Improvement in Listener Comfort Through Noise Shaping Using a Modified Wiener Filter Approach

Vasudev Kandade Rajan\textsuperscript{1}, Christin Baasch\textsuperscript{1}, Mohamed Krini\textsuperscript{2}, Gerhard Schmidt\textsuperscript{1}

\textsuperscript{1}Digital Signal Processing and System Theory, Kiel University, Germany
Email: \{vakr, chrb, gus\}@tf.uni-kiel.de
\textsuperscript{2}paragon AG, Delbrueck, Germany
Email: \{Mohamed.Krini\}@paragon.ag

Abstract

The Wiener filter approach to noise suppression is well known both theoretically and practically. This paper explains the idea of extending this filtering approach to shape the residual noise of the output signal. Motivated by different kinds of stationary and instationary noise sources that appear in an automobile, along with the necessity to address the problem of increasing the listener comfort, an algorithm to achieve a time-frequency varying attenuation in the lower frequencies has been developed. The basic theory and the conditions to achieve this are presented. Details of the developed compressor algorithm as well as the modifications necessary to apply it to the classical Wiener filter are shown. The improvement that the modified filter brings to the noise input signal are significant and analyses of the results clearly show a reduction of the “ruggedness” of the output signal.

1 Introduction

Background noise suppression for acoustic signals recorded via microphones is a widely studied, and a practical problem. In applications such as hands-free systems, mobile phones, and a wide variety of devices, noise suppression plays a vital role in improving the overall quality of the system. When these systems are placed in environments like a car cabin the performance has to be high from the point of view of both safety and end-user expectation. Noise suppression algorithms basically consist of two steps. First is to estimate the background noise with any noise estimation algorithm. Second is to attenuate the estimated noise with an adaptive filter such as the Wiener filter. Depending on the computed filter coefficients the attenuation ranges from 0 dB to a minimum value. For automobile environments the minimum is fixed at a certain dB value and is called spectral floor. During noise only situations the spectral floor determines the shape of the noise-suppressed signal. A usual practice is to have a fixed floor which retains the acoustic information of the background noise. Although this is desirable, there are many situations where a more modified noise sounds more comfortable. Some examples are road bumps noise, highly instationary noise especially in the lower frequencies, etc. By carefully controlling the frequency bins of the noise suppression filter a desired noise shape can be achieved. In particular the noise suppressed output can be shaped by modifying the spectral floor. Puder has shown such a method in \cite{2} where a time-frequency varying spectral floor is based on a curve characteristic between a smoothed Wiener filter characteristic and pre-determined attenuation. The attenuation is applied in a time-varying manner based on the average noise-suppression-filter coefficients. A higher value means a reduced attenuation and vice versa. This method eliminates random openings of the filter due to non-speech components and reduces the effect of over-estimation on speech components.

In this paper a method based on time-frequency variation of the spectral floor, not based on the average filter coefficients, but with a goal to shape the noise suppressed signal to a desired shape. The rest of the paper is organized as follows. Sec.\textsuperscript{2} formally introduces the subband domain Wiener filter approach for noise suppression and the conditions under which noise shaping can be achieved. Sec.\textsuperscript{3} briefly introduces different noise sources in an automobile environment. Sec.\textsuperscript{4} presents an example of such an approach applied to a highly disturbed input signal by controlling the shape in the lower frequencies. Sec.\textsuperscript{5} presents the results of the method shown in Sec.\textsuperscript{4}. Finally a conclusion is presented in Sec.\textsuperscript{6}.

2 Revisiting the Wiener Filter

For the following consider a discrete-time noisy input signal \(y(n)\) modeled as \(y(n) = s(n) + b(n)\) where \(s(n)\) is the clean speech signal and \(b(n)\) is the background noise. The short-time Fourier transform (STFT) is used to obtain the spectrum of the input signal \(Y(\mu,k)\) where \(\mu\) is the sub-band index and \(k\) is the frame index. Through any noise estimator a magnitude estimate of the background noise \(\hat{B}(\mu,k)\) is obtained. The clean speech estimate \(\hat{S}(\mu,k)\) is computed by multiplying the input spectrum \(Y(\mu,k)\) with a noise suppression filter \(H_{\text{classic}}(\mu,k)\):

\[
\hat{S}(\mu,k) = H_{\text{classic}}(\mu,k) Y(\mu,k).
\]

The coefficients of the filter \(H_{\text{classic}}(\mu,k)\) in Eq. (1) are computed using

\[
H_{\text{classic}}(\mu,k) = 1 - \frac{\hat{B}^2(\mu,k)}{|Y(\mu,k)|^2}.
\]

This limits the maximum attenuation of the filter to \(H_{\text{floor}}\) leading to

\[
H_{\text{mode}}(\mu,k) = \max \left\{ 1 - \frac{\hat{B}^2(\mu,k)}{|Y(\mu,k)|^2} \cdot H_{\text{floor}} \right\}.
\]

The value of the maximum attenuation is typically set between -6 and -20 dB. This means the attenuation curve will be limited to 0 dB in parts of the spectrum which contain high signal-to-noise ratio (SNR) and the noise floor of the output spectrum will be limited to \(H_{\text{floor}}\). One way to avoid “musical noise” or opening of the filter in non-speech frequency bins is to multiply the estimate of the noise with a factor, which forces the spectral floor to come into play.
This leads to the second modification in the computation of the filter coefficients
\[
H_{\text{mod2}}(\mu, k) = \max \left\{ 1 - \frac{\Lambda_{\text{o-est}} \hat{B}^2(\mu, k)}{|Y(\mu, k)|^2}, H_{\text{floor}} \right\},
\] (4)
where \(\Lambda_{\text{o-est}}\) is called the overestimation factor.

2.1 Shaping the noise spectrum
When the overestimation is chosen to be high enough, the output of the noise suppression during noise-only situations is an attenuated version of the input spectrum where the attenuation is determined by the parameter \(H_{\text{floor}}\). Under these conditions the spectral floor can be exploited to shape the noise to a desired shape. The word “shape” is used here to imply that the magnitude spectrum of the resulting residual noise is modified. An example scenario for applying such noise shaping is in automobiles. Here noise caused due to a road bump or an uneven road surface is difficult to suppress given its high power and spontaneous occurrence. Even if detected perfectly the suppression is limited by the spectral floor which still makes it audible. To make it inaudible the bump noise must be suppressed with maximum attenuation values much smaller than typical floor values. The solution for suppressing this perfectly detected bump noise is to set those bins in the spectral floor, where the noise is detected, to a very low value. This means that \(H_{\text{floor}}\) is now a function of time and frequency leading to \(H_{\text{floor}}(\mu, k)\). Given this, the conditions under which noise shaping can be performed are the following:

- relatively good noise estimation \(\hat{B}(\mu, k)\),
- high overestimation \(\Lambda_{\text{o-est}}\), depending on the noise estimator, and
- time-frequency varying spectral floor \(H_{\text{floor}}(\mu, k)\).

From the above discussion Eq. (5) is arrived at
\[
H_{\text{mod1}}(\mu, k) = \max \left\{ 1 - \frac{\Lambda_{\text{o-est}} \hat{B}^2(\mu, k)}{|Y(\mu, k)|^2}, H_{\text{floor}}(\mu, k) \right\}. \tag{5}
\]

Effectively the spectral floor itself can be treated like a filter, which modifies the noise. The spectral floor can be modified in various ways depending on the application. It could also be coupled with an adaptive filter algorithm which computes the filter coefficients given a desired noise shape or even be used as a noise whitener by acting like an equalizer.

2.2 Desired noise shape
To present the solution of noise shaping a more complete definition of a desired noise \(B_{\text{desired}}(k, \mu)\) shape is necessary. During noise only segments of the input spectrum \(Y(\mu, k)\) the goal is to obtain an output spectrum such that the magnitude is close to the magnitude of the desired noise
\[
|\hat{S}(\mu, k)| \approx B_{\text{desired}}(k, \mu) \Big|_{S(\mu, k)=0} \tag{6}
\]
and the phase is taken from the noisy input signal. From the above described conditions for noise shaping \(H_{\text{mod1}}(\mu, k)\) is equal to the modified spectral floor \(H_{\text{floor}}(\mu, k)\), which results in
\[
\hat{S}(\mu, k) = Y(\mu, k) \cdot H_{\text{floor}}(\mu, k) \Big|_{S(\mu, k)=0}. \tag{7}
\]
noise suppression filter must track this rising and falling amplitude. The suppression of the amplitude should be enough to handle this kind of noise.

4 Adjusting the Spectral Floor

Given the different kind of noise conditions the objective now is to explore different desired noise shapes. From a clearly defined noise shape $B_{\text{desired}}$, the spectral floor $H_{\text{floor}}$ can be computed by

$$H_{\text{floor}}(\mu, k) = \frac{B_{\text{desired}}(\mu, k)}{|Y(\mu, k)|}. \quad (8)$$

The desired noise shape can be chosen in many ways. It can be white noise, colored noise, noise independent of input spectrum, noise dependent on the input spectrum, etc. Broadly it can be classified into input-independent shapes (e.g. fixed magnitude spectral densities) and input-dependent (leading e.g. to the conventional approach or to a new one – both will be described below).

4.1 Fixed noise shape

The desired noise can have a fixed frequency-dependent shape. For example it can be set to white noise, which has equal power in all frequencies, or gray noise which has equal loudness in all frequencies as perceived by a human ear. For such a shape the desired noise should be set to

$$B_{\text{desired}}(\mu, k) = B_{\text{des-fixed}}(\mu). \quad (9)$$

4.2 Input independent attenuation

Input independent attenuation means that the final shape of the desired noise is an attenuated version of the input spectrum:

$$B_{\text{desired}}(\mu, k) = H_{\text{floor}} \cdot |Y(\mu, k)|. \quad (10)$$

where $H_{\text{floor}}$ determines the amount of attenuation applied. Note, if we combine Eqs. (10), (8), and (5) we end up with the conventional scheme described by Eq. (4).

4.3 Input dependent attenuation

A more robust and flexible choice would be to compute the desired noise based on the input spectrum as shown in Eq. (11)

$$B_{\text{desired}}(\mu, k) = C(|Y(\mu, k)|) \cdot |Y(\mu, k)|. \quad (11)$$

The aim is to design a function $C(...) \gamma$ which computes the attenuation to be applied depending on the magnitude of the input spectrum. A first attempt to design such a function has been made here. From the analysis of the various sources of noise in an automobile, it has been observed that there are high fluctuations in the lower frequencies of the noisy input spectrum. The function $C(...) \gamma$ must be designed to reduce these fluctuations and attenuate the input spectrum. An exact value of the compression to be applied can be computed by the ratio of a desired magnitude level $Y_{\text{des-mag}}(\mu, k)$ and the instantaneous magnitude $|Y(\mu, k)|$. By applying additional compression control parameters, the function $C(...) \gamma$ can be defined as

$$C(...) = \min \left\{ 1, \left( \frac{\Delta Y_{\text{est}} Y_{\text{des-mag}}(\mu, k)}{|Y(\mu, k)|} \right)^{C_{\text{control}}}, H_{\text{floor}} \right\}. \quad (12)$$

$\Delta Y_{\text{est}}$ controls the overestimation applied to the desired magnitude level. $C_{\text{control}}$ is a compression control parameter through which the final attenuation to be applied can be controlled. Fig 2 shows the output of the function $C(...) \gamma$ plotted against the input magnitude to desired level ratio.

4.4 Estimating the desired magnitude

The desired magnitude level can be estimated in two different ways:

1. the input spectrum floor as estimated from a noise estimator can be used or
2. a desired level can directly be computed by slowly tracking the input spectrum magnitude.

The disadvantage with using the noise estimate is that in many cases the noise estimator is parametrized to be very close to the "true" background noise which would then make the desired level sensitive to immediate changes in the background noise. Whereas, slow tracking in the long-term converges to an average magnitude level of the input spectrum. The desired magnitude has to counter the two specific noise properties: fluctuation (along the frequency axis) and modulation in the lower frequencies. From the instantaneous spectrum of the current frame, the fluctuations can be removed by applying a zero phase forward-backward smoothing. For a frame $k$ the smoothing is performed using a first order IIR filter as shown in Eq. (13) along the frequency axis

$$\gamma_{\text{inst}}(\mu, k) = \gamma_{\text{inst-fq}}[|Y(\mu, k)|] + \left(1 - \gamma_{\text{inst-fq}}\right) \gamma_{\text{inst}}(\mu - 1, k), \quad (13)$$

for $\mu = 1, \ldots, N_{\text{bin}}$. The smoothing is performed in the backward direction as well. $\gamma_{\text{inst-fq}}$ can also be chosen in a frequency dependent manner such that high smoothing is applied in the lower frequencies and low smoothing is applied in the higher frequencies. $\gamma_{\text{inst-fq}}$ can be set to 30 dB per Hz in the lower frequencies. For controlling the modulation a multiplicative constant based tracker is employed. The tracker follows the smoothed input $\gamma_{\text{inst}}(\mu, k)$ based on the previous estimate $Y_{\text{des-mag}}(\mu, k - 1)$. A rising constant $\Delta Y_{\text{mag-inc}}$ or a falling constant $\Delta Y_{\text{mag-dec}}$ is multiplied to the previous estimate based on a comparison with...
In this paper the existing Wiener filter modifications have been extended to be applied in a different scenario. Although the idea to modify the spectral floor has been presented before, it was used on a predefined mapping characteristic. This paper took that idea further to compute a curve dependent on the input spectrum. The paper also presented the conditions under which the spectral floor can be used to shape the residual noise of the output. An application example in the form of a compressor was presented. The basic idea of a compressor is to estimate the input spectrum and compute the distance between the instantaneous spectrum level and the desired spectrum level. The method was implemented in a real-time system and a subjective evaluation of the results was presented.

5 Results and Evaluation

The noise suppression technique presented above was implemented on a real-time system. The algorithm was tested on a variety of noise conditions which covered the three different noise cases:

- the impact of compressor on non-fluctuating noise,
- on narrow street noise case, and
- a passing car noise case.

The frame of reference for these tests was the improvement in listener comfort level. Listening tests were performed to evaluate the noise comfort level aspect. Nine listeners participated in the listening tests who rated the level of comfort for all permutations involving the original signal, conventionally suppressed signal, and proposed noise shaping method. A modified method of [3] was used for noise estimation. A Comparison Mean Opinion Score (CMOS) was computed based on the evaluation, in which +3 indicates ‘much better’ and -3 indicates ‘much worse’. Important cases are the scores where the proposed method is compared to the conventional method which have received a score of +2.5 and +1.3 for the narrow street and passing car noise scenarios. The averages and the standard deviations (std.) of these scores are presented in Tab. 1.

The spectrogram of a part of the original input signal, output without the compressor algorithm and output with the

![Spectrogram of input and outputs.](image)

The increment and the decrement constants can be set to about 1 dB per second which would result in a duration of 10 seconds for a change in the level by 10 dB. The tracked magnitude has the information about the average power level of the input spectrum. The attenuation computed in Eq. (11) is applied up to the frequency bin \( \mu_{\text{comp-max}} \). This maximum frequency can be parametrized. The overestimation factor has to be set to a high value up to \( \mu_{\text{comp-max}} \). This ensures that the modified floor can influence the noise in those bins.

\[
Y_{\text{des-mag}}(\mu, k) = \begin{cases} 
Y_{\text{des-mag}}(\mu, k - 1) \Delta \rho_{\text{mag-inc}}, & \text{if } Y_{\text{des-mag}}(\mu, k - 1) < Y_{\text{inst}}(\mu, k), \\
Y_{\text{des-mag}}(\mu, k - 1) \Delta \rho_{\text{mag-dec}}, & \text{else}.
\end{cases}
\] (14)

The smoothed spectrum as in Eq. (14)

\[
Y_{\text{des-mag}}(\mu, k) = \begin{cases} 
Y_{\text{des-mag}}(\mu, k - 1) \Delta \rho_{\text{mag-inc}}, & \text{if } Y_{\text{des-mag}}(\mu, k - 1) < Y_{\text{inst}}(\mu, k), \\
Y_{\text{des-mag}}(\mu, k - 1) \Delta \rho_{\text{mag-dec}}, & \text{else}.
\end{cases}
\] (14)

\[
Y_{\text{des-mag}}(\mu, k) = \begin{cases} 
Y_{\text{des-mag}}(\mu, k - 1) \Delta \rho_{\text{mag-inc}}, & \text{if } Y_{\text{des-mag}}(\mu, k - 1) < Y_{\text{inst}}(\mu, k), \\
Y_{\text{des-mag}}(\mu, k - 1) \Delta \rho_{\text{mag-dec}}, & \text{else}.
\end{cases}
\] (14)

The spectrogram of a part of the original input signal, output without the compressor algorithm and output with the

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Convent. vs.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unprocessed</td>
<td>1.4 / 1.6</td>
<td>2.5 / 0.4</td>
<td>-2.1 / 1.2</td>
</tr>
<tr>
<td>Proposed vs.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unprocessed</td>
<td>1.8 / 1.5</td>
<td>2.1 / 0.7</td>
<td>2.5 / 0.4</td>
</tr>
<tr>
<td>Proposed vs.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conventional</td>
<td>0 / 1</td>
<td>1.3 / 1.4</td>
<td>2.5 / 0.6</td>
</tr>
</tbody>
</table>

Table 1: Result of a CMOS test indicating the listener comfort level.

The spectrogram of a part of the original input signal, output without the compressor algorithm and output with the

This smoothing effect due to the compressor can be clearly seen in the spectrogram plots.

6 Conclusion

In this paper the existing Wiener filter modifications have been extended to be applied in a different scenario. Although the idea to modify the spectral floor has been presented before, it was used on a predefined mapping characteristic. This paper took that idea further to compute a curve dependent on the input spectrum. The paper also presented the conditions under which the spectral floor can be used to shape the residual noise of the output. An application example in the form of a compressor was presented. The basic idea of a compressor is to estimate the input spectrum and compute the distance between the instantaneous spectrum level and the desired spectrum level. The method was implemented in a real-time system and a subjective evaluation of the results was presented.

References