White Noise Reduction of Audio Signal using Wavelets Transform with Modified Universal Threshold

MATKO SARIC, LUKI BILICIC, HRVOJE DUJMIC
University of Split
R.Boskovica b.b, HR 21000 Split
CROATIA

Abstract: - This paper discusses wavelet-based algorithm for audio denoising. We focused on audio signals corrupted with white noise which is especially hard to remove because it is located in all frequencies. We use Discrete Wavelet transform (DWT) to transform noisy audio signal in wavelet domain. It is assumed that high amplitude DWT coefficients represent signal, and low amplitude coefficients represent noise. Using thresholding of coefficients and transforming them back to time domain it is possible to get audio signal with less noise. We are proposing modified universal thresholding of coefficients which results with better audio signal. The main criterion for evaluation of experimental results was objective degree grade (ODG).

*Keywords: a*udio denoising, wavelets, thresholding, noise reduction

1 Introduction

Signal denoising using wavelets was introduced by Donoho [1]. He developed linear denoising for noise consisting of high frequency components and non-linear denoising (wavelet shrinkage) for noise existing in the low frequency as well.

Schremmer, Haenselmann and Bömers [2] implemented a software for real-time wavelet-based denoising of audio signals. Denoising is achieved using soft or hard thresholding of DWT coefficients. The criterion for successful noise removal is difference between original signal and denoised signal. If we hear difference as silence, that means perfect reconstruction of original signal. The noise that has been added to the signal has been removed and signal has not been modified. If difference sounds like music, part of original signal has also been removed with noise. Error estimation is achieved using ratio between square root energy of difference between original and denoised signal and square root energy of added noise.

Donoho's denoising method with new threshold's value searching method was improved in [3]. First step of their method is based on iteration: output signal from wavelet denoiser is used as new input signal and it is denoised again with same threshold value. In second step they use signal processing method based on the enhancement of diversity of signal to be processed. In this case diversity can be enhanced computing for the same signal some different DWT transforms (transforms with different wavelet mother and number of iterations). For every DWT transform thresholding and IDWT is performed. Denoised signal is obtained by computing the mean of all outputs.

Novel speech enhancement system based on a wavelet denoising framework was introduced in [4]. In this system, the noisy speech is first preprocessed using a generalized spectral subtraction method to initially lower the noise level with negligible speech distortion. A perceptual wavelet transform is then used to decompose the resulting speech signal into critical bands.

Overview of complex wavelets, and their application to audio signal processing (noise reduction and signal compression) is presented in [8].

This paper is organized as follows. In chapter 2 the basic concepts of discrete wavelet transform are described. In chapter 3 denoising using DWT is explained. Simulation results obtained using modified universal threshold are presented in Section 4. We conclude in Section 5.

2 Discrete wavelet transform

Fourier transform gives information about frequency content of signal, but it does not show at what times frequency components occur. It is the reason why we use Short term Fourier transform and wavelet transform for analysis of signals like audio.

Wavelet transform has advantage over Short term Fourier transform because it analyzes the signal at different frequency with different resolutions. High frequency components have good temporal localization, but frequency resolution is poor. Low frequency components have good frequency resolution, but they are not localized in time well. This approach is called multiresolution analysis and it makes sense when signal has high frequency components for short durations and low frequency components for long durations. This approach has certain similarities with Bark-scale of human auditory system: human ear has better frequency resolution at low frequencies and lower frequency resolution at high frequencies.

The discretized continuous wavelet transform enables the computation of the continuous wavelet transform by computers, but it is highly redundant and requires significant computation time and resources. Discrete wavelet transform (DWT) provides analysis and synthesis of original signal with significant reduction in the computation time. Decomposition of the signal is obtained by passing time domain signal through half band low pass and high pass filters. Filtering the signal is equivalent to convolution of signal with impulse response of filter:

$$y_k = x_i \otimes h_i = \sum_{i=-\infty}^{\infty} x_i h_{k-i} \qquad (1)$$

where x_i is input signal, h_i is impulse response of filter and y_i is filter output.

This procedure doubles frequency resolution because frequency band of the output of filter spans over half the previous frequency band. After filtering, half of the samples can be eliminated according to the Nyquist's rule because signal highest frequency is now halved. Therefore the signal is subsampled by 2 simply discarding every other sample. This halves the time resolution because only half the number of samples represent entire signal. Because of that subsampling relation (1) is modified:

$$y_k = x_i \otimes h_i = \sum_{i=-\infty}^{\infty} x_i h_{2k-i} \qquad (2)$$

Procedure is started by filtering input signal with high-pass and low-pass filter. Outputs from high-pass filter are called detail coefficients and they are kept apart as level 1 DWT coefficients. Outputs from low-pass filter, approximation coefficients, are filtered again. This procedure is illustrated in Fig. 1. DWT of original signal is then obtained by concatenating all coefficients starting from last decomposition level. Inverse transform is done by filtering approximation and detail coefficients with lowpass and high pass synthesis filter. Outputs from these two filters form approximation coefficients of next level.

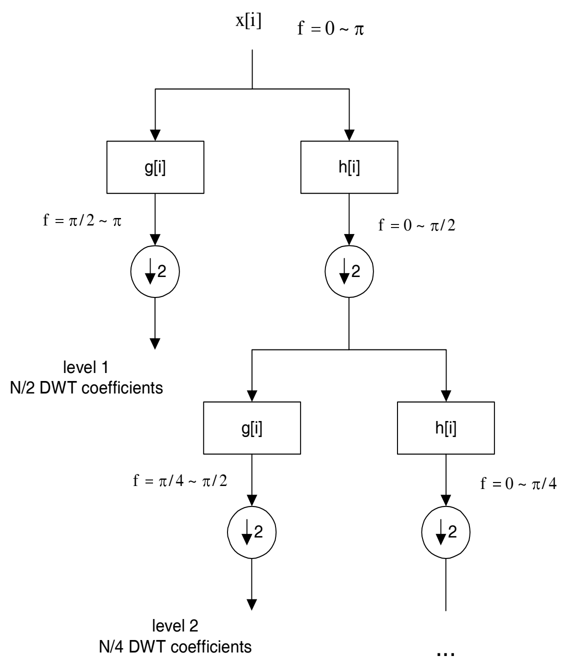

Fig. 1. Discrete wavelet transform (h and g denotes lowpass and highpass filter respectively)

3 Denoising of audio signals using DWT

We assume sampled noisy audio signal y_i

$$y_i = x_i + \sigma_n n_i \qquad i = 1,2,3\ldots N \quad (3)$$

where x_i represents original signal, σ_n is standard deviation of noise and n_i is array of random numbers generated according to Gaussian probability density function with $\mu=0$ and $\sigma^2 = 1$.

Equation (3) in wavelet domain is:

$$W_\Psi\, y_i = (W_\Psi)(x_i+\sigma_n\, n_i) = W_\Psi\, x_i + \sigma_n\,(W_\Psi\, n_i) \qquad (4)$$

where W_Ψ denotes wavelet transform.

If the basis functions of wavelet transform are orthonormal, wavelet transform of the white noise n_i is also white noise w_i of same amplitude. Solving for x_i gives:

$$x_i = \left(W_\Psi^{-1}\right)\left(W_\Psi y_i - \sigma_n w_i\right) \qquad (5)$$

We don't know $\sigma_n\, w_i$, so we estimate it by some value t which gives:

$$\overline{x}_i = \left(W_\Psi^{-1}\right)\left(W_\Psi y_i - t\right) \qquad (6)$$

where $\overline{x}_i$ denotes estimated x_i.

Relation (6) indicates that denoising is in fact removal of noise contribution t from wavelet coefficients. This procedure is called soft thresholding and it is defined with following expression:

$$n_t(z_i) = \begin{cases} \mathrm{sgn}(z_i)(|z_i| - t), & |z_i| > t \\ 0, & \text{else} \end{cases} \qquad (7)$$

where $n_t(z_i)$ is threshold operator and $z_i = W_\Psi y_i$ is wavelet coefficient. $\overline{x}_i$ can be calculated as $\overline{x}_i = \left(W_\Psi^{-1}\right)\left(n_t(z_i)\right)$.

There exist various schemes for selection of threshold t. Their aim is to find threshold value that will efficiently remove noise, but also preserve fidelity of original signal. Too high threshold often cuts part of original signal and causes audible artifacts in denoised signal. On the other hand, too low threshold doesn't remove noise very well.

Denoising algorithm scheme is showed in Fig. 2. First step is windowing of time domain signal because it is usually too long to be processed entirely. First, window length must be chosen: too short window doesn't pick up important time structures of audio signal. On the other side, too long window will loose important short transient details in music. According to [5] we use window length of 4096 samples. Because of nature of DWT algorithm which includes subsampling by factor 2 it is advisable that number of samples is equal to power of two. The simplest windowing function is square equal to 1 over windowing interval and zero elsewhere. This means sharp discontinuities on the edges which can produce large coefficients values in wavelet domain and make it harder to resolve signal from noise. Solution which was used in our work is to add some extra coefficients to window of samples and so form extended window.

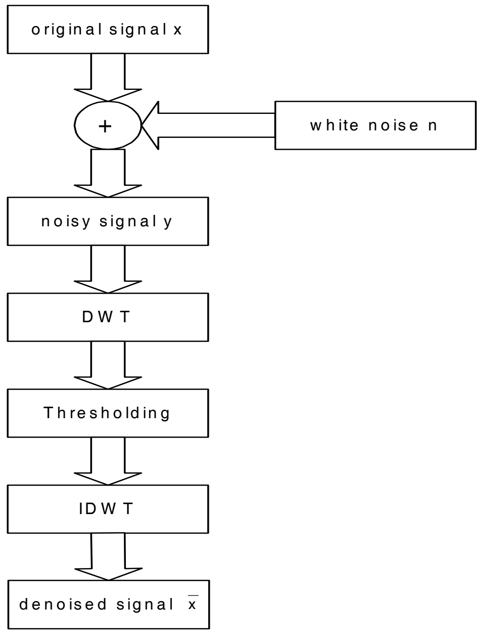

Fig. 2. Noise removal algorithm

Noisy signal is obtained by adding an array $\sigma_n n_i$ to array of original signal x_i. Using DWT degraded signal is transformed into wavelet domain.

One of the first methods for selection of threshold t was developed by Donoho and Jonstone [6] and it is called VisuShrink (universal threshold). Slightly different universal threshold was proposed in [3]:

$$t = \sigma_n \sqrt{2 \log_2(N)} \qquad (8)$$

where N denotes number of samples and σ_n is standard deviation of noise.

During our research it is noticed that threshold obtained by (8) is too high. It must be corrected if we want to get maximum performance especially for low level noise.

After thresholding, inverse DWT is applied to get denoised time domain signal.

4 Simulation and results

We implemented white noise removal algorithm in Matlab which has large collection of functions for wavelet analysis (Wavelet toolbox). We choose 8-level DWT and applied soft thresholding on all levels including approximation coefficients too. For DWT Daubechies filters with 6 coefficients are used. We noticed that filter of higher order increase computational complexity, but also improves denoising results.

Input of our simulation is original signal in Wave format. Noisy signal is created by adding random generated numbers $\sigma_n n_i$ to the orignal signal samples (3). According to window length, blocks of 4096 samples are processed individually.

Criterion used for evaluation of results obtained by denoising algorithm is objective degree grade (ODG), variable obtained according to [5]. We have also compared it with mean square error (MSE) which is widespread used for estimating signal quality.

Input MSE is defined as:

$$\frac{1}{N}\sum_i (x_i - y_i)^2 \qquad (9)$$

where x_i is original signal and y_i is noisy signal.

Output MSE is defined as:

$$\frac{1}{N}\sum_i (x_i - \overline{x}_i)^2 \qquad (10)$$

where $\overline{x}_i$ is estimated x_i (noisy signal y passed through denoising algorithm (Fig. 3)).

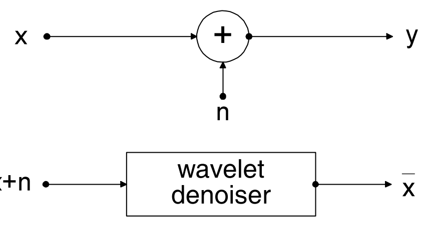

Figure 3. Signals used for calculating mean square error

Lower MSE means the closer match between two signals. According to this criterion denoising is successful if output MSE is lower than input MSE. In that case denoised signal $\overline{x}_i$ is closer to original signal x than noisy signal y.

MSE is conventional method for estimating signal quality, but for audio signal they aren't always in accordance with listening tests. An example which is often mentioned to illustrate limitations is so-called 13 dB miracle. Superimposed noise with spectral structure adapted to audio signal is almost inaudible even if resulting SNR declines to 13 dB [7].

As a main criterion for quality estimation we have used Objective degree grade (ODG) which is calculated according to [7]. ODG corresponds to sbjective listening tests. ODG values range from 0 (imperceptible impairment) to -4 (very annoying impairment). Lower ODG means more annoying impairment.

We investigated maximum gain in objective degree grade for few input signals with different levels of degradation. Amount of white noise added to original signal is controlled with variable σ_n (3). Maximum ODG gain is obtained replacing threshold t (8) with

$$t = k \cdot \sigma_n \sqrt{2\log_2(N)} \qquad (11)$$

where 0<k<1. During our work it is noticed that universal threshold t given by equation (8) is too high for audio signals and it cuts part of original signal too. So it is modified with factor k in order to obtain higher quality output signal. The value of k was changed gradually with steps of 0.1. In that way for every noise level we found k that gives the best result. This is shown in table 1 (example 1: 8.2 s long excerpt from song Sultans of swing, Dire straits.).

σ_n	Input ODG	k	output ODG	input MSE (10^{-6})	output MSE (10^{-6})
0.001	-1.24	0.2	-0.76	0.99	1.558
0.0025	-2.38	0.3	-1.79	6.228	9.399
0.005	-3.34	0.3	-2.79	24.972	29.418
0.0075	-3.58	0.3	-3.34	56.183	55.74
0.01	-3.72	0.5	-3.55	99.624	112.86

Table 1. Results of noise removal algorithm for modified soft thresholding method (example 1)

Table 2 shows results for second example: 15.4 s long excerpt from Spring (Antonio Vivaldi).

σ_n	Input ODG	k	Output ODG	input MSE (10^{-6})	output MSE (10^{-6})
0.0005	-1.92	0.3	-0.82	0.25069	0.46022
0.00075	-2.75	0.6	-1.19	0.56078	2.4435
0.001	-3.23	0.6	-1.46	1.0008	3.9298
0.0025	-3.78	0.7	-2.51	6.2533	2.1687

Table 2. Results of noise removal algorithm for modified soft thresholding method (example 2)

For different audio examples significant enhancement is obtained, but with different k values. It suggests that better denoising can be obtained using modification of threshold but it also means that modification factor k should be calculated for each audio sample.

From results in Table 1 and Table 2 (modified soft thresholding) it is obvious that gain in objective degree grade doesn't correspond to MSE gain. MSE shows little enhancement or even loss while ODG and also informal listening tests prove significant enhancement of signal quality. These results confirm that MSE is not always in accordance with perceptual quality of audio signal. Denoising algorithm works better for lower noise signals. For higher amount of added noise higher threshold must be set, but except noise it removes part of original signal causing audible artifacts in denoised signal.

5 Conclusion

In this paper we have used wavelet transform for denoising audio signal corrupted with white noise. Audio denoising is performed in wavelet domain by thresholding wavelet coefficients. It is shown that universal threshold must be corrected in order to get maximum performance especially for low level noise.

Universal threshold is modified with factor that should be calculated for each audio example. For different audio examples significant enhancement is obtained comparing to universal threshold. However, the perfect denoising isn't possible: higher threshold removes noise well, but the part of original signal is also lost.

Objective degree grade (ODG) was used as main criterion and threshold was adjusted to find maximum ODG gain. For every maximum output ODG MSE gain is also calculated. The results confirm that MSE is not always in accordance with perceptual quality of audio signal.

References:

[1] Donoho, D. L., "Denoising via soft thresholding", IEE*E Trans. on Information Theory,* 41: 613-627, 1995

[2] Claudia Schremmer, Thomas Haenselmann, Florian Bomers: "Wavelet based audio denoiser", *IEEE International Conference on Multimedia and Expo (ICME) 2001*, pp. 145-148, Tokyo, Japan, August 2001

[3] Alexandru Isar, Dorina Isar: "Adaptive denoising of low SNR signals", *Third International Conference on WAA 2003*, Chongqing, P. R. China, 29-31 May, 2003, p.p. 821-826

[4] Fu Qiang, Wan Eric A.: "Perceptual wavelet adaptive denoising of speech", *EUROSPEECH-2003*, Geneve, Switherland, p.p. 577-580

[5] Jonathan Berger, Ronald R. Coifman, Maxim J. Goldberg: "Removing noise from music using local trigonometric bases and wavelet packets", *Journal of the Audio Engineering Society*, 42(10):808– 818, October 1994

[6] D.L.Donoho: "De-noising by Soft Thresholding", Technical Report no.409, Stanford University, December 1992

[7] Method for objective measurements of perceived audio quality, RECOMMENDATION ITU-R BS.1387-1

[8] P.J. Wolfe, S.J. Godsill: "Audio signal processing using complex wavelets", presented at the 114th Convention of the Audio Engineering Society, March 2003