Applied Acoustics 109 (2016) 37–43

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

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

ARTICLE INFO

Article history:
Received 10 September 2015
Received in revised form 6 December 2015
Accepted 1 March 2016
Available online 14 March 2016

Keywords:
Noise reduction
Speech enhancement
Adaptive gain control
Temporal envelope
Musical noise

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.

1. Introduction

Nowadays, wearing Hearing Protection Devices (HPD) in workplaces and noisy environments becomes a necessity for people exposed to high noise levels on a daily basis, to protect them from what otherwise would damage the inner ear and induce hearing loss. Yet, in reality, most wearers will not use their HPD when oral communication is needed [1]. To palliate this problem, we intend to develop a smart HPD (S-HPD) that guarantees protection and discriminates between speech and noise to allow the transmission of enhanced speech signals to the protected ear. For this purpose, the integration of a Digital Signal Processor (DSP), an external microphone and an internal loudspeaker in a passive HPD are required [2]. With one external microphone, single-channel speech enhancement can be performed.

While multi-channel speech enhancement takes advantage of the spatial audio information, single-channel speech enhancement does not benefit from these information and therefore remains a challenging task, especially in low Signal-to-Noise Ratio (SNR). Single channel speech enhancement algorithms can be grouped into four main types [3]: spectral subtractive, linear estimators, non-linear estimators, and subspace algorithms. Despite significant differences in the type of estimated parameters, most of the aforesaid speech enhancement algorithms use the signal's first frames to estimate the noise parameters and update these parameters in non-speech segments using a Voice Activity Detector (VAD). However, most existing VAD algorithms are unreliable in low SNRs [4].

Single-channel speech enhancement algorithms usually perform in the spectral domain [5,6]. However, reducing the noise in the spectral domain may generate musical noise due to a random amplification of frequency bins that varies over time [7,8]. In some cases, musical noise is more annoying than the background noise itself. Much research has been conducted for the reduction of the musical noise using post-filtering techniques or image processing approaches such as in [7,9,10].

Other single-channel speech enhancement algorithms based on modulation filtering have been proposed such as [11–13]. These methods require the constant computation of Discrete Fourier Transforms (DFTs) and Inverse DFTs (IDFTs), making their use incompatible for real-time, low-latency applications in embedded systems.

In [14] a time domain speech enhancement algorithm based on an Adaptive Gain Equalizer (AGE) has been proposed. In this algorithm, a gain is applied for each frequency band based on an SNR estimation to boost the speech signal when the SNR in the

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

http://dx.doi.org/10.1016/j.apacoust.2016.03.001
0003-682X/

frequency 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 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, only 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 $w(n)$:

$$y(n) = x(n) + w(n) \tag{1}$$

The incoming signal is divided into $M = 16$ frequency bands using fourth order band-pass Butterworth filters. Filter bandwidth 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:

$$y_m(n) = (y * h_m)(n) \tag{2}$$

with $h_m(n)$ the impulse response of the mth band-pass filter, and the symbol "*" denoting convolution.

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

$$E_m(n) = \sqrt{y_m(n)^2 + \tilde{y}_m(n)^2} \tag{3}$$

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

$$\tilde{y}_m(n) = y_m(n) * \frac{1}{\pi n} \tag{4}$$

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

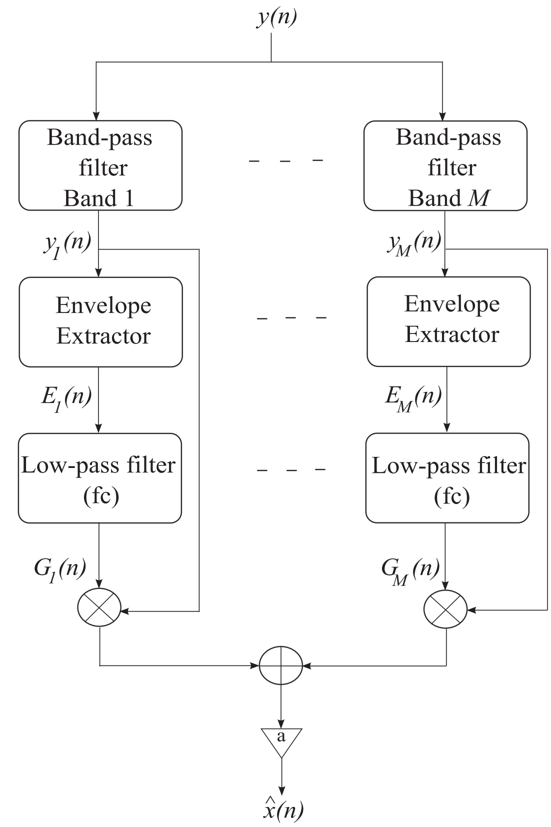

Fig. 1. Block diagram of the proposed speech enhancement algorithm.

- 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:

$$G_m(n) = (E_m * L)(n) \tag{5}$$

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:

$$\hat{x}_m(n) = G_m(n) \cdot y_m(n) \tag{6}$$

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 $\hat{x}(n)$ 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