Dear Author,

Please, note that changes made to the HTML content will be added to the article before publication, but are not reflected in this PDF.

Note also that this file should not be used for submitting corrections.
Noise reduction of speech signals using time-varying and multi-band adaptive gain control for smart digital hearing protectors

Narimene Lezzoum, Ghyslain Gagnon, Jérémie Voix *

École de technologie supérieure, 1100 Notre Dame West, Montréal, Qc H3C 1K3, Canada

ABSTRACT

In this paper, a single-channel speech enhancement algorithm based on non-linear and multi-band Adaptive Gain Control (AGC) is proposed. The algorithm requires neither signal-to-noise ratio (SNR) nor noise parameters estimation. It reduces the background noise in the temporal domain rather than the spectral domain using a non-linear and automatically adjustable gain function for multi-band AGC. The gain function varies in time and is deduced from the temporal envelope of each frequency band to highly compress the frequency regions where noise is present and lightly compress the frequency regions where speech is present. Objective evaluation using the PESQ (perceptual evaluation of speech quality) metric shows that the proposed algorithm performs better than three benchmarks, namely: the spectral subtraction, the Wiener filter based on a priori SNR estimation and a band-pass modulation filtering algorithm. In addition, blind subjective tests show that the proposed algorithm introduces less musical noise compared to the benchmark algorithms and was preferred 78.8% of the time in terms of signal quality. The proposed algorithm is implemented in a miniature low power digital signal processor to validate its feasibility and complexity for smart hearing protection in noisy environments.

* Corresponding author.
E-mail address: jeremie.voix@etsmtl.ca (J. Voix).

http://dx.doi.org/10.1016/j.apacoust.2016.03.001
0003-682X/© 2016 Published by Elsevier Ltd.

Please cite this article in press as: Lezzoum N et al. Noise reduction of speech signals using time-varying and multi-band adaptive gain control for smart digital hearing protectors. Appl Acoust (2016), http://dx.doi.org/10.1016/j.apacoust.2016.03.001
quency band is high. While this algorithm proved to be effective at enhancing speech, it does not significantly reduce the background noise when speech is absent. In the targeted application, where the user wears hearing protection in noisy environments, continuous reduction of background noise is an important feature. It would also be desirable in other applications such as noise-canceling ear-buds. In [15], an Adaptive Gain Control (AGC) based on an SNR estimator was proposed. Unfortunately, it was also shown in [15] that the proposed SNR estimation method is not accurate in low SNR environments (0 dB) and adds artefacts to the enhanced signal. In [16], an AGE applied to the multi-band temporal envelopes was proposed to boost the signal when speech is present. In this method, the gain function is applied to the temporal envelope which is afterward multiplied by the carrier of the signal.

The authors introduced in [17] a single-channel speech enhancement algorithm with a live demonstration using recordings. This paper extends this work with objective and subjective evaluations of this algorithm, its comparison with three other state of the art methods, in addition to its implementation in a miniature low-power digital signal processor (DSP) for smart hearing protection applications.

The proposed algorithm calculates a time-varying and frequency-band dependent gain function from the temporal envelope of each frequency-band. This function enables high compression of frequency bands containing noise and light compression of frequency-bands containing speech. The proposed algorithm operates without any knowledge or estimation of the noise parameters, or assuming that the background noise is additive. It will be shown that this gain function reduces the background noise and improves the quality of the speech signal.

The paper is organized as follows. Section 2 details the proposed noise reduction algorithm. In Section 3, the experimental methodology is presented. Section 4 discusses the objective and subjective results. Section 5 presents the hardware implementation of the method, and Section 6 concludes the paper.

2. Proposed algorithm

Fig. 1 illustrates the architecture of the proposed algorithm. The incoming noisy speech signal \( y(n) \) is composed of clean speech \( x(n) \) and additive noise \( w(n) \):

\[
y(n) = x(n) + w(n)
\]

The incoming signal is divided into \( M = 16 \) frequency bands using fourth order band-pass Butterworth filters. Filter bandwidths are characterized by the equivalent rectangular bandwidth (ERB) [18]. The center frequency of the first and last frequency bands are 125 and 3700 Hz respectively.

The output of each filter is given by:

\[
y_m(n) = y(n) \ast h_m(n)
\]

with \( h_m(n) \) the impulse response of the \( m \)th band-pass filter, and the symbol \( \ast \) denoting convolution.

The Hilbert envelope of the signal \( y_m(n) \) is extracted as per the following equation:

\[
E_m(n) = \sqrt{y_m(n)^2 + \tilde{y}_m(n)^2}
\]

with \( \tilde{y}_m(n) \) the Hilbert transform of \( y_m(n) \), defined as [19]:

\[
\tilde{y}_m(n) = y_m(n) + \frac{1}{2\pi n}
\]

The proposed technique achieves noise reduction using multi-band time-varying gain functions. Our investigation shows that these gain functions must meet three criteria:

- The gain function of each frequency band must be smooth and continuous to avoid abrupt changes in the enhanced signal.
- The gain function must be chosen as a function of the temporal envelope \( E_m(n) \) in order to preserve the quality of speech without adding artefacts.
- The gain function should be near 1 in the frequency bands containing speech and near 0 in the frequency bands containing noise, in order to preserve speech components and attenuate noise components.

A time-varying gain function that fulfills these criteria is the low-pass filtered temporal envelope \( E_m(n) \) of the signal. The gain function is thus:

\[
G_m(n) = (E_m(n) + L(n))
\]

with \( L(n) \) the impulse response of a fourth order low-pass filter. The optimal cut-off frequency \( f_c \) of the low-pass filter is later determined in Section 3.1.

Enhancing the signal in each frequency band consists of multiplying the signal by its smoothed envelope:

\[
x_m(n) = G_m(n) \cdot y_m(n)
\]

This can be seen as non-linear compression: frequency bands with high energy will barely be compressed and frequency bands with low energy will be highly compressed.

The enhanced signal \( x(n) \) of each frame is reconstructed by summing the \( M \) frequency bands, and rescaled using a gain constant "a". In this paper, "a" is the ratio between the RMS (Root Mean Square) values of the noisy and enhanced signals. This gain constant could also be set by the user to adjust the desired listening level.

As an illustrative purpose, Fig. 2 shows the noise reduction effect of the gain function on a 250 ms speech signal corrupted...
by car noise. The clean speech signal is considered as the reference to see how the gain function reduces the background noise continuously in the temporal domain and in the different frequency bands. When speech is absent (from 0 to 0.1 s in the four frequency bands), the background noise is highly compressed. Fig. 2 also shows that speech signal is amplified in the frequency band centered at 698 Hz due to the presence of the first formant.

Fig. 3 illustrates three waveform and spectrogram charts: clean speech produced by a male speaker, speech corrupted by car noise in 0 dB SNR, and the enhanced speech signal. This figure shows, in the temporal and spectral domains, the background noise-reducing effect of the proposed algorithm.

3. Experimental methodology

Although the proposed algorithm can perform in a sample-based approach, it was implemented in Matlab using 250 ms frames with 80% overlap for ease of simulations. Objective and subjective quality tests were conducted to evaluate the performance of the proposed noise reduction algorithm. The results are shown in Section 4, and some audio samples are available online.

3.1. Optimal cut-off frequency of the gain function

The fluctuation rate of the temporal envelope is called the modulation frequency and represents one of the characteristics of speech signal. In [21], a study on the impact of the modulation frequency on speech intelligibility was performed: the speech signal was divided into different frequency bands, and the temporal envelopes and fine structures of each frequency band were extracted. The temporal envelopes have been low-pass filtered with different cut-off frequencies (0, 0.5, 1, 2, 4, 8, 16, 32 and 64 Hz) to determine the most important modulation frequency range for speech intelligibility, knowing that the cut-off frequency are frequency-band independent. In these studies, it has been found that with a modulation frequency of 16 Hz, the speech intelligibility remains the same, and when reducing it, the speech intelligibility starts decreasing.

In this work, the same evaluation has been performed to find the optimal fluctuation rate of the gain function that represents the cut-off frequency $f_c$ of the low pass filtered envelope. This was achieved by using the perceptual evaluation of speech quality (PESQ) [22] as the objective function to maximize.
Signals from the Noizeus corpus [23] were used: 30 speech utterances corrupted by two noise environments, babble and car, in three SNRs (5, 0, and -5 dB). All signals were sampled at 8 kHz. Fig. 4 illustrates the PESQ metric obtained using 6 low-pass filters with cut-off frequencies at 4, 8, 16, 32, 64, and 128 Hz. This figure shows that higher PESQ scores are obtained at 16 Hz in both noise environments and SNRs. In 4 and 8 Hz the temporal envelope is almost at a constant value. Thus, the gain functions calculated using these low cut-off frequencies are equivalent to a constant gain in each frequency band. Low-pass cut-off frequencies of 32, 64, and 128 Hz also gave a PESQ score lower than the 16 Hz, due to their high fluctuation rate, which brings up artefacts in the enhanced signal. From Fig. 4, we conclude that 16 Hz is the optimal cut-off frequency for the gain function low-pass filter.

### 3.2. Objective evaluation

Although the ITU-T P862 standard [22] mentions that the developed PESQ metric has not been validated for noise reduction algorithms, it was shown in [23] that among seven objective speech quality metrics for the evaluation of speech enhancement algorithms, the PESQ metric is the most correlated with the subjective overall quality and signal distortion. Thus, in this paper, the PESQ is used as the objective metric, using the same signals from the NOIZEUS database used in Section 3.1. The performance of the proposed algorithm was compared to three noise reduction algorithms implemented in Matlab, namely: the Wiener filter based on an *a priori* SNR estimation [6], the spectral subtraction [5], and band-pass modulation filtering [12]. The Wiener filter and spectral subtraction codes were taken from [3] (wiener-as and SpecSub), while the code of the modulation filtering was obtained directly from the authors of [12]. The Wiener filter and spectral subtraction codes were taken from [3] (wiener-as and SpecSub), while the code of the modulation filtering was obtained directly from the authors of [12]. In the spectral subtraction, the noise spectrum was estimated and updated from non-speech frames detected using a simple VAD based on segmental SNR, while in the Wiener filter, non-speech frames were detected using *a priori* SNR estimation. The Wiener filter and spectral subtraction were chosen because of their wide use as benchmark algorithms (e.g. [24,25,13]).

Other recent algorithms (e.g. [14–16]) were not selected as benchmarks because, unlike the selected benchmarks, the code was not available to the authors. Implementation intricacies such as non-optimal parameter settings could have potentially led to biased comparisons.

![Fig. 3.](image-url) On the left are the spectrograms and on the right their corresponding waveforms: top panel, the clean speech signal (a male speaking: “the birch canoe slid on the smooth planks”), middle panel, the same speech signal corrupted by car noise in 0 dB SNR, bottom panel, the enhanced signal with the proposed method.

![Fig. 4.](image-url) The PESQ metric calculated with different cut-off frequencies for speech signal corrupted by car and babble noise in 5 and 0 dB SNR.
3.3. Subjective evaluation

In addition to the objective evaluation procedure described above, two series of tests were performed to subjectively compare the proposed algorithm to benchmarks, in terms of level of musical noise and overall quality. To conduct these tests, 20 participants were recruited.

3.3.1. Musical noise assessment

In this test, the proposed algorithm and the three aforementioned benchmark algorithms are evaluated in terms of musical noise generation. Similarly, 30 speech signals corrupted by “car” and “babble” noise with 5, 0 and −5 dB SNR were used. The signals were identified by numbers and presented in a random order to 10 trained participants. These participants were asked to choose from the 4 signals processed with the four algorithms, the signal in which musical noise was the least perceptible. Before this test, participants were trained by being exposed to several speech signals heavily corrupted by musical noise.

3.3.2. Overall quality evaluation

The overall quality was subjectively evaluated to determine user preference among the proposed algorithm, three benchmarks and unprocessed signals. This evaluation was performed using 30 speech signals corrupted by “car” and “babble” noise with 5, 0 and −5 dB SNR, with 10 other participants. These participants were asked to pick the signal that has the best overall quality. During the tests, participants were allowed to repeat the same signal as often as needed.

4. Results and discussion

4.1. Objective test results

Fig. 5 presents a comparison of PESQ results obtained in the two noise environments and three SNRs using the noisy signals, the Wiener filter, spectral subtraction, band-pass modulation algorithm and the proposed algorithm. In car noise, the band-pass modulation filtering, the Wiener filter, and the proposed algorithm improve the quality of speech almost equally. However, in babble noise, the proposed algorithm shows better performances than the other three benchmark algorithms. For instance, in −5 dB SNR with babble noise, the proposed algorithm PESQ score was 1.62 while the Wiener algorithm scored 1.47, the modulation filtering scored 1.36 and the spectral subtraction scored 1.19.

4.2. Subjective test results

4.2.1. Musical noise results

Subjective results conducted for the evaluation of the proposed and benchmark algorithms in terms of musical noise perception were averaged over all participants and signals. The proposed algorithm was chosen 96.3% of the time to be the algorithm with the least musical noise, while spectral subtraction, the Wiener filter and modulation filtering were chosen 1.1%, 1.5% and 1.1% of the time respectively.

4.2.2. Overall quality results

The results of the subjective test evaluating the overall quality of processed and unprocessed signals are illustrated in Table 1. Overall, in both noise environments (car and babble) and the three SNRs (5, 0 and −5 dB), the proposed algorithm was chosen to be the algorithm with the best overall quality 78.8% of the time.

5. Hardware implementation

The proposed algorithm was designed to be implemented in real-time on an embedded system. To validate this goal, we show in this section implementation details of the algorithm on a low power DSP. The main purpose of this hardware implementation is the development of a smart hearing protection device that enables enhanced speech signals to be transmitted to the ear while protecting the S-HPD wearer from background noise.

5.1. DSP overview

The DSP used for the implementation of the proposed algorithm was provided in a small 32-lead, 5 mm × 5 mm package. The Analog to Digital Converter (ADC) and the Digital to Analog Converter

Table 1

Overall quality results for the proposed algorithm, the three benchmarks (SS corresponds to the spectral subtraction, MF to the modulation filtering, and W to the Wiener filter) and the noisy signals, in car and babble noise with 5, 0 and −5 dB SNRs. Results indicate the proportion of users who preferred each algorithm for a given combination of noise and SNR conditions.

<table>
<thead>
<tr>
<th>Noise environments</th>
<th>Proposed (%)</th>
<th>SS (%)</th>
<th>MF (%)</th>
<th>W (%)</th>
<th>Noisy (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>82.2</td>
<td>0.0</td>
<td>15.5</td>
<td>2.3</td>
<td>0.0</td>
</tr>
<tr>
<td>0</td>
<td>80.0</td>
<td>4.4</td>
<td>8.9</td>
<td>0.0</td>
<td>6.7</td>
</tr>
<tr>
<td>−5</td>
<td>66.7</td>
<td>8.9</td>
<td>17.8</td>
<td>4.4</td>
<td>2.2</td>
</tr>
<tr>
<td>Babble</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>97.7</td>
<td>0.0</td>
<td>2.3</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0</td>
<td>86.6</td>
<td>4.4</td>
<td>2.3</td>
<td>6.7</td>
<td>0.0</td>
</tr>
<tr>
<td>−5</td>
<td>60.0</td>
<td>6.7</td>
<td>20.0</td>
<td>4.4</td>
<td>8.9</td>
</tr>
<tr>
<td>Average</td>
<td>78.8</td>
<td>4.1</td>
<td>11.1</td>
<td>3.0</td>
<td>3.0</td>
</tr>
</tbody>
</table>
(DAC) are 24 bit stereo audio converters. They can operate at sampling frequencies ranging from 8 kHz to 96 kHz. In the address space of the DSP three RAMs are encompassed: a program RAM, a coefficient RAM, and a data RAM. The program RAM cannot exceed 1024 instructions per audio frame and governs the execution of the instructions in the core. The parameter RAM stores the initial coefficients of the program and cannot exceed 1024 coefficients. The data RAM is divided into two memory addressing types: modulo and non-modulo memories. Each of the modulo and non-modulo data RAM offer 4096 memory words. This RAM stores audio data-words for processing in addition to some run-time parameters.

### 5.2. Hardware implementation

The DSP and other associated electronics such as audio inputs, audio outputs, and battery are integrated in an Auditory Research Platform (ARP) [26]. This platform is illustrated in Fig. 6. Two earpieces are connected to this platform, and in each earpiece, an external miniature microphone and an internal miniature loudspeaker are integrated for external sound acquisition and sound transmission.

The hardware implementation of the noise reduction algorithm is made following the steps described in Section 2. The resulting number of instructions per audio frame is 333, which is equivalent to a rate of 32.5% from the entire program RAM. The data RAM used by the algorithm is 140 (3.4% from the entire modulo data RAM, and 0% from the non-modulo data RAM), while the coefficient RAM used is 124 (12.1% of the coefficient RAM).

### 5.3. Real-time test

Real-time tests of the proposed algorithm were performed using some noisy speech signals. For this purpose, the audio input of the ARP was connected to the audio output of a computer in which the noisy signals were playing, while the enhanced signals were saved in the computer. Fig. 7 shows the enhancement in the temporal and spectral domains.

### 6. Conclusions

This paper presented a noise reduction algorithm for the development of a smart hearing protection device that enables the transmission of enhanced speech while protecting the S-HPD wearer from noise. It demonstrated that noise reduction and speech quality improvement can be performed using a time-varying and frequency-band dependent gain function estimated from the low-pass filtering of the temporal envelope. The proposed method overcomes two types of problems in speech enhancement: one, the musical noise generally associated with processing in the frequency domain, and two, the amplification and attenuation distortions caused by an imperfect SNR and noise parameter estimation. Objective and subjective results show that the proposed algorithm improves the perceptual quality of speech signals without prior estimation of the noise or speech parameters. The hardware implementation of the proposed algorithm validates its reliability and low complexity for the intended real-time application. The hardware resources show that other tasks can be combined to the noise reduction method such as a voice activity detection algorithm to discriminate between speech and noise and transmit enhanced speech signals to the protected ear, while keeping the ear protected from noise when speech is not present. The proposed solution can also be integrated into active noise control headphones, which are already equipped with external micro-
Acknowledgments

The authors would like to thank Prof. Tiago H. Falk for sharing the code of the modulation filtering algorithm. This work was supported by Sonomax Technologies Inc. and its “Industrial Research Chair in In-ear Technologies”.

References