

# A WAVELET BASED AUDIO DENOISER

Claudia Schremmer<sup>†</sup>, Thomas Haenselmann<sup>†</sup> and Florian Bömers<sup>‡</sup>

<sup>†</sup> Department of Praktische Informatik IV, University of Mannheim, Germany

<sup>†</sup>{schremmer,haenselmann}@informatik.uni-mannheim.de

<sup>‡</sup> florian@bome.com

## ABSTRACT

This article discusses real-time denoising algorithms for digital audio based on the Wavelet Transform. White noise is located in all frequencies and is thus especially hard to detect. We use the locality of the wavelet function to single out the frequency domains of the signal itself and thereby be able to denoise it. Perfect denoising is not possible: the higher the threshold coefficient is set, the more noise is detected, but the more the original signal is affected as well. We have implemented a flexible framework for denoising that includes hard and soft thresholding, different Wavelet Transforms, and different treatment of the padding coefficients. The presented denoiser is a real-time application that allows direct subjective judgments of a parameter setting. An evaluation will conclude the article.

**Keywords.** Denoising, Wavelet, Digital Audio, Evaluation.

## 1. INTRODUCTION

Applications of signal processing all struggle with a major problem, *noise*. A pure and undisturbed signal is superimposed by another – unwanted – signal. How to separate the one from the other without deterioration of the signal itself? This question accompanies the search for a good representation of the signal throughout its encoding process. Though the Fourier Transform (FT) and the Windowed Fourier Transform are successfully used, new methods focus on the Wavelet Transform (WT) in order to overcome the FT's disadvantages. Its construction through multiresolution analysis proves to reflect the frequency resolution of the human ear: lower frequencies are resolved well, while high frequencies are only loosely resolved. Furthermore, the implementation of the WT is fast enough to allow real-time applications.

This article discusses the removal of white noise in digital audio. Two major approaches exist to detect and remove white noise, each with unwanted side-effects. For a given

noise level, an optimal denoising level can be calculated, resulting in minimal artifacts of the denoised signal. We use a real-time implementation of an audio tool to demonstrate the ideas of noise and noise removal. Furthermore, we evaluate the subjective parameter setting in the denoising process against the automatic setting for minimal artifacts.

The article is organized as follows: after discussing some related work in Section 2, we present the basic concepts of digital audio, noise, wavelet theory, and filter banks in Section 3. In Section 4, we present a framework for digital audio processing in real-time. Section 5 details an evaluation and comparison of automatic and subjective parameter settings. Section 6 concludes the article.

## 2. RELATED WORK

Much research on reduction of the noise in cassette recordings has been carried out by Dolby [1]. Dolby noise reduction is based on reducing the perception of *tape hiss*, which comes from the magnetic particles on a tape. The idea is that music be encoded just before recording. Here, the level of soft, high-frequency passages is raised to make them louder than the tape's noise, while loud passages are not altered.

An unusual project has been carried out at the Yale School of Music [2]. A wavelet denoiser has been used to restore a battered recording of Brahms playing his own work in 1889. The original wax recording is one of the rare opportunities to analyze a major composer's interpretation of his own work. Even if the wavelet-restored sound is not yet musically pleasant, enough of the recording has been reconstructed to prove that Brahms took considerable liberties with his own score.

A formal approach to denoising using wavelet analysis was introduced by Donoho [3], who has developed two major wavelet-based algorithms. *Linear denoising* assumes noise to consist of high-frequency components. The corresponding (fine) scales of the Wavelet Transform are set to zero. *Non-linear denoising* or *wavelet shrinkage* assumes noise to consist also of low energy, thus existing in the coarse scales as well. The algorithms presented in this article are based partly on Donoho's work.

---

This work was supported by the Bundesministerium für Bildung und Forschung (BMBF), Germany.

### 3. NOISE IN DIGITAL AUDIO PROCESSING

What the human ear perceives as *sound* is physically small (analog) changes in air pressure that stimulate the eardrum. A digital system can handle only discrete signals though. The conversion analog–discrete is realized through sampling. The level of the current is measured at short intervals so that sound is represented by a sequence of discrete values.

#### 3.1. Wavelet Transform and Filter Banks

As a wavelet is a compact function that vanishes outside a certain interval, the WT is specially adapted to analyze local variations like in audio analysis: the WT offers high temporal localization for high frequencies while offering high–frequency resolution for low frequencies [4]. A high–frequency event (e.g. a cymbal crash) will be analyzed by short, and high–amplitude wavelets. Low tones (e.g. a bass drum) will be analyzed by long, low–amplitude wavelets. An important step for the application of wavelet theory in signal processing, the transition from the mathematical theory to filters, has been presented by Mallat [5] through multiresolution. A multiresolution analysis is implemented via high–pass filters, resp. band–pass filter (i.e., wavelets) and low–pass filters (i.e., scaling functions). Low–pass filters let all frequencies pass that are below a cut–off frequency, whereas the remaining frequency components are removed from the signal. High–pass filters work vice versa. In this context, the Wavelet Transform of a signal can be realized with a filter bank via successive application of a 2–channel filter bank consisting of high–pass and low–pass filters: the detail coefficients of every recursion step are kept apart, and the recursion starts again with the remaining approximation coefficients of the transform. As human perception proves to focus on low–frequency stimuli, this recursion of the WT reflects perfectly the subdivision of a signal according to human perception.

The synthesis of a filter bank works just the other way round. The detail and approximation coefficients of each decomposition level are filtered through two synthesis filters, where each filter application gives an approximation coefficient for the next level. The process is repeated until the original signal has been reconstructed.

The implementation of the discrete WT using Mallat’s algorithm is of complexity  $O(n)$ . In spite of that, current implementations require more processor resources than the FFT, depending on the length of the filters. However, on modern computers it is fast enough for real–time analysis and synthesis of audio data.

#### 3.2. Noise and Wavelets

Noise in signal processing denotes a perturbing signal in some or all frequency bands that is generally unwanted. The less regular this noise is, the more sophisticated the methods have to be that are applied to denoise the signal. One distinguishes noise according to its properties in the time and frequency domains.

The notation *white noise* refers to the distribution of the noise in all frequency domains. *Gaussian vs. uniform noise*: this characteristic refers to the probability density in the time domain. Uniform noise has a constant probability density over a finite interval. Gaussian noise is normally distributed and can be defined over an infinite interval by just two parameters, average and spread. Almost all natural phenomena generate Gaussian noise because of the “law of large numbers”.

White noise is especially difficult to detect and to remove as it is located in all frequencies. A (dyadic discrete) WT, however, makes use of the fact that a transformed signal’s coefficients “live” in scales, and that these scales are proportioned similarly to human audio perception. Harmonic content like music is closely correlated, and thus produces bigger coefficients than noise, which is highly uncorrelated. The removal of the small coefficients thus results in noise removal. There are two wavelet–based variants to handle white noise [3][6]:

- *hard thresholding*: coefficients on some or all scales that are below a certain threshold are set to zero,
- *soft thresholding*: coefficients on some or all scales below a threshold are set to zero, and additionally all coefficients above this threshold are reduced by the value of the threshold.

While hard thresholding exhibits artifacts with an increasing threshold, soft thresholding attenuates the range of the wavelet coefficients and smoothens the signal, thus modifying its energy [7].

### 4. IMPLEMENTATION

We have implemented a tool for real–time digital audio processing on a Windows PC that is presented in detail in [8]. Its purpose is to demonstrate the possibilities and power of wavelet–based audio processing algorithms. We highly recommend to download the tool [9] and listen to the modifications of your own voice or a recorded file, because audible stimuli cannot be completely described in written form.

A graphical user interface enables the user to choose an input device and an output device for the audio stream. Any number of filters can be activated in–between. In this work, we concentrate on noise and the removal of noise via wavelet analysis. In Figure 1, the setting is as follows: a differ-

ence listener is started right at the beginning of the processing. After the application of a Wavelet Transform, a noise generator adds white noise to the original audio file. A parameter determines whether the noise distribution is uniform or Gaussian, and the amount of noise can be adjusted between  $-\infty$  dB (no noise) and 0 dB (maximum noise energy). As for the Wavelet Transform, we have implemented the following filters: Haar, various Daubechies, Coiflets, Symlets, Biorthogonal, Battle–Lemarié, and spline filters.

The wavelet denoiser (cf. Figure 2) is applied before the inverse Wavelet Transform synthesizes the audio data. The wavelet denoiser realizes soft and hard thresholding for white noise. Parameters include the type of algorithm, the cut-off threshold, the number of levels to be treated (up to all levels), and whether the padding coefficients of the boundaries are included for thresholding. Thus, the parameters allow flexible control of the denoising process.

Finally, terminating the difference listener helps one to concentrate on the distortion of the signal. In this setup with difference listeners, the output indicates what has been added to the signal/removed from the signal. Complete silence indicates perfect reconstruction: the noise that has been added to the signal has been perfectly removed, and the signal itself has not been modified. Music in the difference listener, however, indicates data that has mistakenly been cut off during the treatment.

Additionally to the audible output, the presented audio tool can also visualize the denoising performed: the application of a time domain display for the noisy signal, and a second time domain display after removal of the noise is demonstrated in Figure 3. This service is especially suited to demonstrate the denoiser when only visual media are available (like this article).

The real-time wavelet denoiser presented in this article aims to apply experience to the concepts above. Can we perceive the difference between uniform and Gaussian white noise in audio? What is the audible difference between hard and soft thresholding? How much noise can be added to a signal without irreversible deterioration? What is the effect of padding coefficients? As human perception is still not totally understood, and models vary strongly, directly *hearing* the result of a parameter setting – and judging it instantly – is still the easiest and most reliable way to get an answer to these questions.

## 5. RESULTS

We present the evaluation of error estimation for audio data “erroneously” removed during the denoising process. In this test, a Battle–Lemarié wavelet filter with 49 coefficients was used. The denoiser was set to denoise the first 5 levels of wavelet coefficients using soft thresholding, including the padding coefficients. The error estimation was taken from

[10], and required the original signal for comparison. When  $E$  denotes the square root energy of the added noise, and  $\hat{E}$  denotes the square root energy of the difference between original and denoised signal, the error estimation is  $\hat{E}/E$ . The relation  $\hat{E}/E < 1$  stands for successful noise removal, whereas  $\hat{E}/E \geq 1$  means that data of the original signal have been removed together with the noise. Table 1 shows the evaluation results for the music file `dnbloop.wav`, which has a wide frequency usage and some short transients. For the objective assessment, the threshold was set to yield a minimum error [11]. For a fixed  $E$ , this turned out to be the minimal  $\hat{E}$ . The subjective assessment revealed the average rating of five probands, where the least noticeable noise in the setting with the difference listener (cf. Section 4) was indicated.

The error estimation in Table 1 reveals that increasing noise requires also an increasing threshold parameter for the denoiser. Furthermore, the subjectively adjusted threshold is in all cases much higher than the automatically chosen threshold. As the objective assessment was constructed to result in minimal error, the subjective setting by the ear cannot deliver better results. The minimal error thresholds all result in error estimation below 1, the algorithm has thus proved its success. The results of the subjective threshold adjustment can be interpreted as follows: the less noise that is added, the more people will have difficulty to detect it at all. A denoising threshold where the parameter is set too high might then result in “erroneous” removal of audio data, but this is still below audible stimuli.

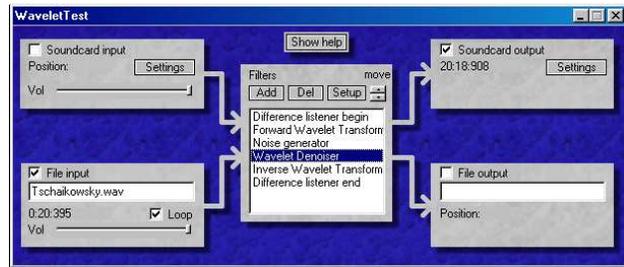
The higher the distortion gets (increasing noise), the better the perceived perturbation, and the better the ideal threshold can be approximated. This result is reflected by the error estimation of the two test series, approaching each other with increasing noise.

## 6. CONCLUSION

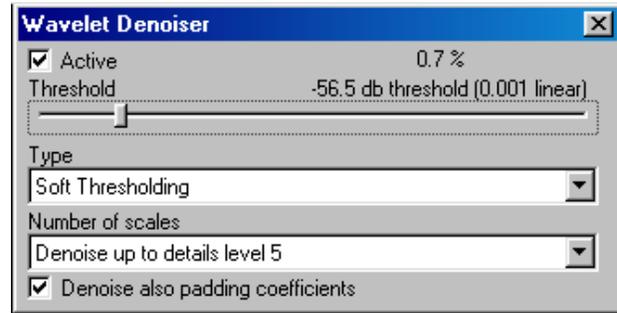
We have presented the basic concepts of digital audio processing and we have discussed the Wavelet Transform and why wavelet analysis is a suitable tool to detect and remove white noise in digital audio. As the choice of threshold parameters is crucial, we have used the real-time aspect of our tool to compare subjective parameter settings to automatic minimal error settings. With low noise, human perception does not hear the artifacts introduced by the denoising algorithms very well. With increasing noise level, subjective assessment approaches objective assessment.

## 7. REFERENCES

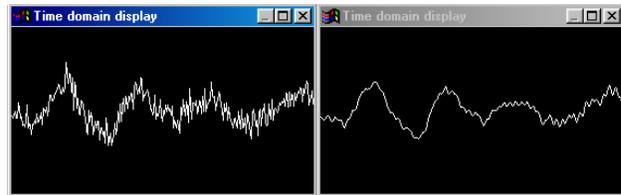
- [1] Dolby, “Making Cassettes Sound Better,” <http://www.dolby.com/cassette/bcsnr/>, 2000.
- [2] Jonathan Berger and Charles Nichols, “Brahms at the piano,” *Leonardo Mus. Journal*, vol. 4, pp. 23–30, 1994.
- [3] David L. Donoho, “Nonlinear Wavelet Methods for Recovering Signals, Images, and Densities from Indirect and Noisy Data,” <http://www-stat.stanford.edu/~donoho/Reports/>, 1993.
- [4] Ingrid Daubechies, *Ten Lectures on Wavelets*, vol. 61, SIAM. Society for Industrial and Applied Mathematics, Philadelphia, Pennsylvania, 1992, ISBN 0-89871-274-2.
- [5] Stéphane Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, San Diego, CA, USA, 1998, ISBN 0-12-466605-1.
- [6] M. Lang, H. Guo, J.E. Odegard, and C.S. Burrus, “Nonlinear processing of a shift invariant DWT for noise reduction,” *SPIE, Mathematical Imaging: Wavelet Applications for Dual Use*, April 1995.
- [7] Curtis Roads, *The Computer Music Tutorial*, The MIT Press, 1996.
- [8] Claudia Schremmer, Thomas Haenselmann, and Florian Bömers, “Wavelets in Real-Time Digital Audio Processing: A Software For Understanding Wavelets in Applied Computer Science,” in *Workshop on Signal Processing Applications (WoSPA)*, <http://www.sprc.qut.edu.au/wospa2000/>, December 2000, Signal Processing Research Center (SPRC) and IEEE.
- [9] Claudia Schremmer, “The Wavelet Tool,” <http://www.informatik.uni-mannheim.de/~cschremm/wavelets/WaveletAudioTool/>, 2000.
- [10] Manojit Roy, V. Ravi Kumar, B.D. Kulkarni, John Sanderson, Martin Rhodes, and Michel van der Stapen, “Simple denoising algorithm using wavelet transform,” *AICHE Journal*, vol. 45, no. 11, pp. 2461–2466, 1999.
- [11] Florian Bömers, “Wavelets in Real-Time Digital Audio Processing: Analysis and Sample Implementations,” M.S. thesis, Universität Mannheim, <http://www.bome.com/personal/thesis.pdf>, 2000.



**Fig. 1.** Graphical user interface of the real-time audio tool. Audio data can be read from soundcard/file and is written to soundcard/file. In-between, a number of filters is applied. Here, we concentrate on the wavelet denoiser.



**Fig. 2.** The wavelet denoiser realizes hard thresholding, and soft thresholding. The number of affected scales can be adjusted from details of level 1 up to all scales. The threshold defines the cut-off frequency, and a flag determines whether the padding coefficients are included.



**Fig. 3.** Display of the time domain before and after application of the wavelet denoiser.

Noise	Objective assessment		Subjective assessment	
	Threshold	$\hat{E}/E$	Threshold	$\hat{E}/E$
-37 dB	-58 dB	0.956	-50 dB	1.121
-34 dB	-50 dB	0.977	-47 dB	1.063
-32 dB	-50 dB	0.921	-45 dB	1.012
-30 dB	-51.5 dB	0.896	-43.5 dB	0.940
-27 dB	-44.5 dB	0.840	-40 dB	0.871

**Table 1.** Evaluation of the wavelet denoiser for `dnbloop.wav`. The noise amplitude is given in dB. Objective assessment yields minimal error estimation  $\hat{E}/E$ . Subjective threshold setting is not optimal, but approaches the minimum with increasing noise.