The Real Time Implementation on DSP of Speech Enhancement Based on Kalman Filter and Wavelet Thresholding

Saoud Safa, Bennasr Mouhamed and Cherif Adnen*

Science Faculty of Tunis, University of Tunis El-Manar, Laboratory of Analysis and Processing of Signals and Electric and Energy System, 1060 Tunis, Tunisia;
Saoud_safa@yahoo.fr, bennasr.mouhamed@gmail.com, adnen2fr@yahoo.fr

Abstract

Background/Objectives: The speech signal is an attractive research issue that solves signal noises. For this reason, the paper main goal is to effectively ensure speech quality. An accurate Optimized Speech Enhancement Algorithm (OSEA) has been proposed and implemented to improve speech intelligibility and quality. The proposed approach was developed and implemented through a Discrete Wavelet Transform (DWT) and kalman filter methods. To assess speech quality, various noisy types for each SNR level were used. The obtained results were compared with a wavelet based speech enhancement using objective assessments like SNR, NRMSE and PESQ. Indeed, the obtained results demonstrated a better speech quality. All prepared tests have been implemented on a TMS320C6416 fixed-point digital signal processor which appears to be suitable platform for real-time requirements.

Keywords: Digital Signal Processor, Discrete Wavelet Transform, Kalman Filter, Speech Enhancement, Real-Time

1. Introduction

The degradation of signals by noise is an ever-present problem. In fact, the background noise damages the intelligibility and quality of speech signals resulting in a harsh drop in performance of speech uses such as: telecommunications, sound recording and teleconferencing. These applications require the suppression of noise and improve the clean signal from noisy signal. Speech enhancement is the main technique in signal processing domain. It reduces noise and improves speech quality and intelligibility.

Over the three decades, noise reduction from speech signals is a very motivating area of researchers in speech processing.

The literature is enriched by many techniques have been proposed for this purpose such as the spectral subtraction approach, the signal subspace approach, adaptive noise canceling, Wiener filter, Kalman filter, notch filter and wavelet methods. The efficiency of these methods depends on the intelligibility and the quality of the enhanced speech signal. The augmentation in the SNR (Speech signal-to-Noise Ratio) is the aim of most methods.

Spectral Subtraction (SS) is the first approach for enhancing speech corrupted by additive noise. It approximates the spectrum of the original speech signal via the subtraction of the magnitude spectrum of the estimated noise from the noisy speech signal magnitude spectrum, whereas keeping the phase spectrum of the noisy speech signal. The shortcoming of this approach is the residual noise.

Another approach is the signal subspace technique which is used for denoising speech signals corrupted by colored additive noise or uncorrelated noise. The main idea of this algorithm is based on the fact that the vector

*Author for correspondence
space of the noisy speech signal is divided into a signal plus noise subspace and an orthogonal noise subspace. Processing is performed just on the vectors in the signal plus noise subspace, although the noise subspace is detached firstly. In fact, the vector space decomposition of the noisy signal is effected via the singular value decomposition or the Karhunen-Loeve Transform (KLT) on the speech signal.

Wavelet base methods are widely used for denoising speech signals corrupted with additive background noise. Although usual approaches often eliminate noise by low-pass filtering that distorts the sharp features in the speech signal, wavelet based methods prove good performance for a large variety of signals. While the speech enhancement based on wavelet transform has been applied in many works. There are numerous issues to be determined for a successful noise suppression application to be successful for real-time speech signals corrupted by additive noise. Previous approach used the DWT thresholding for speech enhancement of real-time speech signals corrupted by Gaussian white noise.

In this context, we present in this paper a combination of discrete wavelet transform and kalman filter used for single channel speech enhancement in a noisy environment. This approach was implemented on a fixed point digital signal processing (DSP) for Texas Instruments to improve the speech quality and intelligibility. Initially, we apply kalman filter to filter the noisy speech signal. Secondly, we conduct the discrete wavelet transform to the filtered frames. Thirdly, in order to track the variation of noisy speech, we use soft thresholding to truncate the noisy discrete wavelet coefficients according to the input noise level, and then we apply the inverse discrete wavelet coefficient to obtain the enhanced signal. In fact the proposed speech enhancement method is properly adapted to reduce the white and the colored noise from the noisy speech. The results of applying the proposed method with TIMIT database are compared with the DWT approach.

This paper is organized as follows: The discrete wavelet transform is presented in Section 2. The proposed speech enhancement method is presented in Section 3. A comparative study of performances is performed using some objective criteria are presented in Section 4. In Section 5, a real-time implementation methodology is presented. Finally, Section 6 concludes this paper.

2. Discrete Wavelet Transform

The Discrete Wavelet Transform (DWT) is a great tool used for speech enhancement and image compression. It gives satisfactory information both for analysis and reconstruction. The DWT decomposed the signal at different frequency bands with different resolutions. The general shape of DWT:

\[
f(t) = \sum_{j=1}^{J} \sum_{k} d_j(k) \psi_j(t) + \sum_{j} a_j(k) \phi_j(t)
\]

where, \( \psi_j(t) \) is the mother wavelet and \( \phi_j(t) \) is the scaling function at level \( L \). The approximation and detail at level \( j \) are defined as:

\[
a_{j+1}(k) = \sum_{m} h_0(m-2k)a_j(m)
\]

\[
d_{j+1}(k) = \sum_{m} h_1(M-2k)a_j(m)
\]

where \( h_0(\cdot) \) and \( h_1(\cdot) \) are the wavelet filters.

Indeed, the recursive pyramid algorithm was developed to compute the DWT that consists in successively decomposing the original signal into low frequency components and high frequency components correspondingly by the mean of low-pass filter analysis and the high-pass filter analysis. The high frequency components (details coefficients) are not analysed any further, however the low frequency components (approximation coefficients) are next decomposed into new approximation coefficients and detail coefficients via the following equations:

\[
Y_{\text{low}}[k] = \sum_{n=0}^{N-1} x(n)h_0(2k-n)
\]

\[
Y_{\text{high}}[k] = \sum_{n=0}^{N-1} x(n)h_1(2k-n)
\]

The three levels DWT decomposition is illustrated in figure 1.
The original speech signal is reconstructed using the inverse discrete wavelet transform (IDWT). Firstly, the approximation and detail coefficients are up-sampled through a factor of 2, and next, in that order filtered by the low-pass analysis filter \( g_1(z) \) and high-pass synthesis filter \( g_0(z) \):

\[
g_1(n) = h_1(N - n - 1) \tag{6}
\]

\[
g_0(n) = h_0(N - n - 1) \tag{7}
\]

Figure 2 shows the three levels DWT reconstruction:

**Figure 2.** Three level DWT synthesis.

### 3. The Speech Enhancement Algorithm

The block diagram of the proposed enhancement system is illustrated in Figure 3. The different steps of our system are detailed in the following paragraphs.

**Figure 3.** Block diagram of the speech enhancement method.

The Kalman filter presents a good solution to the linear MMSE difficulties for the stochastic method by attaining an optimal prediction of the clean speech.

Then, the DWT is applied for each frame, in this step; we will select the mother wavelet, the decomposition level. Numerous criteria are considered for choosing an optimal mother wavelet such as: minimizing the error variance and maximizing the signal to noise ratio between the original signal and the reconstructed signal. In general, the selection of the optimal mother wavelet is based on the average energy concentrated in the approximation part of the wavelet coefficients.

After performing the transformation method, denoising involves thresholding which is the most important step in a transform based denoising; it consists of eliminating the coefficients of the DWT transform inferior to a given threshold. There are different methods of thresholding, such as the hard and the soft thresholding which are the frequently used methods.

In this work, we have used the global threshold proposed by Donoho and Johnstone:

\[
\lambda = \sigma \sqrt{2 \log(n)} \tag{8}
\]

Where \( n \) designates the noisy signal length and \( \sigma \) represents the estimate of the noise standard deviation, given by:

\[
\sigma = \text{MAD} / 0.6745 \tag{9}
\]

With the MAD is the absolute median estimated on the first scale.

The Soft Thresholding “(10)” was carried out on the DWT coefficients before reconstructing the signal.

\[
\text{THR}_{\text{soft}}(Y, T) = \begin{cases} 0 & \text{if } |Y| \leq T \\ \text{sign}(Y)(|Y| - T) & \text{if } |Y| > T \end{cases} \tag{10}
\]

Finally, we conduct inverse DWT to get the reconstructed speech signal by grouping back the approximate and detailed coefficients which are currently noise free.

### 4. Evaluation Criteria

To evaluate the performance of our proposed speech enhancement, we have used objective and subjective criteria.

#### 4.1 Objective Evaluation

The Objective measures are based on mathematical comparison between the original and processed speech signals. The measure of the signal to noise ratio, SNR is one of the most widely used.

- The SNR can be expressed as follows:

\[
\text{SNR} = 10 \log_{10} \left[ \frac{\sum s^2(n)}{\sum (s(n) - \bar{s}(n))^2} \right] \tag{11}
\]

Where \( s(n) \) presents the clean signal and \( \bar{s}(n) \) presents the enhanced signal.
The NRMSE is expressed as follows:

\[ \text{NRMSE} = \sqrt{\frac{\sum_n (s(n) - \tilde{s}(n))^2}{\sum_n (s(n) - \mu s(n))^2}} \]  

(12)

Here, \( s(n) \) is the speech signal, \( \tilde{s}(n) \) is reconstructed speech signal and \( \mu s(n) \) is the mean of speech signal.

Perceptual Evaluation of Speech Quality

PESQ (Perceptual Evaluation of Speech Quality) is an objective quality measure that is approved as the ITU-T recommendation P.862. It is a mean of objective measurement conceived to predict the results of a subjective Mean Opinion Score (MOS) test. Particularly, PESQ was developed to model subjective tests commonly to assess the voice quality by human beings.

4.2 Subjective Evaluation

Mean Opinion Score (MOS)

Several methods for subjective assessment are used in literatures which are described in ITU-T Recommendation P.830. The most commonly used evaluation method is the Mean Opinion Score (MOS) in which a group of listeners are asked to perceptually evaluate enhanced speech signals from noisy signal with various background noises. We present in the table below the correspondence between the scores and the quality judgments.

| Rating | MOS |
|--------|-----|
| 1      | Bad |
| 2      | Poor|
| 3      | Fair|
| 4      | Good|
| 5      | Excellent|

5. Test And Results

In this section, a MATLAB program has been developed to implement the speech enhancement system based on kalman filter with DWT. The database was taken from the TIMIT database, sampled at 16 kHz and recorded by female voice. For this purpose, the used mother wavelet is “db10”, three decomposition levels and soft thresholding was implemented in various noises taken from Noisex-92 database: White noise, F16 cockpit noise, Factory noise and Pink noise with different values of Signal to Noise Ratio (SNR) ranging from 0dB to 15dB were used. To evaluate the efficiency of the developed algorithm, a comparative study between the wavelet based enhancement and the proposed system is performed using three objective criteria: SNR, the PESQ and NRMSE. Also, we used the time domain waveforms and spectrograms for the test and the evaluation.

Throughout the figures bellows, it is observed that the proposed system rates are better than those obtained by DWT based speech enhancement.

Figure 4 presents the SNR of the enhanced algorithms. It’s clear that the suggested algorithm outperforms DWT method in different noise conditions.

A figure 5 depicts the PESQ scores for the denoised speech. Clearly, the performances of the proposed approach using DWT and Kalman filter is better than DWT with different input SNR ranging from 0dB to 15dB.
Figure 6 reveals the superiority of the proposed algorithm. In fact, it gives the lowest NRMSE.

![Figure 6](image_url)

**Figure 6.** Comparison of NRMSE for DWT and proposed approach.

A figure 7 depicts respectively the time evolution and spectrograms of the clean speech signal, the noisy speech signal and enhanced speech signals by DWT algorithm and our proposed approach when signal is corrupted by factory noise at SNR=10dB. The figure shows that the proposed algorithm reduces interfering signals more than the DWT method.

![Figure 7](image_url)

**Figure 7.** The time domain and spectrograms results for (a) clean speech; (b) noisy speech corrupted by factory noise (SNR = 10 dB); (c) Enhanced speech by DWT (SNR =9.77 dB); (d) enhanced speech by the proposed method (SNR = 12.45 dB).

6. Real-Time Implementation Methodology

Real-time test has a great value, particularly for audio applications that have strict timing constraints such as audio streaming. In real-time application, the input signal and the generated output are processed continuously which explicates that the mean processing time per sample is lesser than the sampling period. Thus, we have tested our proposed approach on a flexible platform which is suitable with the particularity of our application. For this reason we have used a developed starter kit containing a DSK board based on DSP-TMS320C6416 and the software tool (Code Composer
6.1 DSK C6416 Overview

The DSK C6416 board incorporates a fixed point digital signal processor TMS320C6416 that works with a clock frequency of 1GHz. Furthermore, it is equipped with audio codec TLV320AIC23 (AIC23) which offers analog-to-digital conversion (ADC) and digital-to-analog conversion (DAC) functions among a selecting sampling rate ranged of alternative settings from 8 to 96 kHz. As indicated in the figure 8 bellow the DSK board has four connections which provide analog inputs and outputs: A microphone input port, a line in port, a line out port, and a headphone port. The DSK board includes 16 MB (megabytes) of synchronous dynamic RAM (SDRAM) and 512 kB (kilobytes) of flash memory. The TMS320C6416 is based on the Very Long Instruction Word (VLIW) architecture that is suitable for numerical exhaustive algorithms. The internal program memory is structured so that a total of eight instructions can be fetched every cycle.

Figure 8 shows the Spectrum Digital DSK board.

6.2 Rapid Prototyping Technology

For a rapid prototyping of real-time applications on DSP processor, we have employed the Embedded Target for C6000 DSPs Platform and RTW (Real-time Workshop) which convert the Simulink model into efficient C code especially for C6000 processors. MATLAB Link for CCS Development Tools interface with TI CCS (code composer studio) to generate an executable which is loaded into the DSK-C6416.

Figure 9 shows the designed Simulink model for implementing the denoising system based on DWT and kalman filter on TMS320C6416.

In order to validate the performances of the proposed speech denoising algorithm, we have conduct informal listening test using the MOS (Mean Opinion Score) method to evaluate the difference between the residual noise characteristics of the enhanced speech. 4 volunteer’s listeners have evaluated the quality of the sentences generated from the headphone connected on the DSP board. The tested sentence is corrupted by (White, Babble) noises at 5 dB. Figure 11 presents the results of real time denoising system.

The MOS values are improved after using kalman filtering with wavelet transform which proves the efficiency of our approach.
7. Conclusion

In this paper, a new speech enhancement algorithm using discrete wavelet transform combined with a kalman filter has been presented. The proposed approach proves its reliability to improve the speech intelligibility without affecting the signal quality referring to the performance evaluation such as SNR, NRMSE, speech quality using the PESQ scores and the time domain waveforms. Finally, the real-time test of speech denoising has been successfully implemented in TMS320C6416 platform and reveals that the proposed algorithm has significantly improved the speech intelligibility.

As a future work, we will tend to implement our system in other real platforms as FPGA and Raspberry-PI to enhance the speech signal devoted to recognition application.

8. References

1. Almajai, Milner B. Visually derived wiener filters for speech enhancement. IEEE Trans. Audio, Speech, and Language Processing. 2011 Aug; 19( 6): 1642–51.
2. Ben Jebara S. A perceptual approach to reduce musical noise phenomenon with wiener denoising technique. IEEE Int. Conf. Acoustics, Speech and Signal Processing, (ICASSP).2006 May; 3: 3.
3. Chen J, Benesty J, Huang Y, Doclo S. New Insights into the noise Reduction Wiener Filter. IEEE Trans Audio, Speech Language Processing. 2006 July;14( 4):
4. Ramesh Babu G. Modified Kalman Filter-based Approach in Comparison with Traditional Speech Enhancement Algorithms from Adverse Noisy Environments. International Journal on Computer Science and Engineering (IJCSE). 2011 Feb; 3( 2):
5. GohZ, Tan KC,Tan BTG. Kalman-Filtering Speech Enhancement Method Based on a Voiced-Unvoiced Speech Model. IEEE Trans. Speech, Audio Processing. 1999 Sep;7(5):
6. Claesson, Rosholm A. Notch filtering of humming GSM mobile telephone noise. in Proc. IEEE Int. Conf. Communications. Signal Processing (ICSP’05); Bangkok, Thailand: 2005 Dec.p. 1320–23.
7. Shavan MS, Shavan MN, Gaikwad MS. Studies on implementation of wavelet for denoising speech signal. International Journal for Computer Applications.2010 June;3(2):1–7.
8. Tabibian S, Akbari A, Nasershari R. A new wavelet thresholding method for speech enhancement based on symmetric kullback- leiblerdivergence. in Proc. 14th Int. CSI Computer Conference, CSICC); 2009 Oct.p. 495–500.
9. Hu Y, Loizou PC. Speech enhancement based on wavelet thresholding the multitaper spectrum. IEEE Transactions on Speech and Audio Processing.2004; 12(1):59–67.
10. Slavy G, Mihov, Ratcho M, Ivanov, Angel N. Popov. Denoising speech signals by wavelet transform. Annual Journal of Electronics.2009;69–72.
11. MalAYERI. Noise speech wavelet analyzing a special time ranges. In Int. Conf. Advanced Communication Technology (ICACT).2010 Feb;1: 525–8.
12. Boll SF. Suppression of acoustic noise in speech using spectral subtraction. IEEE Trans. Acoustics, Speech, Signal Processing. 1979 April;27( 2):
13. Sunitha P, Sekhar T. Speech signal enhancement using wavelet threshold methods. International Journal of Ethics in Engineering and Management Education.2014 March;1(3):59.
14. Donoho D, Johnstone IM, Kerkycharian G, Picard D. Wavelet shrinkage: Asymptota. Journal of the Royal Statistical Society. Serie B.1995;.57:3019–69.
15. Hu Y, Loizou FC. Speech enhancement based on wavelet thresholding the multitaper spectrum. IEEE Trans Audio, Speech, Language Processing.2004 January;12(1):59–67.
16. Shabeera Bee S, Pramod JT, Jilani SAK. Real time speech denoising using simulink and beagle bone black. International Journal of Core Research in Communication Engineering. 2015 April; 2(2):
17. Mallat S. Multiresolution approximations and wavelet orthonormal bases of L2(R). Trans. Amer. Math. Soc. 1989;315: 69–87.
18. Hu Y, Loizou FC. Perceptual Evaluation of Speech Quality (PESQ):An objective method for end-to-end of speech quality assessment of narrowband telephone network and speech codecs. ITUT Recommendation. P. 862, ITU; 2000.
19. Garofolo J, Lamel L, Fisher W, Fiscus J, Pallett D, Dahlgren N, et al . TIMIT Acoustic–Phonetic Continuous Speech Corpus: Linguistic Data Consortium; 1993.
20. Math Works. Real-Time Workshop for Use with SIMULINK∥ User’s Guide. Version 6; June 2004.
21. Texas instruments. TMS320 DSP/BIOS. 5(42): User’ Guide, spru423I, Août 2012.