Adaptive Speech Noise Cancellation using Wavelet Transforms

Peyman Goli\textsuperscript{a} * and Khalil Vaez\textsuperscript{b}

\textsuperscript{a} Department of Electrical and Computer Engineering, Khavaran Higher Education Institute, Mashhad, 9184168619, Iran
\textsuperscript{b} Department of Electrical and Computer Engineering, Khavaran Higher Education Institute, Mashhad, 9184168619, Iran

* Corresponding author. Tel.: +98-513-5230690; email: p.goli@profs.khi.ac.ir

\textbf{ABSTRACT}

A new method is presented for single-channel noise reduction. The proposed method is a combination of wavelet transform and adaptive filter. The speech signal is initially decomposed by wavelet transform into frequency sub-bands and then the noise is removed by an adaptive filter based on the LMS algorithm. Noise is removed separately for approximation and details coefficients. In the adaptive filter, the noisy speech sub-band is considered as the desired signal and the delayed one as the filter input. Filter output is the improved sub-band. After noise reduction, the sub-bands are reconstructed to achieve improved speech. The improved speech quality was evaluated by two objective measures, and also was compared with the unprocessed noisy speech and improved speeches by wavelet and adaptive filtering techniques. The results show that in the noises with low autocorrelation (such as white noise) the proposed algorithm shows much higher performance than other methods for improving speech quality in all noisy conditions.

\textbf{Keywords:} Speech enhancement, noise reduction, adaptive algorithm, wavelet transform.

1. Introduction

Speech is the most important means of communication between individuals and necessary for understanding issues; for this reason, improved quality of speech has been of interest to researchers for its many uses. Various noises, including background noise and ambient noise contaminate the speech signal. Therefore, methods are needed for reducing or removing noise from speech signal. Noise reduction of speech has many applications in automatic speech recognition (ASR), speaker identification, wireless communication, speech improvement in the hearing aids, and improving speech quality recorded in high-noise and highly reflective environments.

One of the oldest ways of speech noise reduction is the spectral subtraction that was introduced in 1979 [1]. A spectral subtraction is used to retrieve the signal power spectrum contaminated by the additive noise by reducing the estimated noise power spectrum from the noisy signal power spectrum [2]. Short-time Fourier transform (STFT) in which the Fourier transform is calculated within short-time segments of speech is introduced to adapt more spectral differential methods with non-stationary nature of speech. In this way, the signal is transmitted to the two-dimensional time-frequency [3]. The problem with this transmission is that window length remains constant at different frequencies, while speech signals require more flexibility. To overcome the problem, wavelet transform is presented. The theory of the noise reduction in the wavelet domain was introduced by Donoho (1995) [4]. The noisy signal is transmitted from the wavelet filter bank (including low-pass and high-pass filters). The output of the filters can be decomposed again and continue to reach to different levels and a desired signal. Then, the noise is removed by soft and hard thresholds, etc. [5]. Recently, comparative threshold methods have been used to remove noise from speech signal by using wavelet transform [6].

The use of adaptive filters is another powerful method for removing noise from speech, which provides statistical model-based methods. In these algorithms, first presented in 1949, a clean signal is estimated based on the optimization of a measure (such as the Wiener measure, or the statistical mean of output error power). In the adaptive filters, the output error power can be minimized by changing the filters weights. Speech improvement is divided into two groups of single-channel methods (i.e., single microphone) and multi-channel methods (i.e., dual-channel, microphone array and distributed microphones) based on the number of available microphones in the adaptive methods. In single-channel methods there is only one microphone to receive speech signals. Thus in this case, the only input of speech enhancement system is a noisy speech signal [7, 8].

In multi-channel methods, in addition to noisy signal, the ambient noise can be obtained by placing different microphones. Given the use of more microphones, these methods have more efficiency in noise reduction than single-channel methods, but the limitation of microphone usage or the condition of having noisy speech alone make single-channel noise reduction methods more useful than multi-channel methods [9].

There are various adaptive algorithms (i.e., search methods) to search optimal weights in the adaptive filters. The least-mean squares (LMS) and the normalized least-mean squares (NLMS) are used for their simple computation and implementation in a wide range of speech processing methods. The recursive least square (RLS) is one of the algorithms used to better represent the convergence of adaptive filters. The implementation of this algorithm is accompanied by complex calculations [10]. Recently, an adaptive filter has been developed that provides a good...
exchange between complexity and convergence rate by using two adaptive algorithms called fast algorithm for repetitive pattern and fast Euclidean search [11].

In the proposed method, a new method for single-channel noise reduction by combining wavelet transform and adaptive filters is introduced, and its efficiency in improving speech quality is evaluated. The following paragraphs first describe wavelet transform and adaptive algorithms for noise reduction of speech signal. Then, the proposed method is discussed and implemented. Finally, the proposed method is evaluated and compared with the wavelet transform and adaptive filter methods.

2. Wavelet Transform for Speech Noise Reduction

One of the ways for noise reduction of speech signal is to use wavelet transform. The wavelet filter bank consists of a number of low- and high-pass filters that the low-pass filters output are referred to as approximation coefficients and the high-pass filters output as details coefficients. The number of signal samples in the output can be reduced by down-sampling and the reconstruction of signal is performed by up-sampling. The decomposition can be continued for approximation and details coefficients until reaching the desired signal at different levels. In discrete wavelet transform (DWT) the scales and positions are in the form of integer powers of 2. By applying a signal such as \( s(n) \) to wavelet transform input, the equations of wavelet transform output coefficients are as follows:

\[
A(n) = \sum_{k=1}^{N} s(n) h(-k + 2n),
\]

where \( A(n) \) is the approximation coefficients, \( N \) is the length of signal and the low-pass filter impulse response is shown by \( h(n) \). For the details coefficients we also have,

\[
D(n) = \sum_{k=1}^{N} s(n) g(-k + 2n),
\]

where \( D(n) \) is the details coefficient and the response high-pass filter impulse response is shown by \( g(n) \).

Given the nature of additive noise signal, which is usually distributed in the dense form in the speech sub-bands, it is possible to remove noise samples by using the different thresholds on the details coefficients in the noisy speech. Accordingly, different thresholds are presented to eliminate the noise components. In Figure (1), signal decomposition and reconstruction are shown for wavelet transform.

Figure 1. a. Wavelet transform decomposition; b. Wavelet transform reconstruction

3. Adaptive Methods for Speech Noise Reduction

The use of adaptive filters is another method for noise reduction of speech signal. In the adaptive filters, the filter parameters are changed with the changing speech conditions and environment so that the error measure (i.e., such as the mean expectation of the error power) is minimized. There is also a desired signal in the adaptive filters in addition to the input and the output signals. The desired signal is the signal in which the filter output should be correlated as much as possible. In other words, the filter coefficients should be changed in such a way as to maximize the correlation between the output process and the desired process. In the adaptive filters, the filter coefficients are determined by an adaptive algorithm to achieve the optimal measure. What is important is to provide an adaptive algorithm that can reach optimal amount with the least computational complexity, the least error and the minimum running time of the algorithm. Algorithm accuracy during its running time can be set with the size of the steps to move to the optimal point. The small step size increases the accuracy and reduces the error, while it reduces the speed of the algorithm. The selection of a large step size while speedup the algorithm will also increase the convergence error. Least Mean Square (LMS) is one of the most important adaptive algorithms formed based on the method of steepest descent (i.e., gradient method) [10]. The block diagram of the single-channel noise reduction by the adaptive filter is shown in Figure (2). \( s(n) \) is the noisy speech signal, \( s'(n) \) is the delayed noisy speech that is the filter input, \( d(n) \) is the desired signal, which is the noisy speech, \( e(n) \) is the error signal and \( y(n) \) is the (estimated) output signal. The value of \( T \) is calculated according to the size of the pitch frequency for a frame [12]. The filter coefficients are estimated so that the filter output adapts to the desired signal segment (i.e., noisy speech) that has a correlation in the input. Given that additive noise usually has a lower autocorrelation with the speech signal, due to the delay in the input of the adaptive filter, it is expected that the filter output be adapted to the desired speech signal segment [10].

![Figure 2. The block diagram of speech noise reduction by adaptive filter along with LMS algorithm.](image)

The adaptive LMS algorithm is as follows to determine the filter coefficients,

\[
w(n) = w(n-1) + \delta s'(n)e(n),
\]

where \( w(n) \) is the filter weight at the \( n^{th} \) moment and \( \delta \) is the motion step. The larger is \( \delta \), the greater is the adaptation speed and steady-state error power. Therefore, \( \delta \) should be selected so that a compromise can be made between them.

4. The Proposed Method

In the proposed method, the frequency sub-bands of noisy speech are applied to the adaptive filter. That is, at first, the noisy speech signal is decomposed by the wavelet transform to approximation and details coefficients, and then each of the coefficients is separately applied to the adaptive filter to remove the noise. Thus instead of optimizing the correlation between the noisy signal and the delayed signal to remove the noise in the adaptive filter, the correlation between the frequency sub-bands of the noisy signal and the delayed ones is optimized. Given that the proposed method is a combination of wavelet transform and
adaptive filter, it will be indicated by WTAF (Waveform Transform Adaptive Filter).

In Figure (3) the block diagram of WTAF method is shown. The noisy speech signal \( s(n) \) is decomposed by wavelet transform. In the wavelet transform output, the number of signal samples is reduced by down-sampling, and the approximation coefficients \( s_A(n) \) and the details coefficients \( s_D(n) \) are generated. In addition, the delayed noisy speech signal \( s'(n) = s(n - T) \) is decomposed by a discrete wavelet transform (DWT) and the approximation coefficients \( s'_A(n) \) and the details coefficients \( s'_D(n) \) are generated. Separate adaptive filters are considered for approximation and details coefficients. Delayed coefficients are considered as inputs and non-delayed coefficients are considered as the desired signal of the corresponding filters. The LMS adaptive algorithm is used to estimate the weights of filters. The signals \( y_A(n) \) and \( y_D(n) \) are respectively the adaptive filters output of approximation and details coefficients. In the final stage, the improved signal is retrieved by the inverse discrete wavelet transform (IDWT).

Figure 3. The block diagram of the proposed method for noise reduction of speech signal.

5. Implementation

According to experiments, the best result (i.e., the best convergence rate and accuracy) was obtained in the step size \( \delta = 10^{-7} \) and the filter order 20, using the Haar wavelet. The results of the implementation of WTAF for a sentence with an approximate length of 2 seconds in the white noise with the signal-to-noise ratio (SNR) of 0 dB are shown in Figure (4). Presentation in the time domain and the spectrogram of original clean speech, noisy speech and improved speech are displayed in Figure (4). The spectrogram of the improved signal represents a noise reduction of speech signal, as shown in the Figure.

In addition, the proposed method was implemented according to the number of levels of wavelet transform in the white noise whose results are shown in Figure (5). The improvement of speech quality is evaluated by PESQ (Perceptual Evaluation of Speech Quality) measure for 1 to 4 wavelet transform levels at SNR = −10, −5, 0 and 5 dB. As the results show, the highest level of speech quality improvement is achieved for two wavelet decomposition levels.

6. Evaluation

To evaluate the performance of the proposed algorithm for improving speech quality, 30 English sentences spoken by three men and three women are used from the IEEE sentence database. All sentences are resample to 8 kHz. The approximate length of sentences is 2.5 s. PESQ (Perceptual Evaluation of Speech Quality) and RMSE (Root Mean Squared Error) are used to evaluate the quality of noisy speech improved by the WTAF algorithm.

Figure 4. Presentation in the time domain of the clean speech signal, noisy speech and speech improved by WTAF with their spectrogram in the white noise and SNR = 0 dB.

The speech quality evaluation measure of PESQ was proposed by Rix in 2002. Clean and noisy speech signals are the inputs of the measure and the output is a number between zero and 4.5. The higher is the number, the much more similar is the improved signal to the clean signal [13]. The RMSE shows the difference between the improved signal and the clean signal [14]. This measure is determines as follows,

\[
RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^{N} (x(n) - \hat{y}(n))^2}
\]

where \( x(n) \) is the clean signal, \( y(n) \) is the improved signal, and \( N \) is the signal length. In this measure, the less root square difference between the clean signal and the improved signal indicates the better speech improvement.

Figure 5. PESQ objective measure scores for different levels of wavelet transform at SNR = −10, −5, 0 and 5 dB in the white noise in WTAF.

To evaluate the improved speech quality, WTAF was compared with noise reduction by using the wavelet transform (WT) and the single-channel noise reduction by using an adaptive filter (AF) and unprocessed noisy speech (briefly expressed as unprocessed). In the WT method, two levels of Haar wavelet transform with soft threshold were used [4]. In the AF method, according to Figure (2)
the single-channel noise reduction was used with delay in the filter input [12]. To evaluate the performance of the proposed algorithm according to the autocorrelation of additive noises, the white noise which has the lowest autocorrelation and the pink noise which has more autocorrelation were used. The results of the evaluation by the two objective measures for $\text{SNR} = -10, -5, 0, 5$ are shown in Figures (6) and (7). As the results illustrate, in both evaluation measures of the speech quality in all SNRs, the WTAFT in the white noise has a much better performance than noisy unprocessed speech and other methods, but in the pink noise which has more autocorrelation, the WTAFT and WT, which remove the noise based on the degree of autocorrelation, are not better in improving speech quality.

![Figure 6](image)

**Figure 6.** The performance evaluation of the proposed method (WTAFT) in PESQ and its comparison with wavelet transform (WT), adaptive filter (AF) and unprocessed noise speech for white and pink noises in $\text{SNR} = -10, -5, 0, 5$ dB.

![Figure 7](image)

**Figure 7.** The performance evaluation of the proposed method (WTAFT) in RMSE and its comparison with wavelet transform (WT), adaptive filter (AF) and unprocessed noise speech for white and pink noises in $\text{SNR} = -10, -5, 0, 5$ dB.

### 7. Conclusion

A new method was presented for single-channel noise reduction by combining wavelet transform and adaptive filter. In this approach, instead of increasing the correlation between the noisy speech signal (i.e., as the desired signal) and the delayed signal (i.e., as the input signal) by an adaptive algorithm, the correlation between their frequency sub-bands are increased to reduce noise. For this purpose, first, the noisy speech signal and the delayed signal are decomposed by wavelet transform to frequency sub-bands, and then the LMS adaptive algorithm is separately used for adapting approximation and details coefficients with their delayed coefficients. The evaluation results of the proposed algorithm performance show that noises with low autocorrelation (i.e., such as white noise), show a very high-speech quality improvement compared to the previous methods of noise reduction, wavelet transform and adaptive filters. While in noises with higher autocorrelation (pink noise) this algorithm does not provide a good performance.

### References