Consistent Wiener filtering: generalized time-frequency masking respecting spectrogram consistency

Jonathan Le Roux, Emmanuel Vincent, Yuu Mizuno, Hirokazu Kameoka, Nobutaka Ono, Shigeki Sagayama

HAL Id: inria-00544072
https://inria.hal.science/inria-00544072

Submitted on 7 Dec 2010

Consistent Wiener Filtering: Generalized Time-Frequency Masking Respecting Spectrogram Consistency

Jonathan Le Roux[1], Emmanuel Vincent[2], Yuu Mizuno[3],
Hirokazu Kameoka[1], Nobutaka Ono[3], and Shigeki Sagayama[3]

[1] NTT Communication Science Laboratories, NTT Corporation,
3-1 Morinosato Wakamiya, Atsugi, Kanagawa 243-0198, Japan
[2] INRIA, Centre Inria Rennes - Bretagne Atlantique,
Campus de Beaulieu, 35042 Rennes Cedex, France
[3] Graduate School of Information Science and Technology, The University of Tokyo,
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
{leroux,kameoka}@cs.brl.ntt.co.jp,emmanuel.vincent@inria.fr,
{mizuno,onono,sagayama}@hil.t.u-tokyo.ac.jp

Abstract. Wiener filtering is one of the most widely used methods in audio source separation. It is often applied on time-frequency representations of signals, such as the short-time Fourier transform (STFT), to exploit their short-term stationarity, but so far the design of the Wiener time-frequency mask did not take into account the necessity for the output spectrograms to be consistent, i.e., to correspond to the STFT of a time-domain signal. In this paper, we generalize the concept of Wiener filtering to time-frequency masks which can involve manipulation of the phase as well by formulating the problem as a consistency-constrained Maximum-Likelihood one. We present two methods to solve the problem, one looking for the optimal time-domain signal, the other promoting consistency through a penalty function directly in the time-frequency domain. We show through experimental evaluation that, both in oracle conditions and combined with spectral subtraction, our method outperforms classical Wiener filtering.

Key words: Wiener filtering, Short-time Fourier transform, Spectrogram consistency, Source separation, Spectral subtraction

1   Introduction

Wiener filtering has been one of the most widely used methods for source separation for several decades, in particular in audio signal processing. To exploit the short-term stationarity of audio signals, it is very often applied on time-frequency representations [1], especially the short-time Fourier transform (STFT). However, classical Wiener filtering does not take into account the intrinsically redundant structure of STFT spectrograms, and its output is actually in general not the optimal solution. We show here that by ensuring that the output spectrograms are "consistent", i.e., that they correspond to actual time-domain signals,

we can obtain a more efficient filtering. Many of the most promising methods for source separation exploit spectral models of the sources (non-negative matrix factorization, Gaussian mixture models, autoregressive modeling, etc.) and, as these models are often based on Gaussian assumptions, they commonly involve Wiener filtering as a post-processing [2]. It is thus of tremendous importance to ensure that the information gathered by these algorithms is best exploited.

Wiener filtering can be formulated as the solution of a Maximum-Likelihood (ML) problem in the time-frequency domain without constraint on the space of admissible solutions. The classical solution then only involves a manipulation on the magnitude part of the spectrograms, leading in general to arrays of complex numbers which do not correspond to any time-domain signal. We generalize here the concept of Wiener filtering to time-frequency masks which can involve a manipulation of the phase as well in order to find the ML solution among consistent spectrograms. Formulating the problem as the minimization of an objective function derived from the Wiener likelihood and explicitely taking into account consistency, we present two methods to solve it: one consists in computing the exact optimum by solving the problem in the time domain; the other relies on a relaxation of the consistency constraints through the introduction of a penalty function promoting consistency of the output spectrogram. We already exploited the idea of consistency-promoting penalty functions for fast signal reconstruction from modified magnitude spectrograms [3] and to improve the modeling accuracy of the complex non-negative matrix factorization framework [4]. It enables us here to develop an efficient algorithm which computes an approximate solution close to the true optimum obtained with the time-domain method.

We evaluate the performance of the proposed methods compared to classical Wiener filtering on two tasks: separation of concurrent speech by two speakers under oracle conditions, and denoising of speech mixed with synthetic and real-world background noises where only the noise mean power spectra are known and the speech spectrum is estimated through spectral subtraction [5].

2   Wiener filtering and consistency

2.1   Maximum-Likelihood formulation

We assume that the observed signal x is the mixture of two signals, a target s_1 and an interference signal s_2. We further assume that the STFT coefficients S_1 and S_2 of the signals s_1 and s_2 at each time frame t and frequency bin ω are modeled as statistically independent Gaussian random variables with variance σ_1^2 and σ_2^2 respectively. For convenience of notation, we shall write $\nu^{(i)} = 1/\sigma_i^2$. Note that the case of several interference signals can be reduced, without loss of generality, to that of two sources only, as we assume in particular that the sources are not correlated.

Denoting by X the spectrogram of x, classical Wiener filtering consists in maximizing the log-likelihood of the STFT coefficients S_1 and S_2, which can be written, under the constraint that $X = S_1 + S_2$, as a function of $S = S_1$ only:

$$\mathcal{L}(S) = -\frac{1}{2}\Big(\sum_{\omega,t} \nu_{\omega,t}^{(1)}|S_{\omega,t}|^2 + \sum_{\omega,t} \nu_{\omega,t}^{(2)}|X_{\omega,t} - S_{\omega,t}|^2\Big) + C(\nu^{(1)}, \nu^{(2)}) \ , \qquad (1)$$

where C is a constant depending only on $\nu^{(1)}$, $\nu^{(2)}$. Introducing the classical Wiener filtering estimate for S_1,

$$\hat{S}_{\omega,t} = \frac{\nu^{(2)}_{\omega,t}}{\nu^{(1)}_{\omega,t} + \nu^{(2)}_{\omega,t}} X_{\omega,t} \ , \tag{2}$$

the ML problem can be reformulated as the minimization of the objective function

$$\psi(S) = \sum_{\omega,t} \alpha_{\omega,t} |S_{\omega,t} - \hat{S}_{\omega,t}|^2, \ \text{where} \ \alpha_{\omega,t} = \nu^{(1)}_{\omega,t} + \nu^{(2)}_{\omega,t} \ . \tag{3}$$

2.2   Wiener filtering with consistency constraint

If no further constraint is assumed on S, the objective function is obviously minimized for $S = \hat{S}$. However, we need to keep in mind that the STFT is a redundant representation with a particular structure. Denoting by N the number of frequency bins and T the number of frames, STFT spectrograms of time-domain signals are elements of $\mathbb{C}^{NT}$, which we shall call "consistent spectrograms", but one of the fundamental points of this paper is that not all elements of $\mathbb{C}^{NT}$ can be obtained as such [6, 3]. If we assume that inverse STFT is performed in such a way that a signal can be exactly reconstructed from its spectrogram through inverse STFT, then we showed in [3] that a necessary and sufficient condition for an array W to be a consistent spectrogram is for it to be equal to the STFT of its inverse STFT. The set of consistent spectrograms can thus be described as the null space $\text{Ker}(\mathcal{F})$ of the $\mathbb{R}$-linear operator $\mathcal{F}$ from $\mathbb{C}^{NT}$ to itself defined by

$$\mathcal{F}(W) = \mathcal{G}(W) - W, \ \text{where} \ \mathcal{G}(W) = \text{STFT(iSTFT}(W)) \ . \tag{4}$$

Going back to the Wiener filtering problem, if we now impose that the solution be consistent, the problem amounts to finding a consistent spectrogram S minimizing ψ, or in other words to minimize ψ under the constraint that $\mathcal{F}(S) = 0$. Imposing consistency is not a mere elegance or theory-oriented concern, but a truly fundamental problem. Indeed, the spectrogram of the signal resynthesized from the classical Wiener filter spectrogram $\hat{S}$ is actually different in general from $\hat{S}$, and no longer maximizing the Wiener log-likelihood (or minimizing ψ), so that the final result of the processing that we are listening to is in fact not the optimal solution. What we really want to do is to find a signal such that its spectrogram minimizes the Wiener criterion ψ, or, formulating this in the time-frequency domain, to minimize the following "true" objective function

$$\tilde{\psi}(S) = \sum_{\omega,t} \alpha_{\omega,t} \big| \mathcal{G}(S)_{\omega,t} - \hat{S}_{\omega,t} \big|^2 \ , \tag{5}$$

where $\mathcal{G}(S)$ is again the spectrogram of the signal resynthesized from S by inverse STFT. We can try to solve the problem directly in the time domain by minimizing $\psi(\text{STFT}(s))$ w.r.t. the time-domain signal s. Another possibility is to relax the consistency constraint by introducing it as a penalty function: if the weight of the penalty is chosen sufficiently large, or is increased during the course of the optimization, the estimated spectrogram should finally be both consistent and minimizing ψ among the consistent spectrograms.

3   Optimization algorithms

3.1   Time-domain formulation

The consistent Wiener filtering optimization problem amounts to minimizing $\sum_{\omega,t} \alpha_{\omega,t} |S_{\omega,t} - \hat{S}_{\omega,t}|^2$ on the subspace of consistent spectrograms, while that of estimating the signal whose STFT spectrogram is closest to the modified STFT spectrogram $\hat{S}$ amounts to minimizing $\sum_{\omega,t} |S_{\omega,t} - \hat{S}_{\omega,t}|^2$ on the same subspace [6]. The latter problem can be transformed through Parseval's theorem into the minimization of a simple quadratic form on the time signal parameters, but the weights α make here the computation of the optimal signal cumbersome as they hinder us from simplifying the product of the Fourier matrix and its transpose. Let A_t be the $N \times N$ diagonal matrix with diagonal coefficients $\alpha_{\omega,t}$, F the $N \times N$ Fourier transform matrix, w_t the $N \times L$ matrix which computes the t-th windowed frame of the signal x (of length L), and $\hat{s}_t$ the inverse transform of the t-th STFT frame of $\hat{S}$. We can show that the optimal signal x is given by

$$\hat{x} = \Big(\sum_t w_t^H F^H A_t F w_t \Big)^{-1} \sum_t w_t^H F^H A_t F \hat{s}_t \ . \tag{6}$$

If A_t were not present, as in the latter problem, then $F^H F$ would simplify to $N\mathrm{Id}$ and we would get the simple weighted overlap-add estimation $x = \sum_t w_t^H \hat{s}_t / \sum_t w_t^H w_t$. However, the simplification cannot be done here, leading to a very large $(L \times L)$ matrix inversion problem. Still, this matrix is band-diagonal (and Hermitian), and solving the system is possible in a reasonable amount of time and using a reasonable amount of memory space. To reduce in particular the memory requirements, we can split in practice the estimation of the signal on overlapping blocks of a few frames, and reconstruct an approximate solution on the whole interval by overlap-add from the locally optimal signals.

3.2   Consistency as a penalty function

For an array of complex numbers $W \in \mathbb{C}^{NT}$, $\mathcal{F}(W)$ represents the relation between W and the STFT of its inverse STFT. Instead of enforcing consistency through the "hard" constraint $\mathcal{F}(W) = 0$, which may be difficult to handle, we can relax that constraint by using any vector norm of $\mathcal{F}(W)$ to derive a numerical criterion quantifying how far W is from being consistent. We consider here the L^2 norm of $\mathcal{F}(W)$, which leads, as shown in [3], to a criterion related to that used by Griffin and Lim to derive their iterative STFT algorithm [6]. Introducing the consistency penalty in (3), the new objective function to minimize reads

$$\psi_\gamma(S) = \psi(S) + \gamma \sum_{\omega,t} \big| \mathcal{G}(S)_{\omega,t} - S_{\omega,t} \big|^2 \ . \tag{7}$$

An efficient optimization algorithm for ψ_γ can be derived through the auxiliary function method [7]. A function $\psi_\gamma^+(S, \bar{S})$ is called an auxiliary function for $\psi_\gamma(S)$ and $\bar{S}$ an auxiliary variable if $\psi_\gamma(S) = \min_{\bar{S}} \psi_\gamma^+(S, \bar{S})$, $\forall S$. The minimization of ψ_γ can be performed indirectly by alternately minimizing ψ_γ^+ w.r.t. S and $\bar{S}$.

  If we assume, as we shall do, that the inverse STFT is performed using the windowed overlap-add procedure with the synthesis window before normalization

equal to the analysis window, it results from [6] that $\mathcal{G}(S)$ is the closest consistent spectrogram to S in a least-squares sense:

$$\sum_{\omega,t} \left| \mathcal{G}(S)_{\omega,t} - S_{\omega,t} \right|^2 = \min_{\bar{S} \in \mathrm{Ker}(\mathcal{F})} \sum_{\omega,t} \left| \bar{S}_{\omega,t} - S_{\omega,t} \right|^2, \ \forall S \ . \tag{8}$$

If we now define the function $\psi_\gamma^+ : \mathbb{C}^{NT} \times \mathrm{Ker}(\mathcal{F}) \to \mathbb{R}$ such that

$$\forall S \in \mathbb{C}^{NT}, \forall \bar{S} \in \mathrm{Ker}(\mathcal{F}), \ \psi_\gamma^+(S, \bar{S}) = \psi(S) + \gamma \sum_{\omega,t} \left| S_{\omega,t} - \bar{S}_{\omega,t} \right|^2 \ , \tag{9}$$

we easily see from (8) that ψ_γ^+ is an auxiliary function for ψ_γ. This leads to an iterative optimization scheme in which, starting at step p from a spectrogram $S^{(p)}$, $\bar{S}$ is first updated to $\mathcal{G}(S^{(p)})$, and the new estimate $S^{(p+1)}$ is simply estimated as the minimum of a second-order form with diagonal coefficients, altogether resulting in the following update equation:

$$S_{\omega,t}^{(p+1)} \leftarrow \frac{\alpha_{\omega,t} \hat{S}_{\omega,t} + \gamma \mathcal{G}(S_{\omega,t}^{(p)})}{\alpha_{\omega,t} + \gamma} \ . \tag{10}$$

4 Experimental evaluation

4.1 Settings and implementations

The sampling rate was 16 kHz. All spectrograms were built with a frame length $N = 1024$, a frame shift $R = 512$ and a sine window for analysis and synthesis.

The time-domain method was implemented as follows: the analytical solution is computed separately on blocks of 64 STFT frames; the blocks have a 50 % overlap, and the resulting short-time signals are cross-faded on a small region (here 16 frames) around the center of the overlap regions in order to discard portions of signal near the block boundaries, expected to suffer from boundary effects. The above values for the block size and the amount of overlap and cross-fade were determined experimentally so as to minimize computation and memory costs while still obtaining solutions with a true Wiener criterion very close to that of the analytical solution computed on the whole interval.

For the penalty-based algorithm, heuristically, the larger γ, the slower the convergence, but the better the solution. We noticed experimentally that $\tilde{\psi}$ monotonically decreased through the update (10) with γ fixed when starting from a point obtained through updates with a smaller γ. We thus designed an update scheme for γ: starting from a small value γ_0 (typically 10^{-5}) for γ, we update S through (10) while slightly increasing γ by δ (initially set to γ_0 as well) until the decrease of $\tilde{\psi}$ becomes slower than 1 %, in which case we update δ to 2δ and restart the process. We stop after two increases of δ without significant improvement of $\tilde{\psi}$, which typically occurred after around 200 iterations.

4.2 Speech separation under oracle conditions

We evaluate here the performance of the proposed methods for the separation of 10 mixtures of two speakers under oracle conditions, i.e., assuming that the true

Table 1. *Performance comparison for speech separation under oracle conditions*

Method	SDR	ISR	SIR	SAR	$\tilde{\psi}(S)$	Time
Wiener	15.0 dB	25.0 dB	24.6 dB	15.6 dB	2.0×10^9	0.05 s
Griffin-Lim	11.4 dB	21.9 dB	27.6 dB	11.4 dB	6.8×10^{12}	18.1 s
Time domain	17.1 dB	28.2 dB	27.5 dB	17.7 dB	6.2×10^4	1423.0 s
Penalty	16.5 dB	27.1 dB	26.7 dB	17.1 dB	2.2×10^5	8.1 s

power spectrograms of both sources are known. The speech signals were taken from the BSS Oracle Toolbox data [8], downsampled to 16 kHz and downmixed to mono before being mixed together to obtain 12 s long 0 dB Signal to Distortion Ratio (SDR) mixtures. For comparison, we also give the results for the classical Wiener filter output $\hat{S}$ ("Wiener") and for the spectrogram whose magnitude is closest to the magnitude of the classical Wiener filter, computed through Griffin and Lim's iterative STFT algorithm [6] run for 400 iterations ("Griffin-Lim"). This way of obtaining a consistent spectrogram through post-processing of the classical Wiener filter magnitude seems indeed a natural method to attempt.

The results are summarized in Table 1. For each method are reported four commonly used objective source separation performance criteria [9], namely the Signal to Distortion ratio, the source Image to Spatial distortion Ratio (ISR), the Signal to Interference Ratio (SIR) and the Signal to Artifacts Ratio (SAR), as well as the computation time and the final value of the "true" Wiener criterion $\tilde{\psi}$. Although the performance of the classical Wiener filter is already very good, with 15.0 dB output SDR, we can see that the proposed methods all lead to significant improvements in both the true Wiener criterion $\tilde{\psi}$ and the objective performance criteria, with in particular the output SDR raised to 17.1 dB for the time-domain method and 16.5 dB for the penalty-based one, while simply reconstructing the phase as a post-processing does not solve the problem (higher $\tilde{\psi}$, lower SDR). The increase in SDR may not seem straightforward, but it can be understood as a result of the fact that with our methods the spectrogram of the resynthesized signal is closer to the intended ML solution. Computation of the analytical time-domain solution is very costly, but enables us to see that the solution obtained in much less time with the penalty-based algorithm is close to optimal. We will use this last algorithm for the noise reduction experiments below.

We also studied the influence of the frame shift on performance, and noticed that the output SDR increases with the amount of overlap between frames in the STFT, especially for the analytical solution. This could be expected as consistency constraints become stronger when overlap increases. Computation time of course increases as well, roughly linearly with the total number of spectrogram frames for all the methods. Detailed results are skipped due to space constraints.

4.3   Real-world background noise reduction

In order to test our method in more realistic conditions, we performed noise reduction experiments on speech mixed with various types of noise, assuming that only the average power spectrum of the noise is known and that the power spec-

Table 2. *Performance comparison for noise reduction. Values are in dB.*

Input SDR		-10 dB				0 dB				+10 dB			
		SDR	ISR	SIR	SAR	SDR	ISR	SIR	SAR	SDR	ISR	SIR	SAR
Ga	Wiener	-3.1	12.9	-3.2	6.1	6.2	20.3	7.3	12.0	14.8	29.0	16.5	20.0
	Penalty	3.4	7.4	11.6	2.5	9.4	15.6	18.8	10.2	15.9	24.7	23.9	17.1
Su	Wiener	-5.9	18.2	-5.4	6.5	3.9	26.9	4.9	11.5	13.6	34.9	14.8	20.3
	Penalty	-2.8	13.6	-0.7	2.3	6.7	23.5	9.8	10.0	15.6	31.9	19.0	19.0
Sq	Wiener	-4.6	14.6	-4.4	5.7	5.1	23.6	6.3	11.5	14.6	32.5	16.0	20.4
	Penalty	-1.7	9.5	0.1	1.0	7.1	17.8	11.4	9.2	15.6	27.1	20.3	18.1
Ca	Wiener	-4.8	9.7	-5.9	5.6	4.7	18.5	5.4	11.0	13.8	29.1	15.1	19.7
	Penalty	-1.0	4.5	-0.6	-0.9	6.1	12.2	11.2	7.2	14.0	24.0	19.2	16.0

trogram of speech is estimated by spectral subtraction [5]. We considered four noise signals: a synthetic Gaussian white noise, and three real-world background noises from SiSEC 2010's "Source separation in the presence of real-world background noise" task [10] recorded near a subway car ("Su"), on a square ("Sq") and in a cafeteria ("Ca"). The stereo signals were downmixed to mono and cut to 10 s length to match the speech signals, which were the same as above. We considered 30 mixtures for each noise, with 10 different speech signals and at three input SDRs: -10 dB, 0 dB and 10 dB. The results for the penalty-based algorithm and the classical Wiener filter, averaged for each noise and input SDR on the 10 corresponding mixtures, are summarized in Table 2.

We can see that the proposed method leads to a significant improvement over Wiener filtering in terms of output SDR, with, averaged on all noises, further gains of 4.1 dB, 2.4 dB and 1.1 dB respectively for -10 dB, 0 dB and 10 dB input SDRs. This can be further analyzed as a strong improvement of the SIR, offset, to a lesser extent, by a deterioration of the SAR and ISR. Note that this trade-off between improvement of SIR and deterioration of SAR and ISR can be tuned through the penalty weight γ depending on the application, as classical Wiener filtering indeed corresponds to $\gamma = 0$. Perceptually, although there remains some musical noise, the residual noise present in the Wiener filter estimates is much weaker with the proposed method.

We believe that the tendency of our algorithm to further suppress the interference compared with the classical Wiener filter is related to the distribution of the time-frequency bins whose power has not been canceled through spectral subtraction. This can be simply understood in the particular case where speech is replaced by silence. If most bins are set to zero, our algorithm will tend to cancel the remaining ones as well, as a consistent solution with most bins equal to zero in a given neighborhood is likely to be zero on the whole neighborhood, an effect similar to block-thresholding [11], shown to be one of the most effective denoising methods to date. This is first confirmed by the fact that our algorithm seems to perform quite well on Gaussian noise, whose power is exponentially distributed and for which 63 % of the bins are thus set to zero when subtracting the mean power. We tested this hypothesis informally by looking at synthetic noises with various power distribution: the improvements of our algorithm over

classical Wiener filtering decreased as the proportion of bins above the mean power increased, although we shall skip the details here for the sake of brevity.

Finally, we note that by comparing the spectral subtraction results, obtained with a rather crude estimate for the noise power spectrum, with the oracle ones, we can expect our method's performance to depend on the reliability of the power spectrum estimates.

5   Conclusion

We presented a new framework for Wiener filtering and more generally time-frequency masking which takes into account the consistency of spectrograms to compute the true optimal solution to the Wiener filtering problem. We presented two methods to find optimal or near optimal solutions, investigated their performance in comparison with previous works, and showed in particular that our method combined with spectral subtraction outperforms classical Wiener filtering. Future works include combining our method with more sophisticated algorithms for the estimation of the noise power spectrum, and extension of the framework to the multichannel case.

References

1. E. J. Diethorn, "Subband noise reduction methods for speech enhancement," in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, Y. Huang and J. Benesty, Eds.   Kluwer, 2004, pp. 91–115.
2. E. Vincent, M. G. Jafari, S. A. Abdallah, M. D. Plumbley, and M. E. Davies, "Probabilistic modeling paradigms for audio source separation," in *Machine Audition: Principles, Algorithms and Systems*.   IGI Global, to appear.
3. J. Le Roux, N. Ono, and S. Sagayama, "Explicit consistency constraints for STFT spectrograms and their application to phase reconstruction," in *Proc. SAPA*, Sep. 2008, pp. 23–28.
4. J. Le Roux, H. Kameoka, E. Vincent, N. Ono, K. Kashino, and S. Sagayama, "Complex NMF under spectrogram consistency constraints," in *Proc. ASJ Autumn Meeting*, no. 2-4-5, Sep. 2009.
5. S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. ASSP*, vol. 27, pp. 113–120, Apr. 1979.
6. D. W. Griffin and J. S. Lim, "Signal estimation from modified short-time Fourier transform," *IEEE Trans. ASSP*, vol. 32, no. 2, pp. 236–243, Apr. 1984.
7. D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Proc. NIPS*2000*.   The MIT Press, 2001, pp. 556–562.
8. E. Vincent, R. Gribonval, and M. D. Plumbley, "BSS Oracle Toolbox Version 2.1," http://bass-db.gforge.inria.fr/bssoracle/.
9. E. Vincent, H. Sawada, P. Bofill, S. Makino, and J. P. Rosca, "First stereo audio source separation evaluation campaign: Data, algorithms and results," in *Proc. ICA*, Sep. 2007, pp. 552–559.
10. S. Araki, A. Ozerov, V. Gowreesunker, H. Sawada, F. Theis, G. Nolte, D. Lutter, and N. Q. Duong, "The 2010 signal separation evaluation campaign (SiSEC2010) –Part II–: Audio source separation challenges," in *Proc. LVA/ICA*, 2010.
11. G. Yu, S. Mallat, and E. Bacry, "Audio denoising by time-frequency block thresholding," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1830–1839, May 2008.