A Comparative Performance Analysis of Wavelets in Denoising of Speech Signals

A. K. Verma  
Electrical Engineering Department  
Dayalbagh Educational Institute  
Dayalbagh, Agra, UP, India

Neema Verma  
Electronics and Instrumentation Deptt.  
Hindustan Institute of Technology & Management  
Keetham, Agra, UP, India

ABSTRACT

It is known that data or signal obtained from the real world environment is corrupted by the noise. In most of the cases this noise is strong causing poor SNR and therefore, need to be removed from the desired signal before further processing of signal. Research in the area of wavelets showed that wavelet shrinkage method performs well and efficiently as compared to other methods of denoising. Here, we present a comparative analysis of performance of various types of wavelets i.e. Haar, Db10, Coif5, Bior3.3 and Sym5 in denoising of speech signals in the presence of White Gaussian noise. In the process of denoising, scale dependent Visu Shrink employing universal threshold selection criteria (square-root-log) method for deciding the threshold levels for truncating the wavelet coefficients with soft thresholding is used. Along with the performance evaluation of different types of wavelets, the effect of wavelet decomposition levels is also investigated. The quality of denoised speech signal is expressed in terms of Peak Signal to Noise Ratio (PSNR) as compared to original noiseless speech signal.

General Terms  
Signal Enhancement, Denoising.

Keywords  
Wavelets, PSNR, Denoising, Universal Threshold, Soft Thresholding, Level of Decomposition.

1. INTRODUCTION

Signal degradation by noise is an important phenomenon. Therefore, for various signal processing fields denoising is a key problem. All audio signals are one dimensional; their range of frequency plot displays the experimental results, section V represents the acknowledgements & section VI represents conclusion followed by references. Several types of wavelets have been developed for various applications depending upon their specific properties. Thus it gives immense motivation to observe the effectiveness of these wavelets in denoising of speech signals. This paper presents a comparative analysis of performance with Haar, coif5, Db10, bior5.3 and sym5 wavelets in denoising.

2. DISCRETE WAVELET TRANSFORM (DWT)

The wavelet transform (WT) is a mathematical tool useful in the analysis of signals. Its representation involves the decomposition of the signals in wavelet basis functions \( \hat{\psi}(t) \) given by,

\[
\hat{\psi}_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right), a, b \in \mathbb{R}
\]

Here \( a, b \) are called scale and position parameters respectively. If scales and positions are chosen based on powers of two, so-called Dyadic scales and positions, then analysis becomes much more efficient and just as accurate. It was developed in 1988 by S. Mallat. In this case, wavelet function becomes,

\[
\psi_{m,n}(k) = 2^{m/2} \psi(2^{-m}k - n) \quad m, n \in \mathbb{Z}
\]

in orthonormal basis for \( L^2(\mathbb{R}) \). For a given function \( f(k) \), the inner product \( f(k), \psi_{m,n} \) then gives the discrete wavelet transform as,

\[
DWT(m, n) = 2^{m/2} \sum_{k=\infty}^{\infty} f(k) \psi^*(2^{-m}k - n)
\]

The multi resolution theory given by S. Mallat and Meyer proves that any conjugate mirror filter characterizes a wavelet \( \psi \) that generates an orthonormal basis of \( L^2(\mathbb{R}) \), and that a fast discrete wavelet transform is implemented by cascading these conjugate mirror filters. The wavelet decomposition of a signal \( f(k) \) based on the multi resolution theory can be done using digital FIR filters as shown in figure 1 [11].
For the purpose of denoising the noisy signal \( x(k) \) is first decomposed into \( N \) levels and then level dependent soft thresholding is performed. The threshold values \( \lambda_j \) are calculated by universal threshold (square root log) method proposed by Donoho and Silverman [12] is given by,
\[
\lambda_j = \sigma_j \sqrt{2 \log(N_j)}
\]
Where, \( N_j \) is the length of the noisy signal at \( j^{th} \) scale and \( \sigma_j \) is Median Absolute Deviation (MAD) at \( j^{th} \) scale given by,
\[
\sigma_j = \frac{\text{mad}(x_j)}{0.6745} = \frac{\text{median(|x|)}}{0.6745}
\]
Where, \( c \) represent wavelet coefficients at scale \( j \). This is the optimal threshold in the asymptotic sense and minimizes the cost function of the difference between the function. In the proposed scheme, Soft thresholding is used over hard thresholding as hard thresholding is a “keep or kill” procedure and sometimes, pure noise coefficients may pass the hard threshold and appear as annoying ‘blips’ in the output. While the soft thresholding shrinks coefficients above the threshold in absolute value. The continuity of soft thresholding and some advantages such as it makes algorithms mathematically more tractable. Moreover, Soft thresholding avoids these annoying ‘blips’ in output. The transfer functions of soft and hard thresholding is shown in figure 4.

**Fig. 4: (a) Hard Thresholding, (b) Soft Thresholding**

Threshold determination is an important problem. When denoising, a small threshold may yield a result closer to the input, but the result may still be noisy. A large threshold on other hand produces a signal with a large number of zero coefficients. This leads to a smooth signal. Paying too much attention to smoothness destroys details (Sharp changes) of signal.

**4. EXPERIMENTAL RESULTS**

As samples, five speech signals each of size 10 seconds duration sampled at 8000 samples per second are analyzed for experiment. These five speech samples contain three of English language \( f_{1e}, f_{2e}, f_{3e} \) and two of Hindi language \( f_{1h}, f_{2h} \). For comparing the performance of various wavelets in speech signals following five wavelets, Haar, Db10, Coif5, bior3.3 and sym5 are used. Besides observing the performance of various wavelets, the effect of decomposition level is also investigated. For the performance comparison and measurement of quality of denoising, the Peak Signal to Noise Ratio (PSNR) is calculated between original speech signal \( f(k) \) and denoised speech signal \( f_d(k) \) given by,
\[
\text{PSNR} = 10 \log_{10} \left( \frac{\text{Max}^2}{\text{MSE}} \right)
\]
Where, \( f_{max} \) is maximum value of signal and is given by,

\[
f_{max} = \max(\max(f(k)), \max(f_d(k)))
\]

And MSE is mean Square Error given by,

\[
MSE = \frac{1}{N} \sum_{k=1}^{N} [f_d(k) - f(k)]^2
\]

PSNR values for various speech signals are shown comparatively in figure 5 to figure 8.

<table>
<thead>
<tr>
<th>Signals</th>
<th>Haar</th>
<th>Db10</th>
<th>Coif5</th>
<th>Bior3.3</th>
<th>Sym5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f_1 )</td>
<td>26.18</td>
<td>30.56</td>
<td>30.67</td>
<td>29.80</td>
<td>30.27</td>
</tr>
<tr>
<td>( f_2 )</td>
<td>23.67</td>
<td>27.86</td>
<td>28.19</td>
<td>27.23</td>
<td>27.82</td>
</tr>
<tr>
<td>( f_3 )</td>
<td>28.77</td>
<td>32.53</td>
<td>32.61</td>
<td>31.69</td>
<td>32.10</td>
</tr>
<tr>
<td>( f_4 )</td>
<td>26.09</td>
<td>33.36</td>
<td>33.45</td>
<td>31.66</td>
<td>32.93</td>
</tr>
<tr>
<td>( f_5 )</td>
<td>28.21</td>
<td>30.82</td>
<td>30.84</td>
<td>29.32</td>
<td>30.49</td>
</tr>
</tbody>
</table>

Figure 5 (a, b): Comparison of PSNR at level 2 decomposition

<table>
<thead>
<tr>
<th>Signals</th>
<th>Haar</th>
<th>Db10</th>
<th>Coif5</th>
<th>Bior3.3</th>
<th>Sym5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f_1 )</td>
<td>24.58</td>
<td>28.98</td>
<td>29.72</td>
<td>27.51</td>
<td>28.93</td>
</tr>
<tr>
<td>( f_2 )</td>
<td>22.32</td>
<td>27.32</td>
<td>27.29</td>
<td>26.13</td>
<td>27.07</td>
</tr>
<tr>
<td>( f_3 )</td>
<td>26.97</td>
<td>29.23</td>
<td>29.30</td>
<td>26.75</td>
<td>28.82</td>
</tr>
<tr>
<td>( f_4 )</td>
<td>28.78</td>
<td>30.55</td>
<td>30.89</td>
<td>29.30</td>
<td>30.48</td>
</tr>
<tr>
<td>( f_5 )</td>
<td>26.51</td>
<td>28.13</td>
<td>28.35</td>
<td>26.03</td>
<td>28.05</td>
</tr>
</tbody>
</table>

Figure 6 (a, b): Comparison of PSNR at level 3 decomposition

<table>
<thead>
<tr>
<th>Signals</th>
<th>Haar</th>
<th>Db10</th>
<th>Coif5</th>
<th>Bior3.3</th>
<th>Sym5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f_1 )</td>
<td>21.78</td>
<td>24.05</td>
<td>24.17</td>
<td>21.45</td>
<td>23.97</td>
</tr>
<tr>
<td>( f_2 )</td>
<td>18.96</td>
<td>20.61</td>
<td>21.19</td>
<td>18.42</td>
<td>20.57</td>
</tr>
<tr>
<td>( f_3 )</td>
<td>25.50</td>
<td>26.61</td>
<td>26.76</td>
<td>24.18</td>
<td>26.69</td>
</tr>
<tr>
<td>( f_4 )</td>
<td>26.88</td>
<td>26.91</td>
<td>27.36</td>
<td>25.86</td>
<td>27.43</td>
</tr>
<tr>
<td>( f_5 )</td>
<td>25.14</td>
<td>26.02</td>
<td>26.23</td>
<td>23.81</td>
<td>26.07</td>
</tr>
</tbody>
</table>

Figure 7 (a, b): Comparison of PSNR at level 4 decomposition

<table>
<thead>
<tr>
<th>Signals</th>
<th>Haar</th>
<th>Db10</th>
<th>Coif5</th>
<th>Bior3.3</th>
<th>Sym5</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f_1 )</td>
<td>22.49</td>
<td>26.06</td>
<td>25.77</td>
<td>23.43</td>
<td>25.38</td>
</tr>
<tr>
<td>( f_2 )</td>
<td>19.78</td>
<td>23.10</td>
<td>24.26</td>
<td>20.68</td>
<td>23.41</td>
</tr>
<tr>
<td>( f_3 )</td>
<td>25.84</td>
<td>27.03</td>
<td>27.33</td>
<td>24.77</td>
<td>27.08</td>
</tr>
<tr>
<td>( f_4 )</td>
<td>27.28</td>
<td>27.29</td>
<td>27.76</td>
<td>26.84</td>
<td>27.75</td>
</tr>
<tr>
<td>( f_5 )</td>
<td>25.50</td>
<td>26.44</td>
<td>26.69</td>
<td>24.47</td>
<td>26.51</td>
</tr>
</tbody>
</table>

5. CONCLUSIONS
In this paper, performance of various wavelets in denoising process is investigated along with effect of level of decomposition in the environment of white Gaussian noise. The method for threshold estimation used is level dependent VisuShrink employing universal threshold. The results show that as level of decomposition increases the value of PSNR decreases. Thus lower levels of decomposition can be preferred for such purposes. As seen from the result, performance of Coif5 and Db10 is better than as compared to...
other wavelets in case of all decomposition levels. The Haar wavelet is performing poorer as compared to others also it gives unwanted distortion in reconstructed voice when heard as it is not a smooth wavelet. Overall from the results, it is clear that Coif5 at level 2 is best suitable giving minimum distortion and maximum PSNR as compared with other wavelets.

6. REFERENCES