Denoising of Audio Data by Nonlinear Diffusion

Martin Welk, Achim Bergmeister, and Joachim Weickert

Mathematical Image Analysis Group
Faculty of Mathematics and Computer Science, Bldg. 27
Saarland University, 66041 Saarbrücken, Germany
{welk,bergmeister,weickert}@mia.uni-saarland.de
http://www.mia.uni-saarland.de

Abstract. Nonlinear diffusion has long proven its capability for discontinuity-preserving removal of noise in image processing. We investigate the possibility to employ diffusion ideas for the denoising of audio signals. An important difference between image and audio signals is which parts of the signal are considered as useful information and noise. While small-scale oscillations in visual images are noise, they encode essential information in audio data. To adapt diffusion to this setting, we apply it to the coefficients of a wavelet decomposition instead of the audio samples themselves. Experiments demonstrate that the denoising results are surprisingly good in view of the simplicity of our approach. Nonlinear diffusion promises therefore to become a powerful tool also in audio signal processing.

1 Introduction

Degradation of signals by noise is a ubiquitous phenomenon. In practically any field of signal processing the removal of noise therefore is a key problem. In the field of image processing, diffusion processes are among the most effective and theoretically best understood denoising techniques [18]. While linear diffusion is highly effective in removing noise, it blurs indiscriminately all image information and therefore removes, or at least severely degrades, important image features such as edges along with the noise. Nonlinear diffusion processes – isotropic as well as anisotropic – have therefore gained increasing interest in the last 15 years [14, 2, 18, 17]. They allow to treat details of different size and contrast differently. Thus they enable the design of image filters which remove noise effectively while at the same time edges are preserved and in some cases even enhanced.

The question therefore arises naturally whether diffusion filters can also be used to denoise other classes of signals. Audio signals are one class of signals which is of similar importance as images. We want therefore to investigate the possibility of denoising digital audio signals by diffusion processes.

First of all, audio signals are one-dimensional, so the range of applicable techniques is constrained to linear and isotropic nonlinear diffusion. Another observation is that music samples, like images, contain well-localised features that should be preserved in the denoising process. It can therefore be expected that a good filtering process should again be inhomogeneous and therefore nonlinear.

Direct application of diffusion to sampled audio signals faces a serious problem. Typically, small oscillatory details are the first structures that a diffusion filter removes from a signal. This is well appropriate for image processing; however, in audio signals the most important features consist of oscillations. It needs therefore a re-consideration which features in an audio signal are typically noise and which are useful information.

Targeting at the denoising of sampled music or speech, we see that useful information basically should consist of periodic oscillations with only a moderate number of different frequencies occurring at the same time while noise is supposed to be made up of irregular oscillations which are not concentrated at single frequencies. It seems therefore reasonable to separate useful signal components from noise by a suitable frequency analysis method. The necessity to keep signal components well localised in time motivates us to prefer wavelet decomposition [5, 13] over Fourier analysis.

An established technique for denoising of data in wavelet representations is wavelet shrinkage [6, 7]. Though it is by far not an optimal denoising technique for audio data, it serves for us as a reference because of its simplicity. We will compare the denoising results of our diffusion-based methods with those of wavelet shrinkage working on the same wavelet representations.

A signal restoration approach that is related to ours since it manipulates wavelet coefficients using variational ideas is described in [4]. Further approaches which combine variational and wavelet ideas in a different manner to denoise signals and images can be found in [1, 8, 12].

Section 2 gives an outline of the wavelet methods that will be needed in this paper. In Section 3 we introduce diffusion processes for wavelet coefficients. Application of these filters to digital audio data is illustrated by experiments in Section 4 which are discussed quantitatively and qualitatively. A summary and outlook in Section 5 conclude the paper.

2 Signal processing with wavelets

Wavelet methods in signal processing rely on the representation of a signal f with respect to a basis consisting of scaling functions φ_i^j and wavelet functions ψ_i^j. All scaling and wavelet functions are shifted and dilated versions of one scaling function φ with low-pass characteristics and one wavelet function ψ with band-pass properties, i.e.

$$\varphi_i^j(z) = 2^{-j/2}\varphi(2^{-j}z - i), \qquad \psi_i^j(z) = 2^{-j/2}\psi(2^{-j}z - i)\ .$$

If the wavelet functions ψ_i^j and scaling functions $\varphi_i^{j_0}$ form an orthonormal basis, we have

$$f(z) = \sum_{i\in\mathbb{Z}} c_i^{j_0}\varphi_i^{j_0}(z) + \sum_{j=-\infty}^{j_0}\sum_{i\in\mathbb{Z}} d_i^j\psi_i^j(z)$$

where $c_i^j := \langle f, \varphi_i^j\rangle$, $d_i^j := \langle f, \psi_i^j\rangle$ with $\langle\cdot,\cdot\rangle$ being the $L_2(\mathbb{R})$ inner product. Note that this representation uses scaling functions only on the coarsest level j_0 while wavelet functions of level j_0 and all finer levels are used. For a discrete signal, represented by a function f which is constant on $[k, k+1)$ for every integer k, only wavelet levels $j \geq 1$ occur.

2.1 Haar wavelet representations

Let us consider the simplest wavelet, the *Haar wavelet* [10] $\psi(z) = \chi_{[0,\frac{1}{2})}(z) - \chi_{[\frac{1}{2},1)}(z)$ with corresponding scaling function $\varphi(z) = \chi_{[0,1)}$, where χ_I is the characteristic function of the interval I. Then, wavelet analysis and synthesis can be efficiently carried out via the *two-scale relations*

$$c_i^j = \frac{c_{2i}^{j-1} + c_{2i+1}^{j-1}}{\sqrt{2}}, \qquad d_i^j = \frac{c_{2i}^{j-1} - c_{2i+1}^{j-1}}{\sqrt{2}}.$$

For a discrete signal as described above, the c_i^0 equal the signal samples.

A wavelet representation of this kind, called *decimated* wavelet decomposition, constitutes a hierarchical subdivision of the domain of definition of the signal, thereby bearing the clear disadvantage of lacking translation invariance. A simple but effective way to overcome this problem in case of a discrete signal of finite length is the *cycle-spinning* procedure [3]. Cycle-spinning essentially means that the decimated wavelet analysis and synthesis of a signal of length N is carried out N times: for the original signal and all $N-1$ possible cyclically shifted copies of it. The analysis step therefore yields N redundant wavelet representations encoding the same signal. In the synthesis step, N concurrent signals are generated. However, if the reconstruction is done with processed wavelet coefficients, these N signals will in general no longer coincide. The final reconstruction result is therefore obtained by averaging the concurrent reconstructions. Filters designed with these transforms are shift-invariant by construction.

2.2 Soft wavelet shrinkage

The processing of a signal is typically performed in three steps. First, the *analysis* step in which the given signal is transformed into wavelet representation; second, some operation on the wavelet coefficients; third, the *synthesis* in which the modified coefficients are used to reconstruct the processed version of the signal.

One class of denoising methods widely studied in literature are *wavelet shrinkage* procedures. *Soft wavelet shrinkage* applies the shrinkage function

$$S_\theta(y) = \begin{cases} y - \theta \operatorname{sgn}(y), & |y| > \theta, \\ 0, & |y| \leq \theta \end{cases}$$

to the wavelet coefficients after the analysis step. The modified coefficients $\tilde{d}_i^j := S_\theta(d_i^j)$ are used with the unchanged scaling coefficients c_i^j in the synthesis step.

In [16] it was shown that the denoising quality of shift-invariant soft Haar wavelet shrinkage is improved by using the level-dependent shrinkage parameter $\theta_j = 2^{-j/2}\theta_0$ on level j instead of one uniform parameter θ.

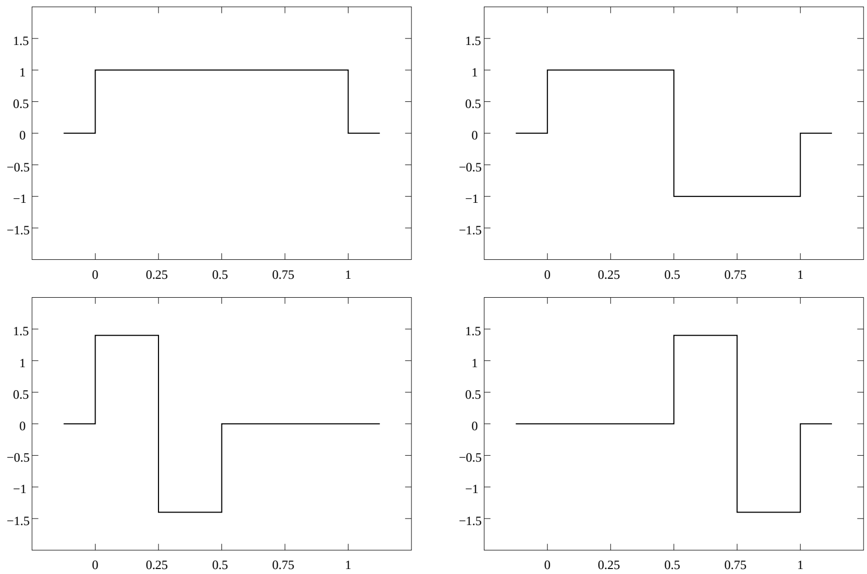

Fig. 1. Basis for a two-level decimated Haar wavelet decomposition. **Top left:** Scaling function φ_0^1. **Top right:** Wavelet function ψ_0^1. **Bottom left:** Wavelet function ψ_0^2. **Bottom right:** Wavelet function ψ_1^2.

2.3 Daubechies wavelets

In image processing, excellent denoising results can be achieved using Haar wavelets, particularly its shift-invariant modification. However, for audio signal processing Haar wavelets are often considered insufficient because they have only one vanishing moment. Daubechies wavelets [5] are often suggested as a better choice in this context. They can be constructed with arbitrarily many vanishing moments. In every case, their wavelet and scaling functions can be found via a recursive procedure. For the simplest Daubechies wavelet (Daubechies-4), one has

$$\varphi(1) = \frac{1+\sqrt{3}}{2}, \qquad \varphi(2) = \frac{1-\sqrt{3}}{2}, \qquad \varphi(k) = 0,\ k \in \mathbb{Z} \setminus \{1,2\},$$

$$\begin{aligned}\varphi(z) = \frac{1}{4}\Big(&(1+\sqrt{3})\varphi(2z) + (3+\sqrt{3})\varphi(2z-1) \\ &+ (3-\sqrt{3})\varphi(2z-2) + (1-\sqrt{3})\varphi(2z-3)\Big) \\ \psi(z) = \frac{1}{4}\Big(&-(1+\sqrt{3})\varphi(2z) + (3+\sqrt{3})\varphi(2z-1) \\ &- (3-\sqrt{3})\varphi(2z-2) + (1-\sqrt{3})\varphi(2z-3)\Big)\end{aligned}$$

for all $z = 2^{-j}k$, with integers j, k, and by continuity on the whole real line (see [5, 11]).

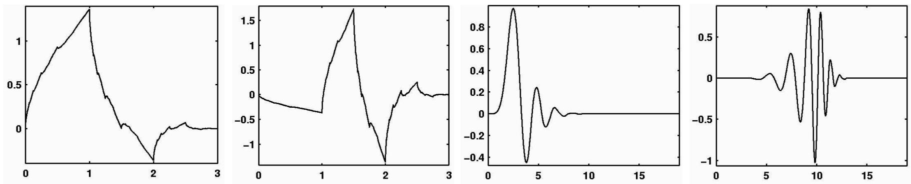

Fig. 2. Left to right: Scaling function for Daubechies-4 wavelet. – Corresponding wavelet function. – Scaling function for Daubechies-20 wavelet. – Corresponding wavelet function. Generated using MATLAB with Wavelet toolbox.

Scaling and wavelet functions of the Daubechies-4 and Daubechies-20 wavelets used in our experiments are shown in Fig. 2.

3 Diffusion of wavelet coefficients

Let a one-dimensional signal $s(z)$ be given. The partial differential equation

$$v_t = \partial_z(g(v_z^2)\, v_z), \qquad (z,t) \in \mathbb{R} \times (0, +\infty) \tag{1}$$

with initial condition $v(z,0) = s(z)$ describes a one-dimensional diffusion process which embeds the signal $s(z)$ into a family $v(z,t)$ of signals, parametrised with $t \in [0, +\infty)$, which constitute smoothed versions of s. The parameter t can be considered as a diffusion time which, however, must be well distinguished from the signal time z. Here, the *diffusivity* $g(y^2)$ should be a bounded, non-increasing, positive function of its nonnegative argument. For $t \to \infty$, the signal will tend to a constant function. To obtain a smoothed signal, it is therefore necessary to choose a stopping time T which determines the degree of smoothing.

Besides the case $g(y^2) = 1$ of linear diffusion we consider the Perona–Malik diffusivity $g(y^2) = \frac{1}{1+y^2/\lambda^2}$ where λ is a threshold parameter [15]. The purpose of non-linear diffusion with such a decreasing diffusivity is to suppress smoothing at locations with large gradients which are supposed to be edge-like discontinuities holding important information.

A simple explicit discretisation of (1) is given by

$$v_i^{k+1} = v_i^k + \frac{\tau}{h}\left(g((\dot{v}_i^k)^2)\dot{v}_i^k - g((\dot{v}_{i-1}^k)^2)\dot{v}_{i-1}^k\right), \qquad \dot{v}_i^k = \frac{1}{h}(v_{i+1}^k - v_i^k)$$

with step sizes τ for diffusion time h for the signal parameter.

In applying the diffusion equation to the coefficients of a decimated wavelet representation, the coefficients of each wavelet level are considered as one channel, such that diffusion does not transfer amplitudes between different frequency bands. It requires attention that the step size h doubles from each level to the next coarser one. Having this in mind, linear diffusion can be implemented straightforward by

$$[d_i^j]^{k+1} = [d_i^j]^k + \frac{\tau}{2^{2j}}([d_{i+1}^j]^k - 2[d_i^j]^k + [d_{i-1}^j]^k)$$

where $[d_i^j]^k$ denotes the wavelet coefficient d_i^j in the decomposition of the signal in the k-th diffusion-time step, and we have assumed that the temporal resolution in signal time is 1 for the samples, i.e. wavelet level 0.

In nonlinear diffusion of multi-channel signals (like colour images) it is essential that the discontinuities where smoothing is suppressed are localised at equal positions in all channels [9]. To achieve this, one uses a common diffusivity which incorporates gradient information from all channels and steers uniformly the diffusion in all of them. The same argument applies also in our model. Here it is assumed that large differences between neighbouring wavelet coefficients signify boundaries of acoustic events extended in time which should not be blurred.

Consequently, we want to use a common diffusivity also in the nonlinear diffusion of our wavelet coefficients. In computing this common diffusivity in a decimated wavelet representation, it needs to be clarified which neighbour differences should contribute to which diffusivities. Since diffusivities are to steer diffusion *between* neighbouring wavelet coefficients of one level, the proper location where to estimate the diffusivity is the central point between the coordinates of the wavelet coefficients themselves. To determine which neighbour differences should enter a particular diffusivity, we look at the influence zones of the wavelet coefficients, i.e. for each coefficient the group of subsequent samples that it depends on. Then each diffusivity is influenced exactly by the neighbour differences of those wavelet coefficients with influence zones starting or ending at the position of this diffusivity, see Fig. 3. Vice versa, the diffusion between two adjacent wavelet coefficients is regulated only by the diffusivity at the single inter-sample location where the two influence zones meet.

This procedure can directly be motivated as a simple subsampling of the coarser wavelet levels. By writing down each wavelet coefficient $[d_i^j]^k$ of the j-th level 2^{j-1} times, a vector-valued signal with the signal-time resolution of the finest wavelet level 1 arises. The i-th vector in this signal reads

$$([d_i^1]^k, [d^2_{\lfloor i/2 \rfloor}]^k, \dots, [d^{j_0}_{\lfloor i/2^{j_0-1} \rfloor}]^k)^{\mathrm{T}}$$

The diffusivity between the i-th and $(i+1)$-th vector in this signal is

$$g_{i+1} := g\left(\sum_{j=1}^{j_0} \left([d^j_{\lfloor (i+1)/2^j \rfloor}]^k - [d^j_{\lfloor i/2^j \rfloor}]^k \right)^2 \right)$$

where $\lfloor z \rfloor$ denotes the largest integer less or equal z. Here we have weighted all wavelet levels equally; one could also use different weights for the wavelet levels.

After computing the new diffusion time-step, the 2^{j-1} copies of the coefficient $[d_i^j]^k$ are no longer identical; the new value $[d_i^j]^{k+1}$ is then obtained by averaging the concurrent values which amounts exactly to

$$[d_i^j]^{k+1} = [d_i^j]^k + \frac{\tau}{2^{2j}} \left(g_{2^{j-1}(i+1)} \cdot \left([d^j_{i+1}]^k - [d_i^j]^k \right) - g_{2^{j-1}i} \cdot \left([d_i^j]^k - [d^j_{i-1}]^k \right) \right) .$$

In the context of decimated wavelet shrinkage, it can be criticised that diffusivities at different locations have different numbers of influencing coefficient

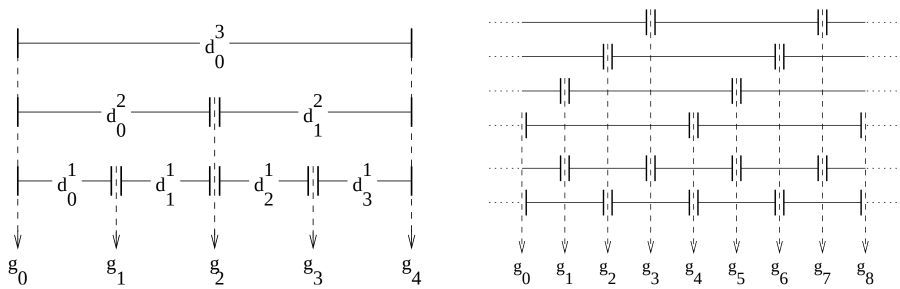

Fig. 3. Influence of differences of neighbouring wavelet coefficients on diffusivities for the coupled nonlinear diffusion. **Left:** Decimated Haar wavelet representation with 3 levels. Wavelet coefficients d_j^k with their influence zones are shown. By $g_0, \ldots, g_4$ subsequent diffusivities are denoted. **Right:** Shift-invariant Haar wavelet representation with 2 levels. Again, $g_0, \ldots, g_8$ are subsequent diffusivities. Variables for wavelet coefficients are omitted.

pairs. One could consider compensation factors to remedy this. On the other hand, as soon as we switch to shift-invariant Haar wavelets, the problem disappears. In this situation, if we count identical wavelet coefficients only once, we have for each inter-sample location exactly one pair of coefficients in each wavelet level whose influence zones meet there; compare Fig. 3.

We emphasise that in the shift-invariant setting only coefficients of one wavelet level which are part of the same decimated wavelet decomposition communicate directly in the diffusion process. These are coefficients which have not only the same frequency but also the same phase. Coefficients of the same level and different phase, as well as those of different levels, belong to different channels which are linked only by the channel coupling.

4 Experimental results

For a first impression of the properties of different wavelet bases in audio processing, we degrade a synthetic 200 Hz sine wave[1] by adding Gaussian noise[2] of 10 % the signal variance, and apply to it soft wavelet shrinkage with non-shift-invariant and shift-invariant Haar wavelets as well as with two Daubechies bases, see Fig. 4. It is evident that the signal shape of the shrinked signal is strongly influenced by the shape of the used wavelets. Audio perception is very sensitive to such details in the wave shape such that the signals denoised with decimated Haar, and Daubechies-4 wavelets do not sound too well. Surprisingly, the quality of shift-invariant Haar wavelet shrinkage is subjectively perceived superior to that of Daubechies-20 shrinkage although the shift-invariant Haar

[1] The sampling frequency of all our audio data is 44,100 Hz.

[2] The choice of Gaussian noise in the experiments presented here is nothing special. Denoising quality was quite similar when, e.g., white noise or recorded noise from technical sources was added.

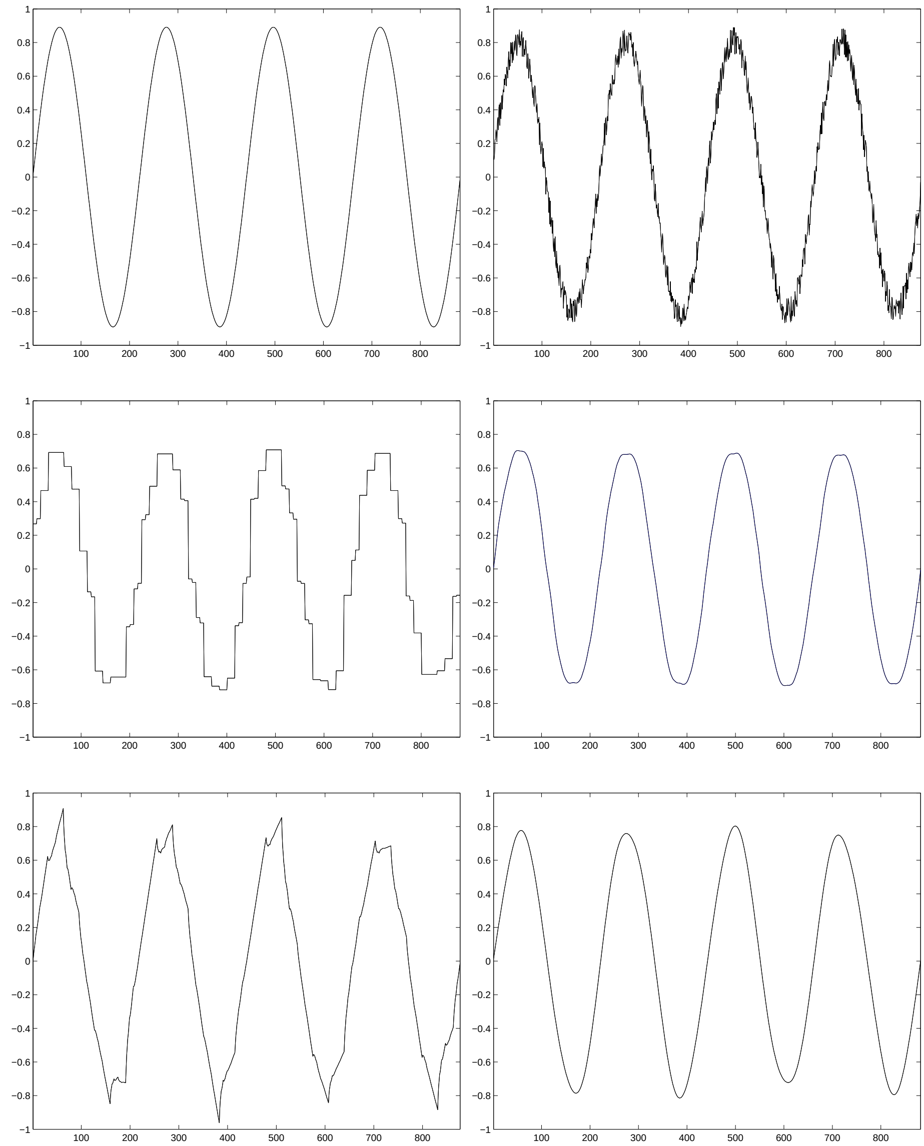

Fig. 4. Soft wavelet shrinkage with different wavelet decompositions, shrinkage parameter always $\theta = 10000$. **Top left:** Original signal (sine wave of 200 Hz). **Top right:** Same with Gaussian noise, noise variance ca. 10 % of signal variance. **Middle left:** Denoised using decimated Haar wavelet shrinkage. **Middle right:** Shift-invariant Haar wavelet shrinkage. **Bottom left:** With Daubechies-4 wavelets. **Bottom right:** Daubechies-20 wavelets.

Table 1. SNR measured for different denoising methods

Denoising method	SNR (dB), drum	SNR (dB) instr.
Wavelet shrinkage, shift-invariant Haar	12.50	13.54
Nonlin. diffusion, decimated Haar	12.57	14.13
Nonlin. diffusion, Daubechies-4	12.81	14.49
Nonlin. diffusion, Daubechies-20	13.01	15.05
Nonlin. diffusion, shift-invariant Haar	12.92	13.04
no denoising (noisy signal)	11.16	13.63

wavelet shrinkage result shows a slight deformation around the peaks. However, the Daubechies-20 denoised sample displays a jitter of amplitude which is indeed an audible perturbation.

In the further experimental validation of our diffusion denoising model, we use two musical signals: First, a short drum loop; second, a clipping of instrumental music (brass accompanied by strings)[3]. We add Gaussian noise to each of the signals. The noise variance is about 10 % of the signal variance for the drum loop and about 5 % for the instrumental piece. Table 1 compiles results of signal-to-noise ratio (SNR) measurements for selected denoising methods. The SNR is computed as

$$\mathrm{SNR}(u, f) = 10 \log_{10} \frac{\mathrm{var}(f)}{\mathrm{var}(f-u)} \,\mathrm{dB}$$

where f is the original and u the noisy signal.

According to subjective perception, our nonlinear diffusion method in general leads to better denoising results than soft wavelet shrinkage. The SNR, however, favours in some cases wavelet shrinkage (or even the noisy signal!) which indicates that it might not be an adequate criterion for denoising quality. A characteristic difference between shrinkage and diffusion denoising is shown in Fig. 5. Looking at the right part of the original signal clipping displayed, one notices small high-frequent oscillations which are visible particularly near the extrema of the low-frequent base oscillation. These components are indeed essential for the characteristic timbre of the drum. By removing these signal components along with the noise, wavelet shrinkage compromises the timbre much more than nonlinear diffusion which keeps at least part of these components.

In the remainder of this section, we discuss qualitatively a few more observations made during our experiments. As to the choice of the wavelet basis, auditory impression as well as, in part, the SNR measurements suggest that introducing shift-invariance into Haar wavelet shrinkage raises the quality to a level comparable that of (decimated) Daubechies wavelets. This is observed both for shrinkage and diffusion algorithms.

The number of wavelet levels which are used in our wavelet diffusion process is less important than it appears. Most denoising is achieved just in the finest five wavelet levels; with ten levels, no significant enhancement is encountered.

[3] Audio samples are available via the first author's web page, `http://www.mia.uni-saarland.de/welk`

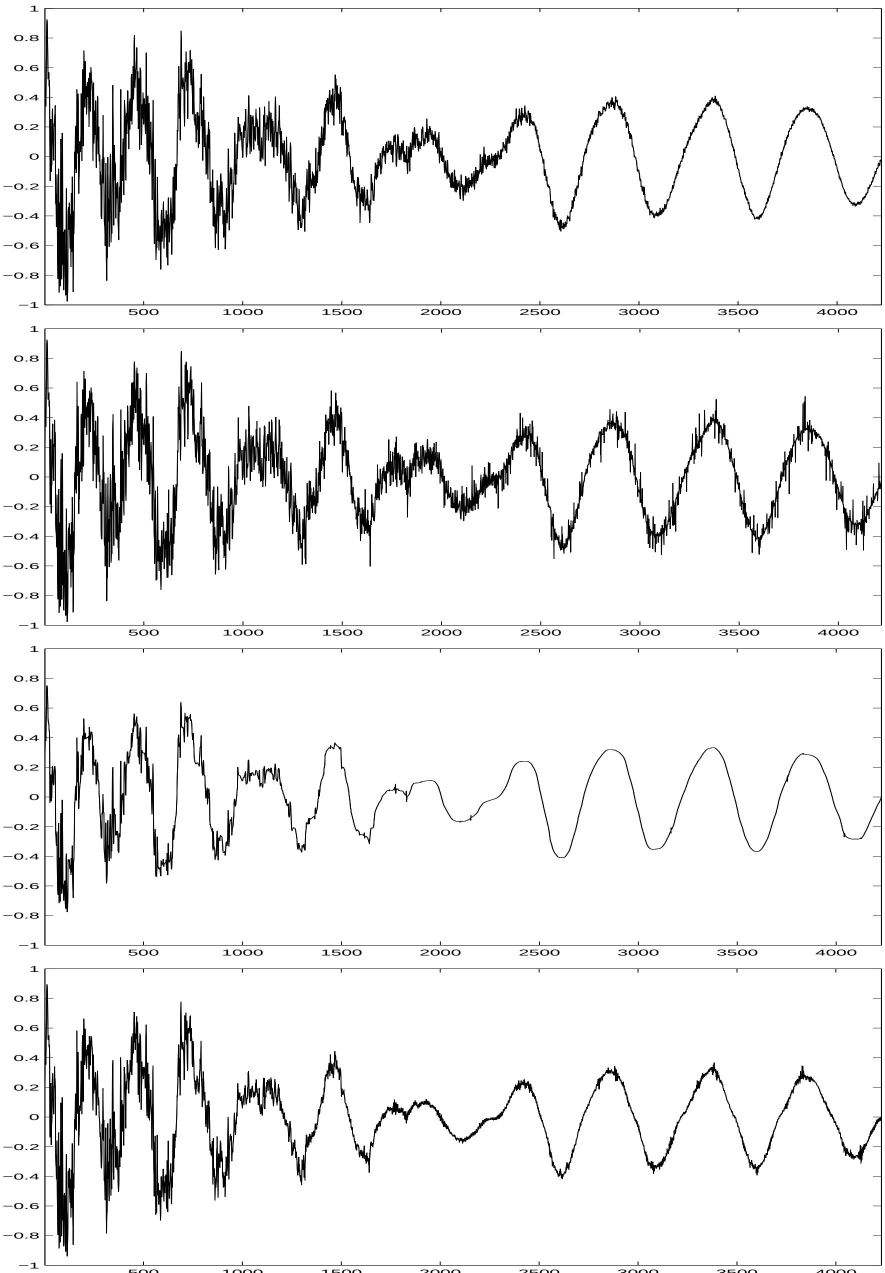

Fig. 5. Denoising of a drum loop signal. Parameters are chosen such that similar degrees of noise reduction are perceived. **Top to bottom:** A drum loop. – Same with Gaussian noise. – Denoised by shift-invariant soft Haar-wavelet shrinkage ($\theta = 5000$). – Denoised by channel-coupled nonlinear diffusion on shift-invariant Haar wavelet decomposition ($t = 100$, $\lambda = 500$, $\sigma = 1.0$).

The reason is that the diffusion process is not very effective on coarse levels because of their low resolution and the very small updates which are therefore made per diffusion-time step. Low-frequent (rumbling) noise should therefore be handled by other measures.

Denoising with Haar wavelet methods often creates a characteristic audible artifact. It consists in a slightly rough-ringing noise of specific pitch which is composed of frequencies standing in octave relations to the sampling frequency. This noise gains intensity the longer diffusion acts or the more coefficients are shrunk by wavelet shrinkage. It is also observed that many signals tend to be flattened even by our nonlinear diffusion process for large t. We assume that these two phenomena are two sides of the same medal: By admitting only transfer between wavelet coefficients of equal frequency and phase, our diffusion process tries to keep signal amplitude in the separated frequency and phase components and to avoid extinction. However, this works perfectly only for frequencies which are subdivisions of the sampling frequency by powers of two (and which therefore in some sense are "in resonance" with the wavelet decomposition). Other frequencies are still weakened during the process, inducing the tendency to flatten signals. On the other hand, even some noise is kept in the resonance frequencies and can be perceived with its pitch as soon as the other frequencies are gone.

In agreement with this reasoning, improvements in the algorithm which lead to a better preservation of signal amplitude reduce the sound artifact at the same time. Starting from a simple linear diffusion process on decimated Haar wavelet coefficients, each of the following steps observably reduces both the diminishing of signal amplitude and the appearance of the artifact tone: first, switching to shift-invariant Haar wavelets; second, making diffusion in each channel nonlinear; and third, establishing of the channel coupling.

5 Summary and outlook

In this paper, we have introduced a method for the denoising of audio signals by nonlinear diffusion. Because of the specifics of audio data compared to images, the diffusion process is not formulated for the digital audio samples but for the wavelet coefficients of a suitable wavelet representation.

Comparisons with wavelet shrinkage techniques reveal a fairly good performance of our method despite its simplicity. Due to the high sensitivity of human auditory perception for even tiny perturbations, the denoising achieved is not satisfactory enough for immediate application Nevertheless, our results clearly indicate that nonlinear diffusion can be successfully adapted to the denoising of audio data.

Ongoing work concentrates on the reduction of artifacts and improvement of the homogeneity of denoising over the frequency range. We are also interested in establishing better quantitative measures for denoising quality that reflect the perceived quality as accurate as possible.

References

1. E. J. Candés and F. Guo. New multiscale transforms, minimum total variation synthesis: Applications to edge-preserving image reconstruction. *Signal Processing*, 82(11):1519–1543, 2002.
2. F. Catté, P.-L. Lions, J.-M. Morel, and T. Coll. Image selective smoothing and edge detection by nonlinear diffusion. *SIAM Journal on Numerical Analysis*, 32:1895–1909, 1992.
3. R. R. Coifman and D. Donoho. Translation invariant denoising. In A. Antoine and G. Oppenheim, editors, *Wavelets in Statistics*, pages 125–150. Springer, New York, 1995.
4. R. R. Coifman and A. Sowa. Combining the calculus of variations and wavelets for image enhancement. *Applied and Computational Harmonic Analysis*, 9(1):1–18, July 2000.
5. I. Daubechies. *Ten Lectures on Wavelets*. SIAM, Philadelphia, 1992.
6. D. L. Donoho. De-noising by soft thresholding. *IEEE Transactions on Information Theory*, 41:613–627, 1995.
7. D. L. Donoho and I. M. Johnstone. Minimax estimation via wavelet shrinkage. *Annals of Statistics*, 26(3):879–921, 1998.
8. S. Durand and J. Froment. Reconstruction of wavelet coefficients using total-variation minimization. *SIAM Journal on Scientific Computing*, 24(5):1754–1767, 2003.
9. G. Gerig, O. Kübler, R. Kikinis, and F. A. Jolesz. Nonlinear anisotropic filtering of MRI data. *IEEE Transactions on Medical Imaging*, 11:221–232, 1992.
10. A. Haar. Zur Theorie der orthogonalen Funktionensysteme. *Mathematische Annalen*, 69:331–371, 1910.
11. A. K. Louis, P. Maass, A. Rieder. *Wavelets: Theory and Applications.* Wiley, New York, 1997.
12. F. Malgouyres. Minimizing the total variation under a general convex constraint for image restoration. *IEEE Transactions on Image Processing*, 11(12):1450–1456, 2002.
13. S. Mallat. *A Wavelet Tour of Signal Processing.* Academic Press, San Diego, second edition, 1999.
14. P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. In *Proc. IEEE Computer Society Workshop on Computer Vision*, pages 16–22, Miami Beach, FL, Nov. 1987. IEEE Computer Society Press.
15. P. Perona and J. Malik. Scale space and edge detection using anisotropic diffusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:629–639, 1990.
16. G. Steidl, J. Weickert, T. Brox, P. Mrázek, and M. Welk. On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and SIDEs. *SIAM Journal on Numerical Analysis*, 42(2):686–713, 2004.
17. J. Weickert. A review of nonlinear diffusion filtering. In B. ter Haar Romeny, L. Florack, J. Koenderink, and M. Viergever, editors, *Scale-Space Theory in Computer Vision*, volume 1252 of *Lecture Notes in Computer Science*, pages 3–28. Springer, Berlin, 1997.
18. J. Weickert. *Anisotropic Diffusion in Image Processing.* Teubner, Stuttgart, 1998.