Speech signal denoising with wavelet-transforms and the mean opinion score characterizing the filtering quality

Alauldeen S. Yaseen[a,b], Alexey N. Pavlov[a,c,e] and Alexander E. Hramov[c,d,e]

[a] Physics Dept., Saratov State University, Astrakhanskaya Str. 83, Saratov, 410012, Russia;
[b] University of Technology, Alsina'a Str., Baghdad, Iraq;
[c] Research and Educational Center "Nonlinear Dynamics of Complex Systems", Saratov State Technical University, Politechnicheskaya Str. 77, Saratov 410054, Russia;
[d] Faculty of Nonlinear Processes, Saratov State University, Astrakhanskaya Str. 83, Saratov, 410012, Russia;
[e] Saratov Branch of Kotel'nikov Institute of Radio-Engineering and Electronics of Russian Academy of Sciences, Zelyonaya Str. 38, Saratov, 410019, Russia

ABSTRACT

Speech signal processing is widely used to reduce noise impact in acquired data. During the last decades, wavelet-based filtering techniques are often applied in communication systems due to their advantages in signal denoising as compared with the Fourier-based methods. In this study we consider applications of a 1-D double density complex wavelet transform (1D-DDCWT) and compare the results with the standard 1-D discrete wavelet-transform (1D-DWT). The performances of the considered techniques are compared using the mean opinion score (MOS) being the primary metric for the quality of the processed signals. A two-dimensional extension of this approach can be used for effective image denoising.

Keywords: wavelet-transform, denoising, thresholding, speech signal

1. INTRODUCTION

Wavelet theory has numerous applications for denoising of signals and images allowing to improve the quality of information received in communication systems. An important applied problem is, e.g., speech enhancement for the needs of small portable devices such as the hearing-aid devices.[1] Effective noise filtering can be provided by different ways, and elaboration of appropriate methods of data processing remains an actual technical problem. There are various sources of noise presented in the environment, and the corresponding fluctuations may occupy either some specific frequency band or the entire frequency band. When frequency bands of noise and information signal are coincide (at least, partly), it becomes very difficult to remove noise without losing information about the signal. Therefore, noise filtering with keeping original features of the processed signal is a challenging task.[2] Wavelet-based denoising techniques have advantages in solving this problem due to their good time-frequency localization and the possibility of performing the multiresolution analysis.

Denoising can be provided by the correction of the wavelet-coefficients with soft or hard thresholding approaches. The standard filtering technique is the 1D-DWT. Within this approach, large decomposition coefficients are mainly associated with the signal, while small wavelet coefficients are caused by fluctuations. Denoising is achieved by selecting an appropriate threshold to remove noise-related coefficients. An effective denoising method is based on the 1D-DWT with the soft thresholding.[3]

However, 1D-DWT has essential disadvantages that are discussed in many review papers and monographs. In order to improve the quality of data processing, various modifications and extensions of the standard denoising methods were proposed. Among these new tools, a 1D-DDCWT should be mentioned.[4] It uses frames or overcomplete expansions that provide a possibility of better data compression and noise reduction due to applying smoother basic functions. Another tool is the dual-tree complex wavelet transform (1D-DTCWT) that also

Further information: (Send correspondence to A.N.P.) E-mail: pavlov.lesha@gmail.com, Telephone: +7 8452 210 710

Dynamics and Fluctuations in Biomedical Photonics XIII, edited by Valery V. Tuchin, Kirill V. Larin, Martin J. Leahy, Ruikang K. Wang, Proc. of SPIE Vol. 9707, 970719
© 2016 SPIE · CCC code: 1605-7422/16/$18 · doi: 10.1117/12.2211384

improves some shortcomings of the 1D-DWT. Both, the 1D-DDCWT and the 1D-DTCWT have their own distinct characteristics and advantages. Due to this, they are sometimes used within combined tools for signal denoising.[4]

In order to provide a characterization of speech signal denoising, the MOS test is often applied that gives an index for speech quality rating. Recently, several objective MOS assessment methods were proposed, such as the perceptual evaluation of the speech quality (PESQ). It evaluates the audible distortions based on the perceptual domain representation of two signals, namely, the original signal and the filtered signal.

In this study we compare abilities of 1D-DDCWT and the standard 1D-DWT. Our analysis is based on numerical measures of the denoising quality such as the root mean square (RMS) error and the MOS. All estimations are performed using MATLAB softwares.

2. METHODS

2.1 1-D Discrete Wavelet Transform

The simplest wavelet-based approach to signal denoising uses the critically sampled DWT and the pyramidal decomposition scheme. Wavelet-coefficients reflecting features of the signal's structure are corrected before performing the inverse DWT. The 1D-DWT approach assumes the perfect reconstruction with two filter banks.

An analyzed function $f(t) \in L^2(R)$ can be decomposed using the wavelet $\psi(t)$ and the scaling function $\phi(t)$ as follows

$$f(t) = \sum_k c_{j_0}(k)\phi_{j_0,k}(t) + \sum_{j \geq j_0} \sum_k d_j(k)\psi_{j,k}(t), \tag{1}$$

where j_0 is an arbitrary starting scale, $c_{j_0}(k)$ are the approximation or the scaling coefficients, and $d_j(k)$ are the detail or the wavelet coefficients. The expansion coefficients are calculated using a pyramidal algorithm.[5–8]

The result of 1D-DWT is a multilevel decomposition, in which the signal is decomposed into approximation and detail coefficients at each level of resolution. Within this decomposition, the application of scaling function and wavelets can be treated as the low-pass and the high-pass filtering of data series.[9]

An important part of signal denoising is the thresholding of wavelet-coefficients $d_j(k)$. It is typically assumed that small coefficients are mainly related to noise while large coefficients are associated with important signal's features. Two main thresholding techniques are typically applied, namely, the hard and the soft thresholding.[10] Hard thresholding is a "keep or kill" procedure; it removes noise-related coefficients leaving unchanged coefficients associated with the information signal

$$p(x) = \begin{cases} x, & \mid x \mid \geq C, \\ 0, & \mid x \mid < C. \end{cases} \tag{2}$$

Since it is quite difficult to select an appropriate threshold C, distortions of the filtered signal occur. Soft thresholding corrects all coefficients

$$p(x) = \begin{cases} x - C, & x \geq C, \\ x + C, & x \leq -C, \\ 0, & \mid x \mid < C, \end{cases} \tag{3}$$

however, the continuity of the function (3) has some advantages. It makes algorithms mathematically more tractable. Sometimes, pure noise coefficients may pass the hard threshold and appear as annoying "blips" in the processed signal. Soft thresholding shrinks these false structures.

2.2 1-D Double Density Complex Wavelet Transform

An extension of 1D-DWT, namely, the 1-D double density DWT (1D-DDDWT) was recently proposed. It is based on a single scaling function and two distinct wavelets where the two wavelets are designed to be offset from one another by one half, the integer translations of one wavelet fall midway between the integer translations of the other wavelet. In this way, 1D-DDDWT approximates the continuous wavelet transform (having more wavelets than necessary thus providing a closer spacing between adjacent wavelets within the same scale). 1D-DDDWT is two-times expansive regardless of the number of scales implemented potentially much less than for the standard 1D-DWT. 1D-DDDWT has twice as many coefficients as the critically sampled 1D-DWT. To construct 1D-DDDWT, the oversampled filter banks are used. Let the filter $h_0(n)$ is the low-pass (scaling) filter, while $h_1(n)$ and $h_2(n)$ are the high-pass (wavelet) filters.

A three-channel filter bank is used to develop 1D-DDCWT corresponding to a wavelet frame based on the scaling function $\Phi(t)$ and wavelets $\Psi_1(t)$ and $\Psi_2(t)$

$$\begin{aligned}\Phi(t) &= \sqrt{2}\sum_n h_0(n)\Phi(2t-n),\\ \Psi_i(t) &= \sqrt{2}\sum_n h_i(n)\Phi(2t-n), \quad i=1,2.\end{aligned} \tag{4}$$

1D-DDDWT has the following important properties: 1) it employs one scaling function and two distinct wavelets which are designed to be offset from one another by one half; 2) 1D-DDDWT is overcomplete by a factor of two; 3) it is nearly shift-invariant if complex wavelets (with real and imaginary parts related via the Hilbert transform) are considered for signal denoising.[11] In the latter case the approach known as the 1D-DDCWT is considered. Within the 1D-DDCWT the input signal is processed by parallel iterated filter banks for real parts $h_i(n)$ and imaginary parts $g_i(n)$, where $i = 0, 1, 2$. A general structure of such decomposition is shown in Figure 1 for the analysis stage (and similar diagram can be used for the synthesis stage).

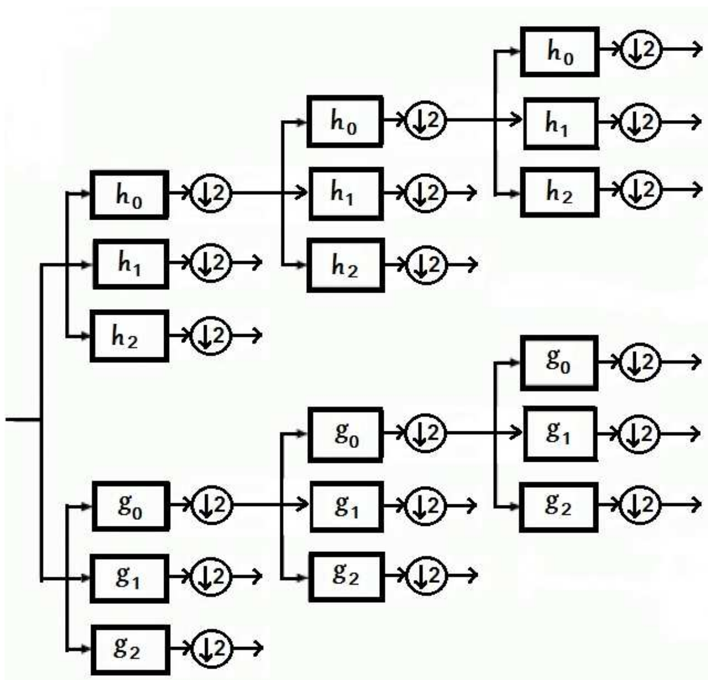


Figure 1. Filter banks diagram of 1-D DDCWT

2.3 Mean Opinion Score

Mean Opinion Score (MOS) – ITU-T P.800 was used as a primary metric for the quality of the signal after its processing. This is a dimensionless quantity with values ranging from 1 to 5. It gives an estimate of human perception of sound quality. Estimating MOS with humans is a long and expensive procedure, involving many humans listening to the records and giving their subjective opinions. For this reason, MOS is not suitable during the stage of the algorithm development. Here, we used an objective perceptual evaluation of the sound quality (PESQ) – ITU-T P.862. It produces similar results to MOS in the same scale (1 to 5) to give an estimate of human perception of sound quality. Our analysis was based on the Matlab implementation of PESQ algorithm.[12]

3. RESULTS

A wavelet transform must be specified by its analysis and synthesis filter banks, single-level convolutions and boundary treatment, and the total number L of iterated multiresolution levels. For the purpose of this research of the potentials of wavelet transforms in denoising speech signals, several experiments were performed. Within the 1D-DWT approach we used a 3-level decomposition with the Daubechies wavelet D^8 (Figure 2). Thus the decomposed signal is characterized by approximation coefficients (at level 3) and detail coefficients at levels 1, 2, 3.

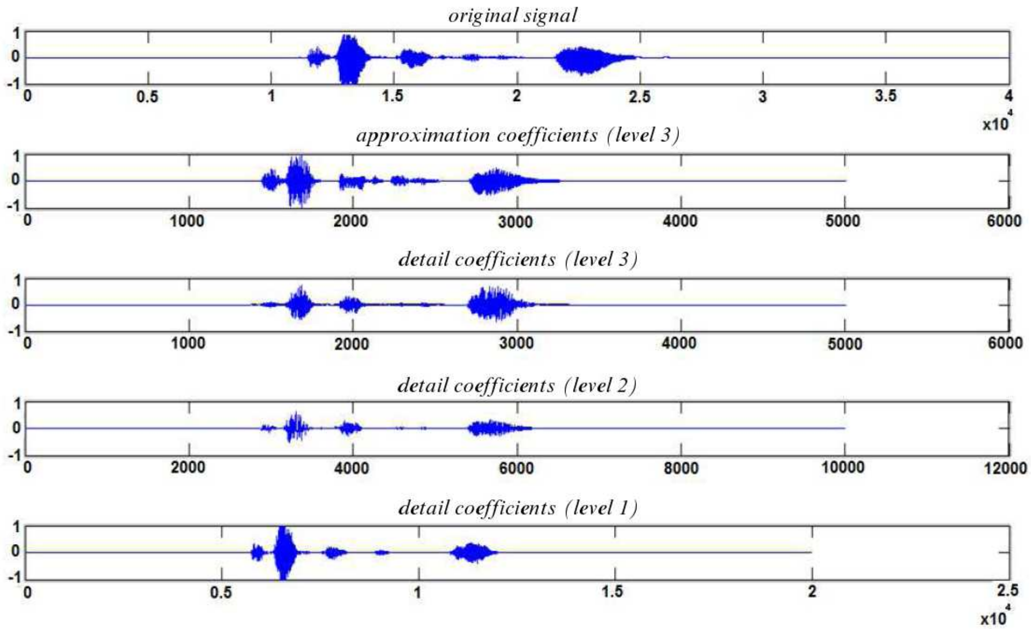
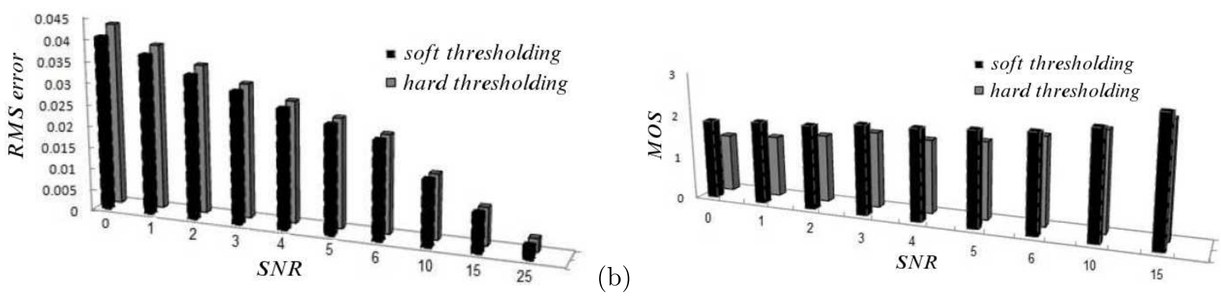


Figure 2. Decomposition of a speech signal using 1D-DWT and 3 levels of resolution.



(a) (b)

Figure 3. RMS error (a) and MOS (b) for 1D-DWT approach versus SNR, dB.

The considered test speech signal was the phrase “Hi, how are you?” (in Russian). It was saved as *.wav-file and sampled at 44100 kHz. Aiming to compare the quality of filtering provided by different methods for noise reduction, this signal was corrupted by Gaussian noise with SNR changed from 0 dB till 25 dB. RMS error between the original (noise-free) and the filtered signal was used to quantify the filtering performance. Best noise reduction was associated with the minimal RMS error.

At the first stage, we compared the results of the 1-D DWT with Daubechies wavelet D^8 for two types of thresholding (hard and soft thresholding – Figure 3). According to the obtained results, the soft thresholding is more efficient than the hard thresholding for all noise levels. At the second stage, we applied the 1D-DDCWT approach. Figure 4 illustrates advantages of this tool. Here, we show results for high noise levels (SNR=0, 1, 2 dB).

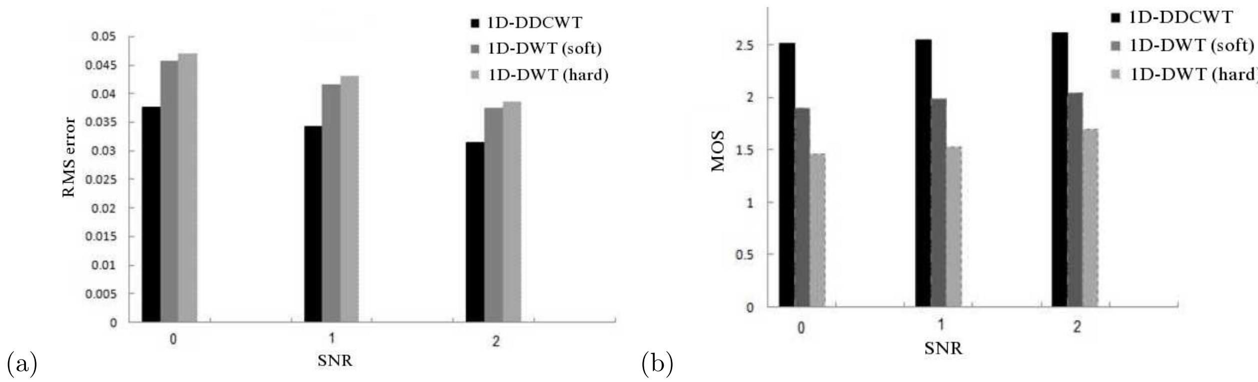


Figure 4. RMS error (a) and MOS (b) versus SNR, dB, for different filtering techniques.

Further we compared the results of noise reduction depending on the decomposition level L. Figure 5 shows the values of RMS error and MOS for the 1D-DWT with high and soft thresholding. According to Figure 5a, RMS error slightly decreases with L, and the reduction of this quantity is small for $L > 3$. Increased decomposition level increases the computational complexity of the denoising algorithm. However, higher decomposition levels are not significantly improve the method's quality. In practice, the decomposition level should be limited to increase the speed of data processing (e.g., $L \leq 6$). The value $L = 3$ represents a good compromise between increased computational complexity and reduced filtering errors. Figure 5b also confirms advantages of the soft thresholding strategy.

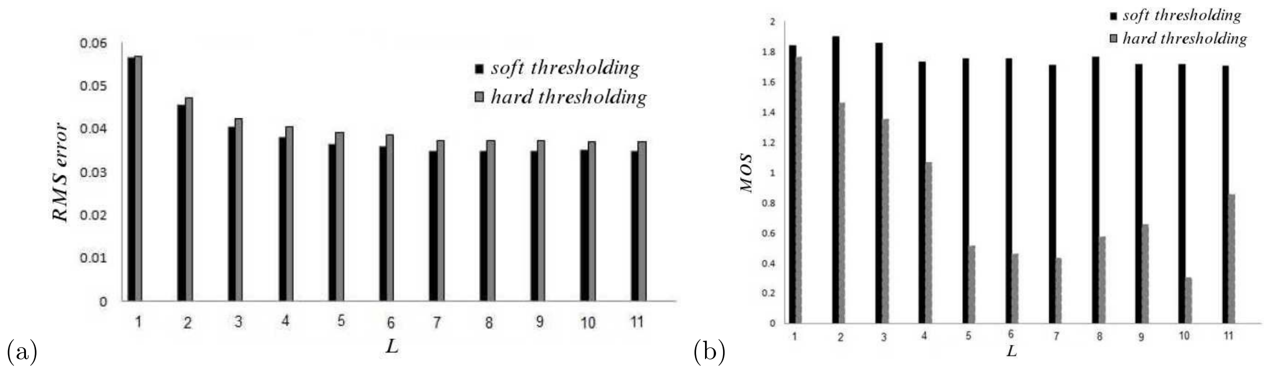


Figure 5. RMS error (a) and MOS (b) depending on the decomposition level for 1D-DWT approach.

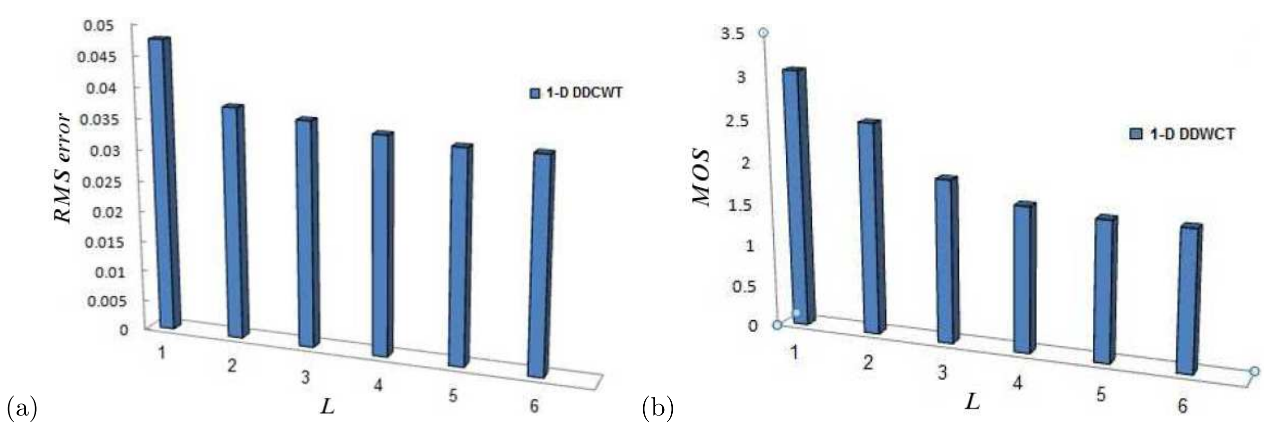


Figure 6. RMS error (a) and MOS (b) depending on the decomposition level for 1D-DDCWT approach.

Analogous results are obtained for 1D-DDCWT approach (Figure 6). Again, the RMS error decreases with the resolution level. Unlike Figure 5b, MOS is significantly higher for the 1D-DDCWT technique confirming advantages of this tool. A good reconstruction is obtained at the first decomposition level with MOS=3.06 for 1D-DDCWT (SNR=0 dB) compared with the best MOS=1.89 for 1D-DWT approach (at level 2 – Figure 5b).

4. CONCLUSIONS

In this study we considered abilities of the 1D-DWT and the 1-D DDCWT for denoising of speech signals corrupted by noise. Signal denoising was performed in the wavelet domain by thresholding of the wavelet coefficients. We showed that the soft thresholding is preferable than the hard thresholding because the soft thresholding provides a reduced error of signal restoration. An important measure of quality of signal denoising is MOS that is estimated based on the perceptual evaluation of speech quality. We demonstrated that the 1-D DDCWT method is more effective even for highly corrupted signals, and this approach can significantly improve the quality of noise reduction. Increased decomposition level does not always provide remarkable improvement of the signal quality. For practical reasons, it is preferable to be limited by 6-th level for the 1D-DWT, while smaller levels typically provide better results in the case of the 1-D DDCWT.

ACKNOWLEDGMENTS

This work was supported by the Russian Science Foundation (Agreement 14-12-00324).

REFERENCES

[1] Mihov, S., Doychev, D. and Ivanov, R., "Practical investigation of specific types of noise signals for the purpose of their suppression in hearing-aid devices," *Proceedings of ICEST-2009* **2**, 399–402 (2009).

[2] Verma, N. and Verma, A. K., "Performance analysis of wavelet thresholding methods in denoiseing of audio signals of some indian musical instrument," *International Journal of Engineering Science and Technology* **4**, 2047–2052 (2012).

[3] Donoho, D. L. and Johnstone, I. M., "De-noising by soft-thresholding," *IEEE Transactions on Information Theory* **41**, 613–627 (1995).

[4] Selesnick, I. W., "The double-density dual-tree DWT," *IEEE Trans. on Signal Processing* **52**, 1304–1314 (2004).

[5] Gonzolez, R. C. and Woods, R. E., [*Digital Image Processing (2nd edition)*], Prentice Hall, New York (2002).

[6] Gonzolez, R. C., Woods, R. E., and Eddins, S. L., [*Digital Image Processing Using Matlab*], Prentice Hall, New York (2004).

[7] Acharya, T., and Ray, A. K., "Image Processing: Principles and Applications," John Wiley & Sons, New York (2005).

[8] Usevitch, B. E., "A tutorial on modern lossy wavelet image compression: foundations of JPEG 2000," *IEEE Signal Processing Magazine* **18**, 22–35 (2001).

[9] Babunaik, B. C. and Anuradha, B., "Speech signal denoising using different wavelet techniques," *Journal of International Academic Research for Multidisciplinary* **2**, 578–590 (2014).

[10] O'Shaughnessy, D., "Enhancing speech degraded by additive noise or interfering speakers," *IEEE Communication Magazine* **27**, 46–52 (1989).

[11] Dewangan, S. and Bhonsle, D., "Image denoising using complex double density dual tree wavelet transform," *International Journal of Engineering Research & Technology* **2**, 906–916 (2013).

[12] Loizou, P. C., [*Speech Enhancement Theory and Practice*], Taylor & Francis Ltd (2007).