Combined Spectral Subtraction and Wiener Filter Methods in Wavelet Domain for Noise Reduction

Farid Ykhlef, A. Guessoum, and D. Berkani

Abstract— **The corruption of speech due to presence of additive background noise causes severe difficulties in various communication environments. This paper presents a novel noise reduction technique based upon a combination of cascaded spectral subtraction and Wiener filter methods in wavelet domain. The scheme's performance is illustrated by experiments in a noisy car environment, in comparison with spectral subtraction and Wiener filter.**

I. Introduction

NOISE reduction is a subject to research in many different fields [1]. Depending on the environment, the application, the source signals, the noise, and so on, the solutions look very different. Here we consider noise reduction for speech signals, and concentrate on common acoustic environments such an office room or inside a car. The goal of the noise reduction is to reduce the noise level without distorting the speech, thus reduce the stress on the listener and ideally increase intelligibility [2].

There are many different ways to perform the noise reduction. Principally, the solutions can be split in two classes: single and multi-microphone systems. Whereas multi-microphone systems exploit the spatial properties of speech and noise, a single-microphone system usually rely on the temporal characteristics. A fundamental requirement for most single-microphone systems is that speech and noise are additive and results of uncorrelated statistical processes, and that the spectral characteristics of the noise changes markedly slower than those of the speech.

The noise reduction method discussed in this paper is a single channel method based on converting successive short segments of speech into the frequency domain. In the frequency domain, the noise is removed by adjusting the discrete frequency "bins" on a frame-by frame basis, usually by reducing the amplitude based on an estimate of the noise. The various methods (differentiated by the suppression rule, noise estimate and other details) are collectively known as, Short-Time Spectral Amplitude (STSA), Spectral Weighting, or Spectral Subtraction methods [3] [4]. Other approaches have been reported in the literature for speech noise reduction, such as the signal subspace approach in [5] and the human auditory system model-based approaches in [6] and [7]. This paper addresses the problem of noise reduction of additive background noise in speech based on the combination in cascade of the spectral subtraction and the Wiener filter in wavelet domain. The rest of this paper is organized as follows; Section II and Section III give a review of the noise reduction strategies based on spectral subtraction and Wiener filter. Section IV discusses and develops the new proposed scheme. Results are presented in Section V. Section VI gives the conclusion.

II. Noise Reduction by Spectral Weighting

Spectral weighting means that different spectral regions of the mixed signal of speech and noise are attenuated with different factors. The aim of this process is a speech signal which contains less noise than the original one. Besides requiring a minimal distortion of the original speech, it is also important that the residual noise, i.e. the noise remaining in the processed signal, does not sound unnatural. The spectral weighting is usually performed in a transformed domain (the frequency domain). A common transform is the Fourier transform which provides an equidistant frequency solution.

Let $s(t)$ and $n(t)$ denote speech and uncorrelated additive noise signals, and let $x(t)$ represent the noisy observed signal. We can write:

$$x(t) = s(t) + n(t) \tag{1}$$

In the short-term Fourier domain we have:

$$X(m, f) = S(m, f) + N(m, f) \tag{2}$$

where m is the current frame and f is the frequency index.

The actual spectral weighting is now performed by multiplying the spectrum $X(m, f)$ with a real weighting function $G(m, f)$ >= 0. We call $G(m, f)$ a *weighting function* or *weighting rule*. The result $\hat{S}(m, f)$ is then,

$$\hat{S}(m, f) = G(m, f)X(m, f) \tag{3}$$

and the cleaned output signal $\hat{s}(t)$ of the system is obtained by transforming $\hat{S}(m, f)$ back into the time domain.

Because of the short-time stationary of speech, the processing has to be done on a frame-by-frame basis.

A basic system for this is shown in Fig. 1. In addition to the functions shown, other ones such as framing, windowing and overlap-and-add, are also necessary [8]. Because $G(m, f)$ is a real function, only the magnitude of $X(m, f)$ is changed. The phase is retained for the reconstruction.

Manuscript received October 19, 2005.

Farid Ykhlef and A. Guessoum are with the LATSI laboratory, Department of Electronic, Faculty of Engineer Sciences, University of Blida, Algeria
(e-mail: f_ykhlef@yahoo.fr).

D. Berkani is with the Signal and Communication Laboratory - (ENP), Algiers, Algeria.

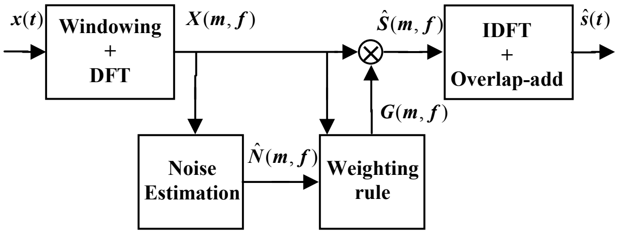


Fig. 1. The principle of spectral weighting.

The weighting function $G(m,f)$ is usually a function of the magnitude spectra $\left|\hat{S}(m,f)\right|$ and $\left|N(m,f)\right|$, or of the power spectral densities $\left|\hat{S}(m,f)\right|^2$ and $\left|N(m,f)\right|^2$. Thus, to calculate $G(m,f)$ some estimate of the noise which should be reduced is necessary. The spectrum of the noise during speech periods is not exactly known. The basic idea is to measure the noise spectrum only when there is no speech. However, it can be estimated, since the noise is assumed to be a short-time stationary process. The estimate of the noise is taken from the speech pauses which are identified using a voice activity detector (VAD).

A working VAD (*voice activity detection*) in hand, giving values of zero and one as indicators of the voice activity in each frame, enables us to update the estimate of the background noise spectrum during the frames that have zero VAD, using the formula:

$$\left|\hat{N}(m,f)\right|^2 = \lambda\left|\hat{N}(m-1,f)\right|^2 + (1-\lambda)\left|X(m,f)\right|^2 \quad (4)$$

where $\left|X(m,f)\right|^2$ is the spectrum of the noisy speech and λ is the forgetting factor.

As a result of these approximations (necessary to follow the time varying nature of the useful signal) undesirable distortions can occur, the most notable being known as "musical noise" in which statistical fluctuations in the frequency components of noise lead to random tonal artifacts in the processed signal. Various techniques have been applied to mask or eliminate these distortions [9] [10].

III. Spectral Weighting Rules

In this section we will briefly discuss two weighting rules for speech noise reduction. These two rules are necessary for our proposed hybrid system.

A. Spectral Subtraction

One of the first weighting rules proposed for speech noise reduction was the spectral subtraction [11]. One version of it is the *magnitude spectral subtraction.* Basically it means that an estimate $\left|\hat{N}(m,f)\right|$ of the noise magnitude spectrum is subtracted from the instantaneous input magnitude spectrum $X(m,f)$ such that:

$$\left|\hat{S}(m,f)\right| = \left|X(m,f)\right| - \left|\hat{N}(m,f)\right| \quad (5)$$

Written as a weighting rule we have:

$$G(m,f) = 1 - \frac{\left|\hat{N}(m,f)\right|}{\left|X(m,f)\right|} \quad (6)$$

To prevent $\left|\hat{S}(m,f)\right|$ from being negative, $\left|\hat{N}(m,f)\right|$ must not be greater than $X(m,f)$.

Although the noise level is reduced by the spectral subtraction, a serious disadvantage is that there will remain an unnatural sounding residual noise. It can be easily explained by the statistical nature of the noise [12]. For example, consider some frequency of the instantaneous spectrum which does not contain any speech, $X(m,f) = N(m,f)$. On the one hand, the effect of too small a noise estimate, $\left|\hat{N}(m,f)\right| < \left|N(m,f)\right|$, is a remaining excitation at this frequency.

On the other hand, if the noise is estimated to be higher than it actually is, the result will be zero due to the necessary bounding, $\hat{S}(m,f)$ = 0. The result is short sinusoids randomly distributed over time and frequency, which remain in the processed signal. Normally this kind of noise is called *musical noise.*

B. The Wiener Filter

The Wiener filter rule is derived from the optimal filter theory [1]. It is based on minimizing the mean squared error between the speech $S(m,f)$ and the estimate $\hat{S}(m,f)$:

$$E[(S(m,f) - \hat{S}(m,f))^2] \quad (7)$$

It is assumed here that the speech and the noise obey normal distribution and do not correlate; otherwise we are in a dead end.

So assume:

$$E[\left|X(m,f)\right|^2] = E[\left|N(m,f)\right|^2] + E[\left|S(m,f)\right|^2] \quad (8)$$

After some processing of the equations (by which the expected values also disappear) we end up with the filter:

$$G(m,f) = \frac{\left|S(m,f)\right|^2}{\left|N(m,f)\right|^2 + \left|S(m,f)\right|^2} \quad (9)$$

The result when using the Wiener rule above also suffers from musical noise. However, the Wiener rule can be implemented in other ways which help to reduce the amount of musical noise [9] [10].

IV. Combined Spectral Subtraction and Wiener Filter in Wavelet Domain

A novel noise reduction structure is proposed based upon a combination of cascaded spectral subtraction and Wiener filter methods in wavelet domain. The proposed hybrid system needs the use of the Discrete Wavelet Transform. For that, we will provide a brief introduction of wavelet transforms.

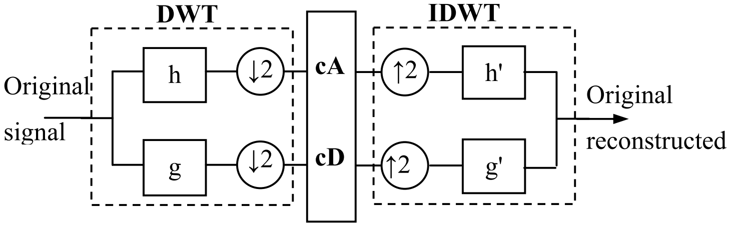


Fig. 2. Decomposition and reconstitution Algorithm
h = low-pass decomposition filter; g = high-pass decomposition filter; ↓2 = down-sampling operation.
h’ = low pass reconstruction filter; g’ = high-pass reconstruction filter; ↑2 = up-sampling operation

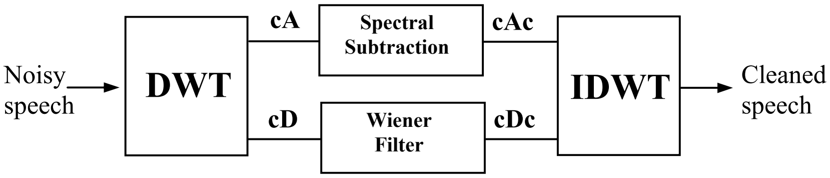


Fig. 3. Hybrid system.

A. Discrete Wavelet Transform

Wavelet transform is a new and promising set of tools and techniques for speech processing. Wavelets have generated a tremendous interest in both theoretical and applied areas, especially over the past few years. There exists an extensive literature addressing the wavelet transform.

The discrete wavelet transform DWT can be simply thought of in terms of filter banks. A filter bank is defined as a set of filters which are applied to a signal together with changes in sampling rates. The simplest case is the two-channel filter bank which consists of a low-pass and a high-pass filter, represented by the coefficients **h** and **g** respectively. An efficient way to implement this scheme using filters was developed in 1988 by Mallat [13].

The low-pass coefficients **cA** can be thought of as representing a coarser approximation of the data and are known as the approximation coefficients. Correspondingly, the high-pass coefficients **cD** represent more detailed information in the data and are known as the detail coefficients.

The other half of the story is how those components can be assembled back into the original signal with no loss of information. This process is called *reconstruction*, or *synthesis*. The mathematical manipulation that effects synthesis is called the *inverse discrete wavelet transform* (IDWT). Where wavelet analysis involves filtering and downsampling, the wavelet reconstruction process consists of upsampling and filtering. The algorithm of wavelet signal decomposition and reconstruction is illustrated in Fig. 2. Moreover, we must add that there are different types of wavelets such as: Haar, Daubechies, Coiflets, Symlet, Biorthogonal and etc. In our case, we chose the Daubechies wavelet.

B. Hybrid System

The proposed speech enhancement scheme is illustrated by Fig. 3. The idea consists to enhanced the approximation coefficients (**cA**) and the detail coefficients (**cD**) resulting from DWT transformation of the noisy speech signal by Spectral subtraction and Wiener filter respectively.

The cleaned approximation and detail coefficients (**cAc**,**cDc**) are transformed in time domain using IDWT transformation.

V. Experimental Results

Experiments are carried out for the three different methods discussed in Section IV. We will use for these experiments frames of 25 ms with an overlap between tow successive frames of 40 %. To demonstrate the usefulness of the proposed scheme in the context of noise reduction application, we will compare the Spectral Subtraction and Wiener Filter with the proposed scheme.

For the hybrid system, we will choose Daubechies wavelet of an order equal to 7.

The test sentence was originally recorded under controlled conditions at a sampling frequency of 16 kHz, using 16-bit. The noise was taken from the interior of an automobile in rainy conditions. Signal to Noise Ratio measures are used for this experiment. It is defined as:

$$\boldsymbol{SNR}_{out} = 10\log_{10}\left(\frac{\sum_{t}\left|s(t)\right|^{2}}{\sum_{t}\left|s(t)-\hat{s}(t)\right|^{2}}\right) \quad (\boldsymbol{dB}) \qquad (10)$$

Table I shows the results obtained using different input Signal to Noise Ratio $\boldsymbol{SNR}_{input}$ (vehicle interior noise).

We can notice that the hybrid system exhibits better results than those obtained with the Spectral Subtraction and Wiener Filter in terms of $\boldsymbol{SNR}_{out}$ and also of noise reduction.

TABLE I

$\boldsymbol{SNR_{out}}$ (dB)

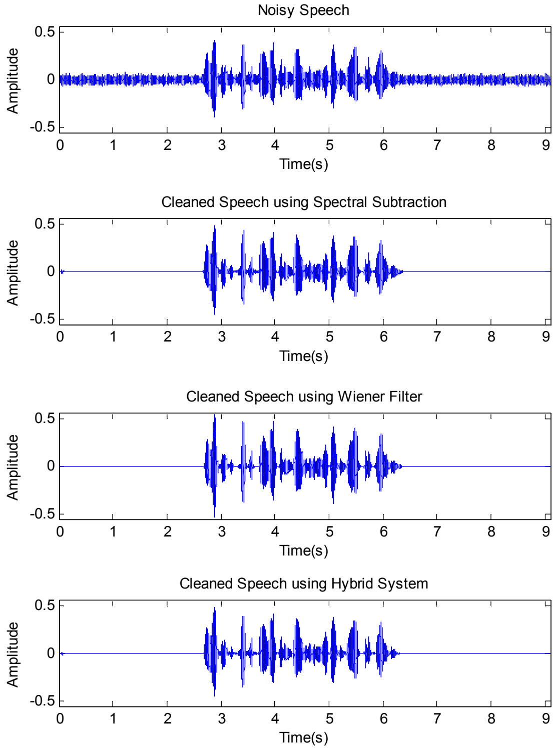
$\boldsymbol{SNR_{input}}$ (dB)	Spectral Subtraction	Wiener Filter	Hybrid System
10	10.54	9.00	10.62
5	10.68	8.95	10.77
0	10.77	8.82	10.88
-5	10.50	8.42	10.65
-10	9.01	7.17	9.34

For a second experiment, the noisy signal is recorded in real conditions at 16 kHz sampling rate in a car environment. The Fig. 4 and Fig. 5 show the time evolutions and the spectrograms of a recorded noisy speech signal with the cleaned speech using the three different methods.

VI. Conclusion

In this paper a new scheme based upon a combination of cascaded spectral subtraction and Wiener filter methods in wavelet domain was proposed for noise reduction fields.

A comparative study between the different methods was carried out to evaluate the performances of the proposed system. The experimental results have shown that our proposed hybrid system is capable of reducing noise and is an adequate procedure to improving the quality of the speech enhancement application.

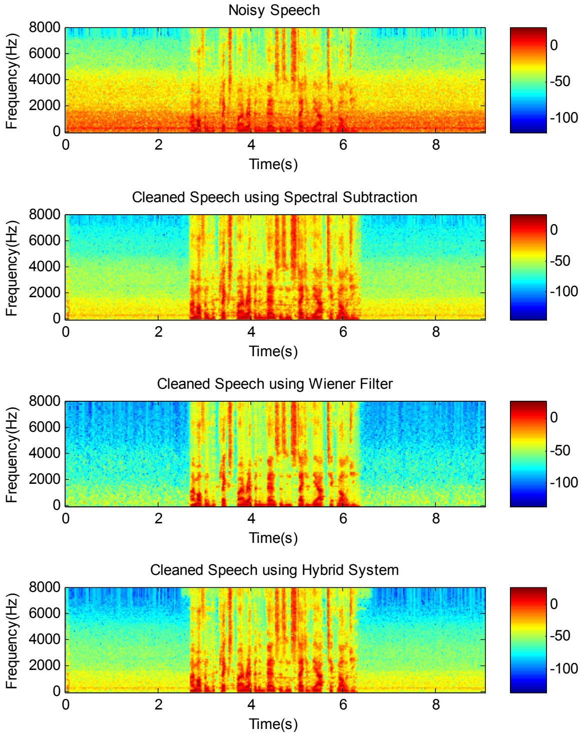

Noisy Speech

Cleaned Speech using Spectral Subtraction

Cleaned Speech using Wiener Filter

Cleaned Speech using Hybrid System

Fig. 4. Temporal representations.


Noisy Speech

Cleaned Speech using Spectral Subtraction

Cleaned Speech using Wiener Filter

Cleaned Speech using Hybrid System

Fig. 5. Spectrograms.

References

[1] S. V. Vaseghi, Advanced Signal Processing and Digital Noise Reduction. Wiley Teubner, 1996.

[2] S. J. Godsill and P. J. W. Rayner, Digital Audio Restoration. Springer Verlag, 1998.

[3] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," *IEEE Trans. on ASSP*, 1984, pp. 11091121.

[4] Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error log-spectral amplitude estimator," *IEEE Trans. on ASSP*, 1985, pp. 443-445.

[5] Y. Ephraim, "A signal subspace approach for speech enhancement," *IEEE Trans. on speech and audio processing*, 1995, pp. 251-266.

[6] N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system," *IEEE Trans. on speech and audio processing,* 1999, pp.126-137.

[7] D. E. Tsoukalas, J. N. Mourjopoulos, and G. Kokkinakis, "Speech enhancement based on audible noise suppression," *IEEE Trans. on speech and audio processing*, 1997, pp. 497-514.

[8] R. E. Crochiere, "A weighted overlap-add method of short-time Fourier analysis/synthesis," *IEEE Trans. Acoustics, Speech, Signal Processing,* Vol. 28, February 1980.

[9] O. Cappé, "Elimination of the musical noise phenomenon with the Ephraim and Malah noise suppressor," I*EEE Trans. on speech and audio processing*, 1994, pp. 345-349.

[10] A. A. Azirani, R. L. B. Jeannes and G. Faucon, "Speech Enhancement Using a Wiener Filtering Under Signal Prescence Uncertainty," in *Proceedings European Signal Processing Conference*, Trieste, Italy, September 1996.

[11] S.F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 27, April 1979, pp. 113-120.

[12] P. Vary, "Noise suppression by spectral magnitude estimation—mechanism and theoretical limits," *EURASIP Signal Processing,* Vol. 8, 1985, pp. 387-400.

[13] S. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Pattern Anal. and Machine Intell.*, vol. 11, N. 7, 1989, pp. 674-693.