WAVELET SHRINKAGE DENOISING APPLIED TO REAL AUDIO SIGNALS UNDER PERCEPTUAL EVALUATION

Luiz W. P. Biscainho, Fábio P. Freeland, Paulo A. A. Esquef and Paulo S. R. Diniz
COPPE/PEE & EE/DEL, UFRJ,
CP 68504, CEP 21945-970, Rio de Janeiro, RJ, BRAZIL
Tel: +55 21 2605010; fax: +55 21 5900788
e-mail: {wagner,freeland,esquef,diniz}@lps.ufrj.br

ABSTRACT

This paper addreses wavelet denoising of audio recordings from a practical viewpoint. First, a set of real high-quality audio signals with distinct characteristics is artificially corrupted by pseudo-white noise. Then, the classical wavelet shrinkage denoising method is applied to them under varied settings, including different wavelets, numbers of scales, threshold application and computation methods etc. At last, an adapted version of the Perceptual Audio Quality Measure (PAQM) is used as an objective quality index to compare the obtained results. This way, the work provides some insight on the performance of wavelet shrinkage applied to corrupted high-quality signals, the performance of the PAQM in this application and the relative importance of different parameters in the final quality determination.

1 INTRODUCTION

The Discrete Wavelet Transform (DWT) [1, 2] has become a popular signal processing tool, especially after their connection to discrete-time systems through filter banks. Its applications include signal analysis, coding, compression, denoising etc.

Wavelet shrinkage [3, 2] is a method to recover a signal from noisy data, closely linked to signal compression, that selects a reduced number of DWT coefficients yet capable of accurately representing the signal.

Denoising audio signals corrupted by broadband noise is a difficult task, which has led to techniques of different degrees of complexity [4, 5]. The problem is even more challenging when the underlying signal has wide dynamic and spectral ranges, as its subtleties must be preserved (common examples are magnetic-tape matrices of hi-fi recordings from the early 60's). An issue often associated to wavelet shrinkage is its capability to preserve details, which suggests it can handle denoising of high-quality audio at low computational cost.

Assessing the performance of these methods is a critical point. Several simple standard signals are frequently used [6], employing objective performance measures like signal-to-noise ratio (SNR). However, audio processing algorithms should be ideally evaluated through the typically more complex real audio signals. Furthermore, in spite of their clear mathematical meaning, objective measures can fail to reflect subjective quality, whereas reliable human assessment requires a large set of tests under the same conditions. A good compromise is reached by objective measures based on Psychoacoustics.

This work applies wavelet shrinkage to a set of real modern audio signals corrupted by broadband noise, and uses the Perceptual Audio Quality Measure (PAQM) proposed in [7]—originally meant to evaluate how much processed signals depart from their original versions, e.g. after compression—to evaluate the results of the restoration of each signal for different settings of the algorithm. Our objective is to set a framework which, while not claiming to be complete, includes:

- the performance of wavelet shrinkage applied on corrupted versions of high-fidelity signals;
- the performance of the perceptual measure in this application;
- the weight of each parameter in the attained overall quality.

In the following, Section 2 reviews the theoretical background linked to the work and Section 3 presents experimental tests. Conclusions are drawn in Section 4.

2 RELATED BACKGROUND

2.1 DWT and Wavelet Shrinkage

Consider the set $\tilde{\psi}(t)$, $\psi(t)$, $\tilde{\varphi}(t)$ and $\varphi(t)$ of analysis and synthesis mother-wavelets and analysis and synthesis scaling functions, respectively. If the time-domain $f(t)$ has no detail beyond scale $j = J$, its DWT [1] expansion gives

$$f(t) = \sum_{k=-\infty}^{\infty} c_k \varphi_k(t-k) + \sum_{k=-\infty}^{\infty} \sum_{j=0}^{J} d_{j,k} \psi_k(t-k),$$

with coefficients obtained by

$$c_k = \int f(t)\tilde{\varphi}_k(t)dt \quad \text{and} \quad d_{j,k} = \int f(t)\tilde{\psi}_{j,k}(t)dt,$$

where $\{\cdot\}_{j,k}(t)$ are j-scaled and k-shifted versions of $\{\cdot\}(t)$. When analysis and synthesis functions are the same, the system is called orthogonal; otherwise, the general system, as presented above, is called biorthogonal. The family of orthogonal wavelets which achieves minimum time support (associated to localization capability) for a given regularity (a property linked to compactness of representation) is the Daubechies, whose simplest version is the Haar wavelet [2].

The wavelet shrinkage method [3, 2] consists in DWT-expanding the signal corrupted by white noise, discarding all coefficients whose magnitude is below a threshold λ and re-synthesizing the signal by Inverse DWT. Two different thresholding strategies can be adopted [1, 2]:

- Hard thresholding, which preserves the remaining DWT coefficients;
- Soft thresholding, which reduces their magnitudes by λ.

There are several ways to calculate the threshold [3, 8], like:

- Minimax, which makes λ proportional to an estimate of the noise standard deviation $\hat{\sigma}$;
- Stein's Unbiased Risk Estimate (SURE), which chooses λ such that the signal estimate mean-square-error (MSE) is minimized;
- Hybrid—a combination of Minimax and SURE.

2.2 PAQM

This measure [7] takes into account some perceptual issues, two of which have to be defined:

- the Critical Band around a tone is the bandwidth that contributes for masking the tone;
- Masking is the property a sound has to inhibit the perception of another one, near to the former in time or frequency.

The implied computation of the perceived sound representation is performed along successive frames of the sampled signal. For each frame:

1. the signal energy spectrum is divided in contiguous critical bands;
2. masking effects are modeled and applied to each band;
3. the perceived sound intensity is obtained for each band by an adequate energy compression.

Based on that representation calculated for both the original signal and a corrupted version of it, the PAQM gives an overall index of dissimilarity: PAQM = 0 means that both signals would be perceived as identical.

In our application, this index is calculated between the original and the noisy version of one signal, as an initial measurement of perceptual noise effects. After that, new indexes are calculated between the original and the restored version of the signal for different sets of parameters (see Fig. 1). Lower PAQM (i.e. closer

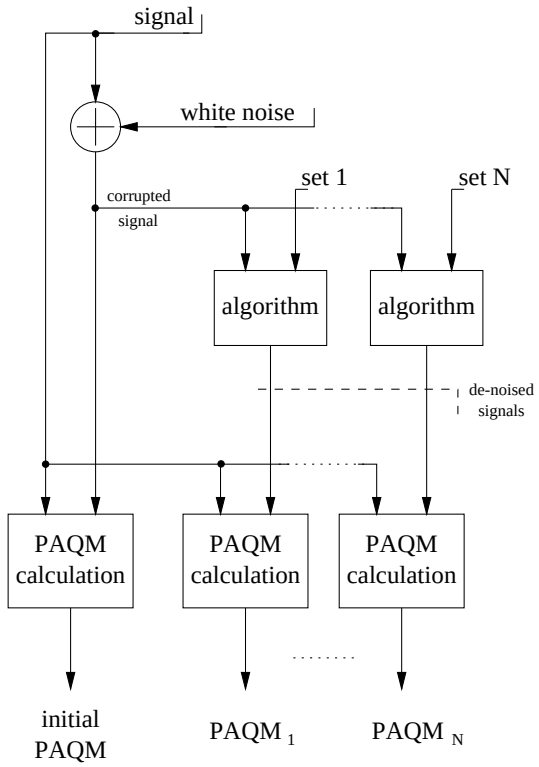

Figure 1: PAQM used to compare the quality of restoration results.

similarity) indicates better performance.

3 TESTS

3.1 Signals and Strategy

Tests employed a set of high-quality music signals, five of them purely instrumental (I1 to I5) and the sixth an unaccompanied singing voice (V). Their choice intended to cover some common scenarios in real audio:

- I1 has rich spectral content and reduced power variations;
- I2 spectrum concentrates on low and high frequencies and exhibits reduced power variations;
- I3 can be considered stationary for long time intervals;
- I4 can be considered stationary only for short time intervals;
- I5 has rich spectral content and exhibits extreme power variation;

Table 1: PAQM of the corrupted signals.

Signal	SNR = 30 dB	SNR = 40 dB
$I1_c$	0.5236	0.1462
$I2_c$	0.2944	0.0610
$I3_c$	0.2541	0.0335
$I4_c$	0.2249	0.0330
$I5_c$	0.2317	0.1247
V_c	0.1845	0.0471

Table 2: PAQM of de-noised versions of signals with diferent SNR values ($J = 2$, soft, SURE, Daubechies 32).

Signal	SNR = 30 dB	SNR = 40 dB
$I1_d$	0.2118	0.0766
$I2_d$	0.1205	0.0373
$I3_d$	0.0955	0.0247
$I4_d$	0.1076	0.0197
$I5_d$	0.2014	0.1170
V_d	0.1439	0.0401

- V aims to explore the processing effects on particular voice characteristics like intelligibility.

Pseudo-white noise sequences were used to obtain additively corrupted versions (indexed $\{\cdot\}_c$) of each signal with different values of mean SNR, on which tests were performed. Their original PAQM values are shown in Table 1. Perceptually, those signals with SNR = 30 dB could be considered strongly contaminated.

For each corrupted signal, restoration was accomplished for several versions of the algorithm, combining different values of number of scales, threshold computation and application methods, wavelet families and wavelet time supports. Each restored signal was compared to its uncorrupted version by the PAQM.

3.2 Results and Discussion

From the extensive set of simulations performed, a group of selected PAQM values related to the de-noised signals (indexed $\{\cdot\}_d$) are presented in the following for easier confrontation. First, Table 2 shows the results of restoration of each signal referred in Table 1, employing the basic settings $J = 2$, soft thresholding, SURE calculation method and wavelet Daubechies 32. Then, each parameter is individually varied, while the others are maintained fixed. The corresponding results, relative to the signals with SNR = 30 dB, are summarised in Tables 3 to 6:

- Table 3 shows the results for $J = 2$ (3 scales) and 7 (8 scales);
- Table 4 compares soft and hard thresholding;

Table 3: PAQM of de-noised signals obtained for different numbers of scales (SNR = 30 dB, soft, SURE, Daubechies 32).

Signal	$J = 2$	$J = 7$
$I1_d$	0.2118	0.2123
$I2_d$	0.1205	0.1205
$I3_d$	0.0955	0.0955
$I4_d$	0.1076	0.1076
$I5_d$	0.2014	0.2018
V_d	0.1439	0.1432

Table 4: PAQM of de-noised signals obtained by soft and hard thresholding (SNR = 30 dB, $J = 2$, SURE, Daubechies 32).

Signal	Soft	Hard
$I1_d$	0.2118	0.4395
$I2_d$	0.1205	0.2322
$I3_d$	0.0955	0.1847
$I4_d$	0.1076	0.1811
$I5_d$	0.2014	0.2075
V_d	0.1439	0.1633

Table 5: PAQM of de-noised signals obtained through different threshold calculation methods (SNR = 30 dB, $J = 2$, soft, Daubechies 32).

Signal	Minimax	SURE	Hybrid
$I1_d$	0.6546	0.2118	0.2118
$I2_d$	0.6592	0.1205	0.1205
$I3_d$	0.4699	0.0955	0.0955
$I4_d$	0.3741	0.1076	0.1076
$I5_d$	0.4256	0.2014	0.2014
V_d	0.5309	0.1439	0.1439

- Table 5 compares Minimax, SURE and Hybrid threshold calculation methods;
- Table 6 presents the results for wavelets Haar, Daubechies 16, 32 and 64 and Biorthogonal 3.1 and 3.7 [1].

A first glance at Table 1 highlights the incoherence between SNR and perceptual evaluation, since different signals with the same SNR are judged quite differently according to PAQM.

Subjective evaluation of results showed that wavelet shrinkage not always attains good performance on noisy versions of high-quality signals, especially those with very low SNR. This fact is confirmed in Table 2, where final PAQM is worse for signals with lower original SNR.

Table 6: PAQM of de-noised signals obtained for diferent wavelet families and time supports (SNR = 30 dB, $J = 2$, soft, SURE).

Signal	Haar	Daubechies			Biorthogonal	
		16	32	64	3.1	3.7
$I1_d$	0.4920	0.2371	0.2118	0.1972	0.1250	0.0955
$I2_d$	0.2742	0.1242	0.1205	0.1279	0.1854	0.1514
$I3_d$	0.2133	0.1009	0.0955	0.1021	0.1862	0.1303
$I4_d$	0.1932	0.1138	0.1076	0.1064	0.1618	0.1309
$I5_d$	0.2173	0.2023	0.2014	0.2004	0.2564	0.2223
V_d	0.1556	0.1449	0.1439	0.1432	0.3715	0.3133

A common perceptual effect is the original homogeneous high-amplitude noise changing into a disturbingly varying, yet low-amplitude, residual.

Voice and signals with large power variations were difficult to deal with, whereas spectrum shape and stationarity did not seem to be critical factors. Accordingly, a comparison between Tables 2 and 1 indicates that signals I1 to I4 were much more improved than I5 and I6.

The PAQM algorithm emulated quite well subjective opinion, considering that it measures the difference between the signal under test and its original version, and not directly the lack of quality of the former.

Increasing the number of scales had little effect on performance, after $J = 2$ (see Table 3).

As expected, soft thresholding led to better results than hard thresholding (see Table 4).

Minimax is known as a better method for low SNR. The fact that SURE calculation method was the best choice and led to the same PAQM values as the Hybrid method (see Table 5) means that 30 dB is being considered a 'high' SNR, in this context. Even in additional tests performed on a version of $I1_c$ with SNR = 20 dB (which in this application is considered too low an SNR value), the Minimax and Hybrid methods were fairly efficient, but not as much as SURE.

Among the wavelet families, Daubechies achieved the best performance, with no need for long time supports (which can be loosely related to more selective frequency separation), and a Biorthogonal wavelet similar in form to Daubechies gave similar results (see Table 6). It became evident that the wavelet shape plays a fundamental role in the processing quality.

4 CONCLUSIONS

This work investigated, through a large number of simulations, aspects of wavelet shrinkage applied in denoising of high-quality real audio signals and its performance evaluation through a perceptual measure. Limitations of the restoration method under some circumstances, such as large power variations and voice, were highlighted. Testing of the involved parameters showed the importance of choosing an appropriate wavelet family in the final quality determination, besides some clearly recommended options (like soft thresholding and SURE threshold calculation). Therefore, searching the best wavelet bases for audio is a natural concern. We also showed that PAQM can be a reliable, yet indirect, performance test for audio restauration techniques, replacing human assessment, since reference signals are available.

References

[1] C. S. Burrus, R. A. Gopinath, and H. Guo, *Introduction to Wavelets and Wavelet Transforms - A Primer*. New Jersey, NJ, USA: Prentice Hall, 1997.

[2] S. Mallat, *A Wavelet Tour of Signal Processing*. San Diego, CA, USA: Academic Press, 1997.

[3] D. L. Donoho, I. Johnstone, G. Kerkyacharian, and D. Picard, "Wavelet shrinkage: Asymptopia?," *J. Roy. Stat. Soc. B*, vol. 57, no. 2, pp. 301–337, 1995.

[4] S. J. Godsill and P. J. W. Rayner, *Digital Audio Restoration*, ch. 6, pp. 135–149. London, UK: Springer-Verlag, 1998.

[5] J. Berger, R. R. Coifman, and M. J. Goldberg, "Removing noise from music using local trigonometric bases and wavelet packets," *J. Audio Eng. Soc.*, vol. 42, pp. 808–818, October 1994.

[6] D. L. Donoho, "Wavelet shrinkage and w. v. d.: A ten-minute tour," Tech. Rep. 416, Department of Statistics, Stanford University, Stanford, CA, USA, June 1993.

[7] J. G. Beerends and J. A. Stemerdink, "A perceptual audio quality measure based on a psychoacoustic sound representation," *J. Audio Eng. Soc.*, vol. 40, pp. 963–978, December 1992.

[8] D. L. Donoho and I. Johnstone, "Adapting to unknown smoothness ideal via wavelet shrinkage," Tech. Rep. 425, Department of Statistics, Stanford University, Stanford, CA, USA, July 1994.