A Demonstration of a Single Channel Blind Noise Reduction Algorithm with Live Recordings

Narimene Lezzoum, Ghyslain Gagnon, and Jérémie Voix
École de technologie supérieure
Université du Québec, Montréal (Qc) Canada
narimene.lezzoum@ens.etsmtl.ca, (ghyslain.gagnon)(jeremie.voix)@etsmtl.ca

Abstract

Currently, most noise reduction algorithms are based on an a priori information such as signal-to-noise ratio (SNR) or noise parameters estimation. They are mostly performed in the spectral domain to reduce the background noise at each frequency bin. However noise reduction in the spectral domain may introduce musical noise and artefacts which are in some cases perceptually more annoying than the background noise itself. In this "show and tell", we present a demonstration of a noise reduction algorithm based on dynamic range compression (DRC) using a time-varying and frequency-band dependant gain function deduced from the low-pass filtering of the temporal envelopes. The algorithm is considered as blind since it requires neither SNR nor noise parameters estimation. A graphical user interface (GUI) built under Matlab shows interactively the noise reduction in the temporal (waveform) and spectral (spectrogram) domains using live speech recordings mixed to pre-recorded noise signals.

1. Introduction

Noise reduction algorithms are nowadays used in multiple areas such as hearing aids, cochlear implants, telecommunication systems and human/robot interaction devices. Most of existing noise reduction algorithms perform in the spectral domain in order to reduce the background noise differently in each frequency bin, for instance, the spectral subtraction [1], the Wiener filter [2] [3], and the bandpass modulation filtering [4]. However, enhancing the speech in the spectral domain may introduce musical noise which is well known in the field of speech enhancement, and represents a random amplification of frequency bins [5].

Anderson [6] proposed a frequency-band dependant and time-varying gain function instead of frequency-varying gain function for fast dynamic range compression (DRC), also suggesting that the concept of frequency-band time-varying gain function can be used in some audio processing systems such as noise reduction. However, to our knowledge, methods and results of such as approach has never been demonstrated for noise reduction applications.

In this "show and tell", a noise reduction algorithm using frequency-band dependant and time-varying gain function is proposed. The proposed method employs dynamic range compression theory in order to reduce the dynamic range independently in each frequency band using a time-varying gain function. This function is deduced from the temporal envelope and tends to preserve the natural quality of the incoming signal.

We demonstrate that the use of a time-varying and frequency-band dependant gain function enables noise reduction and speech quality improvement without introducing musical noise. In addition, this demonstration shows that speech enhancement can be performed without any knowledge, assumption, or estimation of the noise parameters.

2. Scientific and Technical Description

Figure 1 illustrates the architecture of the proposed noise reduction algorithm.

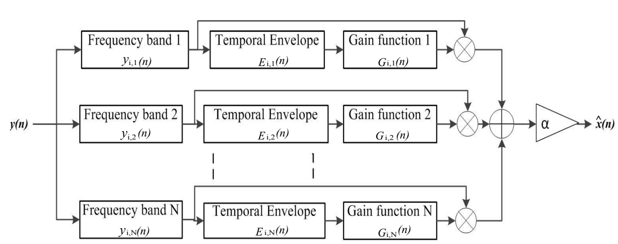

Figure 1. Block diagram of the proposed speech enhancement algorithm.

The proposed algorithm performs in real-time using

250 ms frames with 80% overlap.

The incoming signal is decomposed into N=16 frequency-bands using a gammatone filterbank [7]. From each frequency-band, the temporal envelope is extracted using the Hilbert Transform [8]:

$$E_{i,m}(n) = \sqrt{y_{i,m}(n)^2 + \tilde{y}_{i,m}(n)^2} \tag{1}$$

with i the frame number and m the frequency-band number, and:

$$\tilde{y}(n) = y(n) * \frac{1}{\pi n} \tag{2}$$

with * denoting the convolution.

A gain function is deduced from the temporal envelope of each frequency-band (see section 2.1), and is thereafter multiplied by the incoming signal $y_{i,m}(n)$ of the same frequency-band. The enhanced signal $\hat{x}(n)$ of each frame is reconstructed by summing the 16 frequency-bands, and amplified by a rescaling constant α.

2.1. Time-Varying Gain Function Calculation

When combining the concepts of noise reduction and DRC used in hearing aids [9], [6], a multi-band time-varying noise reduction method can be obtained. According to preliminary results in our research, the multi-band time-varying gain function for noise reduction must meet three criteria:

- The gain function of each frequency-band should 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 $E_{i,m}(n)$ 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 fulfils all the above cited criteria is a low-pass filtered temporal envelope, which represents a smoothed version of the temporal envelope $E_{i,m}(n)$:

$$G_{i,m}(n) = E_{i,m}(n) * L(t) \tag{3}$$

with $L(t)$ the impulse response of a lowpass filter with a 16 Hz cut-off frequency.

3 Objective Validation of the Proposed Method

The proposed method is evaluated using 30 noisy speech signals corrupted by “car” and “babble” noise in 5, 0, and -5 dB SNR from the Noizeus corpus [10]. The performance improvement of the proposed algorithm is compared to noisy signals in addition to a modulation filtering speech enhancement algorithm [4] (benchmark algorithm) using the Perceptual evaluation speech quality (PESQ) metric [11] (see figure 2).

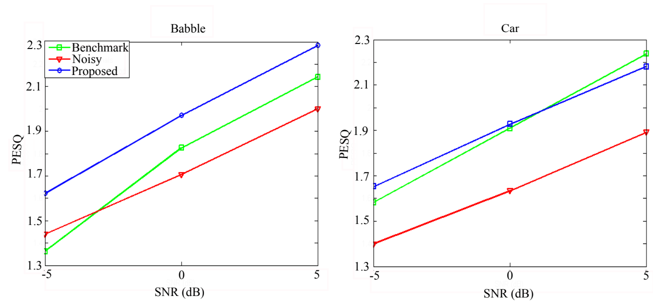

Figure 2. PESQ results for the unprocessed signals, the benchmark algorithm, and the proposed algorithm.

4. System Demonstration

A graphical user interface (GUI) is built in Matlab for an interactive demonstration of the proposed noise reduction algorithm. Figure 3 presents a screenshot of the user interface: part (1) shows the instructions that the user should follow when using pre-recorded speech signals, or recording a live speech signal. Part (2) shows the experimental settings, while part (3) displays the enhancement in the spectrogram and waveform frame by frame: part (3-a) illustrates the enhanced part while par (3-b) illustrates the noisy part which will be enhanced. Part (4) of the interface presents the PESQ results for each noisy/enhanced signal.

A video of this demonstration is available online: http://critias.etsmtl.ca/ts2014. This demonstration runs on a laptop with professional headphones, and a microphone for live speech recordings.

5 Conclusions and Future Developments

In this “show and tell”, we demonstrate that the use of a time-varying and frequency-band dependant gain function reduces the background noise and improves the quality of the speech signal. In addition, we show that good noise

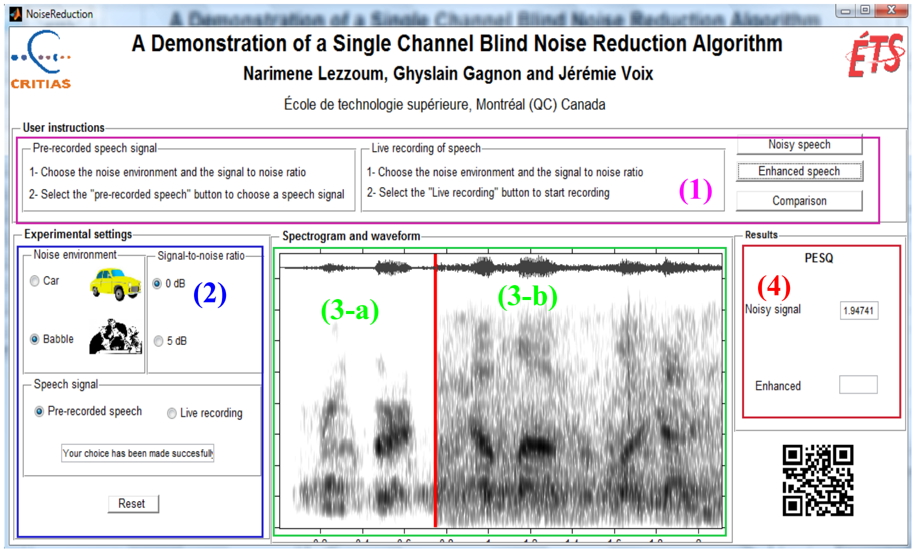

Figure 3. Screenshot of the graphical user interface

reduction performance can be achieved without any knowledge, assumption, or estimate of the noise and speech parameters.

We are currently working towards the implementation of this algorithm in a digital signal processor (DSP) for a real-world embedded application.

Acknowledgement

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

References

[1] S. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” *IEEE Transaction on Acoustics, Speech, and Signal processing*, no. 2, pp. 113–120, 1979.

[2] N. Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series with Engineering Applications.*, cambrigde, ed., 1949.

[3] J. Scalart, P. and Filho, “Speech enhancement based on a priori signal to noise estimation.” in *IEEE Int. Conf. Acoust. , Speech, Signal Processing (ICASSP)*, 1996, pp. 629–632.

[4] T. H. Falk, S. Stadler, W. B. Kleijn, and W.-y. Chan, “Noise suppression based on extending a speech-dominated modulation band,” in *INTERSPEECH 2007*, pp. 2–5.

[5] C. Leitner and F. Pernkopf, “Musical noise suppression for speech enhancement using pre-image iteration,” in *19th International Conference on Systems, Signals and Image Processing (IWSSIP),*, 2012, pp. 464 – 467.

[6] D. V. Anderson, “A modulation view of audio processing for reducing audible artifacts,” *2010 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, no. 1, pp. 5474–5477, 2010.

[7] A. Aertsen and P. Johannesma, “The Specto-temporal receptive field: a functional characteristics of auditory neurons,” *Biological Cybernetics*, vol. 143, pp. 133–143, 1981.

[8] S. L. Marple, “Computing the Discrete-Time ”Analytic” Signal via FFT,” *IEEE Transactions on Signal Processing*, vol. 47, no. 9, pp. 2600–2603, 1999.

[9] G. Kim and P. C. Loizou, “Gain-induced speech distortions and the absence of intelligibility benefit with existing noise-reduction algorithms.” *The Journal of the Acoustical Society of America*, vol. 130, no. 3, pp. 1581–96, Sep. 2011.

[10] Y. Hu and P. C. Loizou, “Evaluation of objective quality measures for speech enhancement,” *IEEE Transactions on audio, speech, and language processing*, vol. 16, no. 1, pp. 229–238, 2008.

[11] ITU-T, “Subjective test methodology for evaluating speech communication systems that include noise suppression algorithm.” Tech. Rep., 2003.