. The adaptive block attenuation performs well against the conventional thresholding operators. It results in sharper note transitions than the estimate with short-time Fourier. However, the short-time Fourier denoising outperforms the wavelet counterpart for the stationary parts when high pitch is involved (eg. musical signals). This is because the short-time Fourier has higher frequency resolution than wavelet representation in high frequency bands.

## 6. CONCLUSION AND FUTURE WORK

The block attenuation provides a good approximation of the attenuation with oracle and its connection to the decision-directed *a priori* SNR estimator of Ephraim and Malah is studied. An adaptive block attenuation based on the dyadic CART algorithm is introduced. The proposed method eliminates the residual noise artifacts and preserves transients of signals better than short-time Fourier based methods do.

We are currently working on an algorithm based on an adaptive time-frequency representation that is able to work on the Gabor wavelet atoms (for transient parts of signals) or the short-time Fourier atoms (for stationary parts of signals). Using grouping bandlets [13] may also help to solve the problem that wavelet representation does not have frequency resolution high enough in high frequency bands.

## 7. REFERENCES

[1] M. Bahoura and J. Rouat, “A new approach for wavelet speech enhancement”, *Proceedings of Eurospeech-2001*.

[2] L. Breiman, J. Friedman, R. Olshen, and C.J Stone, *Classification and Regression Trees*, Belmont, CA: Wadsworth, 1983.

[3] T. Cai and B.W. Silverman, “Incorporation information on neighboring coefficients into wavelet estimation”, *Sankhya*, 63, 127-148, 2001.

[4] O. Cappe, “Elimination of the musical noise phenomenon with the Ephraim and Malah Noise Suppressor”, *IEEE Trans. Speech and Audio Processing*, vol2, pp345-349, Apr.1994.

[5] I. Cohen, “Speech enhancement using a noncausal a priori SNR estimator”, *Signal Processing Letters, IEEE*, vol. 11, Issue 9, pp. 725-728, Sept. 2004.

[6] I. Cohen, “Enhancement of speech using Bark-scaled wavelet packet decomposition”, *Eurospeech 2001 - Scandinavia*.

[7] D. Donoho and I. Johnstone, “Idea Spatial Adaptation via Wavelet Shrinkage”, *Biometrika*, vol. 81, pp. 425-455, 1994.

[8] D. L. Donoho, “CART and best-ortho-basis: a connection”, *Ann. Statist. 25 1870–1911*.

[9] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean square error short-time spectral amplitude estimator”, *IEEE. Trans. Acoust. Speech Signal Process*, vol. ASSP-32, pp. 1109-1121, Dec. 1984.

[10] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean square error log-spectral amplitude estimator”, *IEEE Trans. on Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 443-445, Apr. 1985.

[11] J.S. Lim and A.V. Oppenheim, “Enhancement and bandwidth compression of noisy speech”, *Proc. of the IEEE*, vol.67, Dec.1979.

[12] S. Mallat, *“A Wavelet Tour of Signal Processing, 2nd edition”*, New York Academic, 1999.

[13] S. Mallat, “Geometrical grouplets”, to be submitted, 2006.

[14] R.J. McAulay and M.L. Malpass, “Speech enhancement using a soft-decision noise suppression filter”, *IEEE Trans. on Acoust. Speech and Signal Processing*, 28:137-45, April 1980.

[15] H. Sheikhzadeh and H. R. Abutalebi “An improved wavelet-based speech enhancement system”, *Proceedings of Eurospeech-2001*, pp. 1855-1858.

[16] P. J. Wolfe and S. J. Godsill, “Simple alternatives to the Ephraim and Malah suppression rule for speech enhancement”, *IEEE Workshop on Statistical Signal Processing*, pp. 496-499, Aug. 2001.

[17] P. J. Wolfe and S. J. Godsill, “Audio signal processing using complex wavelets”, Preprint 5829, presented at the 114th Convention of the Audio Engineering Society, 2003.

[1] The audio denoising samples (speech and music) are available online at http://www.cmap.polytechnique.fr/˜yu/research/audio/samples.html.