International Journal of Informatics and Applied Mathematics
e-ISSN:2667-6990 Vol. 3, No. 1, 70-83

Denoising Speech Signal Using Decision Directed Approach

Ouardia Abdelli and Fatiha Merazka

LISIC Laboratory, Telecommunications Department, USTHB University, Algiers, Algeria
ouar_ing80@hotmail.com and fmerazka@usthb.dz

Abstract. This article deals with the problem of improving speech in noisy environments using the decisional approach (DD). The decision approach (DD) uses a priori estimation of the signal-to-noise ratio (SNR) for speech improvement and is used to estimate the time-varying noise spectrum, which results in better performance in terms of intelligibility and a reduction in musical noise. In this article, we propose recursive estimators for the a priori SNR and the spectral components of speech. We introduce a new statistical model which takes into account the temporal correlation between the successive vocal spectral components, while keeping the resulting algorithms simple. This model provides new information on the DD approach and allows the extension of existing speech improvement algorithms.

Keywords: Spectral Attenuation · Signal and Noise · Suppression Rule · SNR · HRNR · TSNR

1 Introduction

With the growth of technology in the field of mobile telecommunications, the need to improve the sound, particularly by reducing the noise annoyance, became increasingly present. Noise reduction techniques are subject to a compromise between the actual level of reduction and distortion that affects the speech signal [1], [2]. On current performance, it is desirable to remove more noise while main tailing an acceptable level of degradation of the restored signal, especially when the noise level is important.

The quality of the speech signal transmitted to the remote party to increase its intelligibility and reduce fatigue of the latter, it appears important to develop noise reduction systems whose purpose is to extract useful information by performing a treatment on the noisy observation signal. In addition to these applications of spoken communication, improving the quality of the speech signal is also required for speech recognition, whose performance is highly altered when the user is immersed in a noisy environment [3].

The techniques that have generated the most interest in recent years are the short-term spectral attenuation approaches that involve modifying a short-term transform of the noisy signal using a suppression rule [1], [2]. The development of this family of techniques is mainly due to the fact that they allow to meet real time constraints and complexity inherent in applications of spoken communication ease of use. A popular statistical model for speech enhancement was proposed in [3]. Accordingly, the individual short-term spectral components of the speech and noise signals are modeled as statistically independent Gaussian random variables.

In this paper, the decision directed (DD) approach is used to estimate the time varying noise spectrum which results in better performance in terms of intelligibility and reduced musical noise. However, the a priori signal to noise ratio (SNR) estimator of the current frame relies on the estimated speech spectrum from the earlier frame. So, we can formulate a short time spectral gain using Wiener filtering with DD approach in which frame delay results in an annoying reverberation effect. The problem is solved by temporal SNR (TSNR), wherein, a second step is formulated so as to remove the delay and Harmonic Regeneration Noise Reduction (HRNR) algorithm which is used to regenerate the harmonics in the reconstructed signal.

This paper is organized as follows. In Section II, we present the parameters and rules of speech enhancement techniques by method short-term spectral attenuation; we introduce a tool useful to analyze the SNR estimators. In Section III, we recall the principle of the DD approach and analyze it and we present and analyze the TSNR and HRNR techniques. Finally, in Section III, we demonstrate the improved performance of the Harmonic Regeneration Noise (HRNR) and TSNR compared to Wiener and (MMSE) methods.

2 Additive noise model

The single-channel case considered, the goal is to estimate the useful speech signal $s(n)$ the latter being disturbed by additive noise $b(n)$ assumed to be independent of the speech signal, from the one observed signal $\chi(n)$ [1], [2], [3]. It can be discretized and mathematically represented as:

$$\chi(n) = s(n) + b(n) \tag{1}$$

The approaches based on short-term spectral attenuation realize the noise reduction in the frequency domain (or spectral). If the signals are stationary then from the temporal relationship can be written:

$$\gamma_\chi(f) = \gamma_s(f) + \gamma_b(f) \tag{2}$$

while $\gamma_\chi(f)$, $\gamma_(f)$, $\gamma_b(f)$ represent the power spectral densities (PSDs) [3], of the respective signals $\chi(n)$, $S(n)$ and $b(n)$. This representation of the power spectral density (PSD) is unfortunately not usable because of the non-stationary of speech signal. Indeed, if it is acceptable to consider the stationary noise, speech cannot be considered as over short durations. It then becomes possible to exploit the quasi-stationarity of the speech on frames of duration of the order of 20 to 40 ms. this is one reason why a majority of noise reduction techniques are based on spectral attenuation.

The Fourier transform (FT) and Short-term Fourier transform (STFT) are qualified and have allowed a fast and numerically inexpensive work[4].

Each frame from the time signal $\chi(n)$ can be represented in the frequency domain by its module $|\chi(p,k)|$ and its associated phase $\Phi_\chi(p,k)$ where p is the time index of the current analysis frame and k the frequency channel of index otherwise discrete frequency f_k in the frequency domain.

$$|\chi(p,k)|\mathrm{e}^{i\Phi_\chi(p,k)} = |S(p,k)|\mathrm{e}^{i\Phi_S(p,k)} + |B(p,k)|\mathrm{e}^{i\Phi_B(p,k)} \tag{3}$$

The purpose of the spectral attenuation then is to estimate the short-term spectrum of the speech signal $S(p,k)$. We assume that is still possible to estimate the PSD of the noise.

2.1 The SNR

SNR is a key parameter that governs the quality of noise reduction techniques. Its various estimators are, however, subject to certain limitations. Perform noise reduction using an ideal SNR estimator (knowing of course all signals). The result is startling quality. If one does not reach that of the clean signal it is however in very close. This test confirms that there is room for tremendous growth in the estimation of SNR [7].

In reality, there is no a simple solution to the spectral estimate of $S(p,k)$, so generally a spectral gain $G(p,k)$ that depends on the SNR is obtained and applied to the noisy spectrum $\chi(p,k)$:

$$\hat{S}(p,k) = G(p,k)\chi(p,k) \tag{4}$$

The spectral gain $G(p,k)$ has always the following asymptotic behavior. An important value of the SNR indicates that a strong speech component is presented with respect to the noise level, the gain $G(p,k)$ must therefore be close to 1 to preserve this component.

A low value of SNR indicates that speech is absent or very low compared to the noise level. The gain $G(p,k)$ must therefore make an important attenuation ($G(p,k) \ll 1$) to reduce the effect of noise.

The problem is thus to estimate the SNR. Depending on the assumptions chosen for expressing the spectral gain, two types of SNR are used, the priori and posteriori SNR [5], [6].

$$SNR_{post}(p,k) = \frac{|\chi(p,k)|^2}{\hat{\gamma}_b(k)} = \frac{|\chi(p,k)|^2}{E[|B(p,k)^2|]} \tag{5}$$

$$SNR_{prio}(p,k) = \frac{\hat{\gamma}_s(k)}{\hat{\gamma}_b(k)} = \frac{E[|S(p,k)^2|]}{E[|B(p,k)^2|]} \tag{6}$$

The quantity of $SNR_{post}(p,k)$ represents the SNR of the current frame taking into account the modulus squared of the noisy signal and therefore time dependent.

The quantity of $SNR_{prio}(k)$ for its part does not depend on time because it expresses the long-term SNR assuming statistics useful speech signal known a priori.

From SNR_{post}, we can also set the instantaneous SNR corresponding to a local estimate (or short-term) of the a priori SNR by DD of the PSD noise squared modulus of the noisy signal [5], [6].

$$SNR_{inst}(p,k) = \frac{|\chi(p,k)|^2 - \gamma_b(k)}{\gamma_b(k)} = \frac{|\chi(p,k)|^2 - E[|B(p,k)^2|}{E[|B(p,k)^2|} = SNR_{post}(p,k) - 1 \tag{7}$$

These three terms are theoretical SNR in so far as only the quantity $|\chi(p,k)|^2$ is known. First, the quantity $E[|B(p,k)^2|$ must be estimated from the noisy signal which is devoted to estimation techniques PSD noise. On the other hand, we also estimate the quantity $E[|S(p,k)^2|$.This estimate is quite problematic and gives rise to various techniques for estimating the SNR a priori [5].

3 The DD approach

The estimate of the PSD of noise is an indispensable preliminary to calculate the SNR and posterior SNR are then obtained as follows [8]

$$\widehat{SNR}_{post}(p,k) = \frac{|\chi(p,k)|^2}{\hat{\gamma}_b(k)} \tag{8}$$

$$\widehat{SNR}_{prio}^{DD}(p,k) = \beta \frac{|\hat{S}_{DD}(p-1,k)|^2}{\hat{\gamma}_b(k)} + (1-\beta) max(\widehat{SNR}_{post}(p,k) - 1, 0) \quad (9)$$

$$\widehat{SNR}_{prio}^{DD}(p,k) = \beta \frac{|\hat{S}_{DD}(p-1,k)|^2}{\hat{\gamma}_b(k)} + (1-\beta) max(\widehat{SNR}_{post}(p,k) - 1, 0) \quad (10)$$

Where $\hat{S}_{DD}(p-1,k)$ is the spectrum of the speech signal estimated in the previous frame. This estimator SNR a priori appointed decision-directed (DD) which means directed by the decision, was proposed in [6] and its behavior is controlled by the parameter b (always close to 1 and typically 0, 98). Finally, in general, the spectral gain is a function that depends on the SNR and optionally a priori SNR

$$G_{DD}(p,k) = g(\widehat{SNR}_{prio}^{DD}(p,k), \widehat{SNR}_{post}(p,k)) \quad (11)$$

The spectrum of the restored speech signal is then obtained

$$\hat{S}_{DD}(p,k) = G_{DD}(p,k)\chi(p,k) \quad (12)$$

In the following, by default, the chosen gain functions correspond to the Wiener filter which leads to:

$$G_{DD}(p,k) = \frac{\widehat{SNR}_{prio}^{DD}(p,k)}{1 + \widehat{SNR}_{prio}^{DD}(p,k)} \quad (13)$$

The diagram of Figure 1 summarizes the principle of the DD approach.

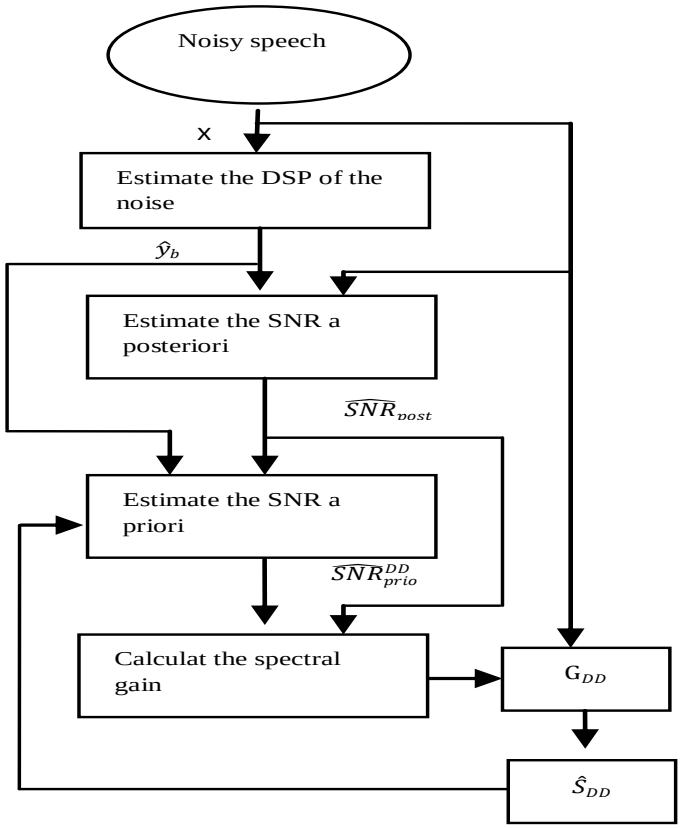


Fig. 1. Diagram of the general principle of the DD approach

3.1 Proposed algorithm

1. The spectral attenuation is to estimate the short-term spectrum of the speech signal $\hat{S}(p,k)$ [1], [2].
2. It is assumed that it is always possible to estimate DSP noise [1], [2].
3. perform noise reduction using an estimator ideal SNR, usually a spectral gain $G(p,k)$ that depends on the SNR is obtained and applied to the noise spectrum $\chi(p,k)$:

$$\hat{S}(p,k) = G(p,k)\chi(p,k) \tag{14}$$

4. Determination of the VAD, voice activity detection.
5. A short-term spectral attenuation or spectral gain calculation requires the estimate of the SNR (a posteriori and / or a priori) [3], [4].
6. Two types of SNR are used, the SNR posteriori and SNR apriori [3], [4].
7. Refine the estimate of the a priori SNR by TSNR approach: this is a technique in two passes corresponding to the DD estimator [6], [7], [8], [9], [12].

8. Estimation errors PSD noise and the impact of Phase generate harmonic distortion corresponds approach HRNR. [9], [10], [11].
9. The speech signal module and estimated —S (p, k)— and the phase of the noisy signal are then used to return to the time domain using an inverse DFT (IDFT) [12].
10. The output signal is finally synthesized from a treatment technique by OLS block type (for overlap and save) and OLA (for overlap and add) This last step is the inverse STFT (TFCTI) [13]

3.2 Test and Results

In this section, the each variant of approach DD method is evaluated and compared with other variants. The speech datasets used in our simulations are from the NOIZEUS corpus18 [13].

The corpus is sampled at 16 KHz and 8KHz, filtered to simulate receiving frequency characteristics of telephone handsets.

Noise signals have different time-frequency distributions, and therefore a different impact on clean signal. For that reason, the NOIZEUS comes with various non-stationary noises at different levels of SNRs. The non-stationary noises are babble, train. In our evaluation, we have used the speech degraded by babble noise at global SNR levels of -5 dB and 5 dB. We also generate a corresponding stimulus set degraded by additive white Gaussian noise (AWGN), stationary noise, at two SNR levels: 5 dB. The performance of this method, tests on such noisy speech samples. Consider the signal speech "ns_bab_m5dB_S_01_01_16KHz" and the noisy speech "clean_S_01_01_16KHz" at -5db and Signal "clean_S_01_01.wav" and the noisy speech "ns_bab_m5dB_S_01_01.wav". Figures 2 and 3 below represent the Periodograms of the speech signals.

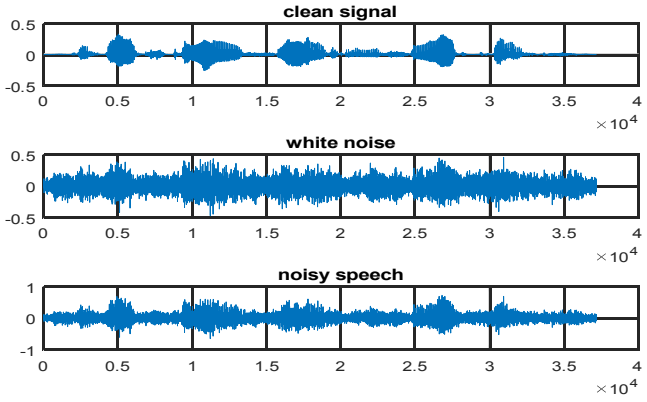


Fig. 2. Periodogram of clean signal1 "clean_S_01_01" and white noise and noisy speech

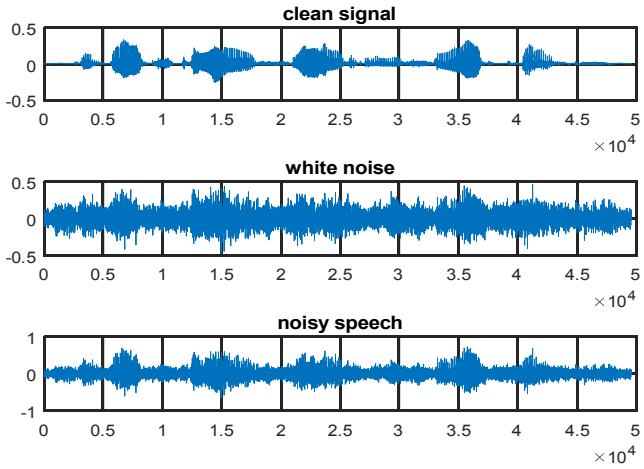


Fig. 3. Periodogram of clean signal2 "clean_S_01_01_16KHz.wav" and white noise and noisy speech

3.3 Results analysis

To illustrate the behavior and performance of the implemented techniques, the spectrograms after each step are plotted as shown in Figures 4 to 7.

Figures 4 (a) and 4 (b) show the spectrograms of the original, noisy speech signal and enhanced by the DD approach of two signals (signal1 and signal2) where we can observe that in the signal spectrogram noisy, the yellow color noise distributed over the entire spectrogram unlike the enhanced signal, the yellow and diminished color and similarly for Figures 6 (a), 6 (b) give the spectrograms of the DD approach with TSNR and HRNR, these figures show that speech is improved by removing much of the noise.

Figures 5 (a) and 5 (b) and Figure 6 (a) and 6 (b), give the spectrograms of the original signal, noisy and enhanced by two signals (signal1 and signal2) by the Wiener method and the logMMSE method these figure present an important space of the yellow color therefore the noise is always important in the restored signals.

Figure 7 shows speech improvement using Wiener filtering with the DD approach where we can see that in addition to noise suppression compared to Wiener filtering and logMMSE, some of the harmonics are suppressed. In Figure 7 (a), we can observe that noise elimination is better with the TSNR approach than that with the DD approach, but the harmonics are still not preserved. The Figure 7 (b) shows the improvement of speech spectrogram using the DD approach with TSNR.

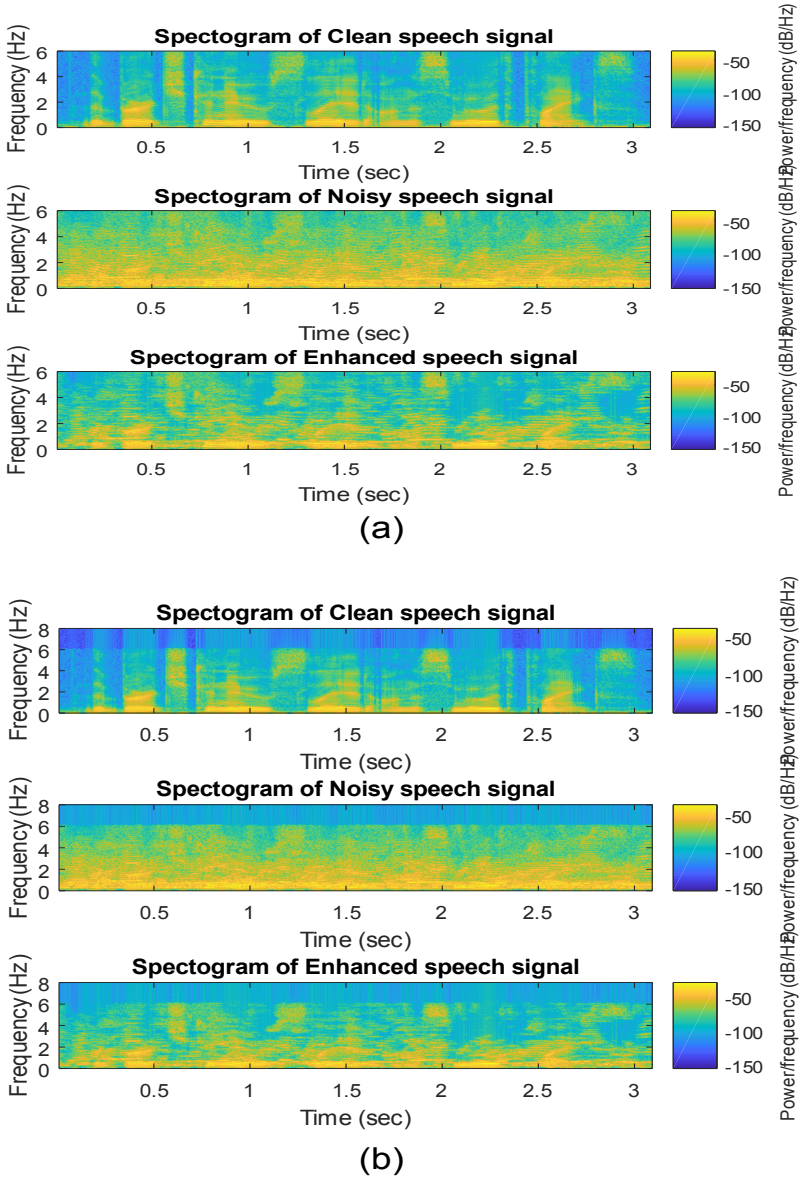
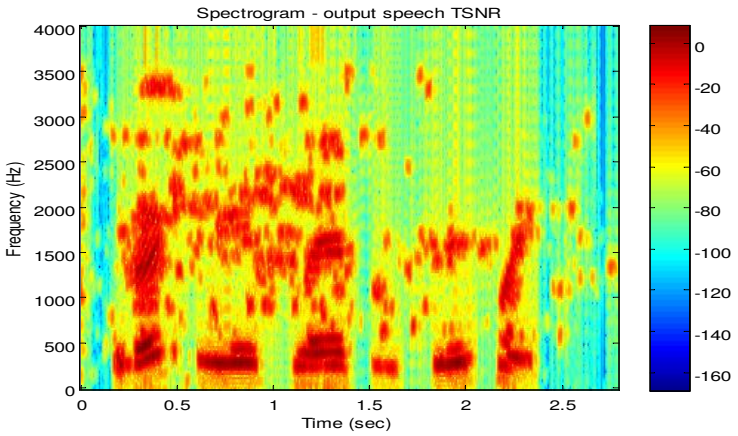
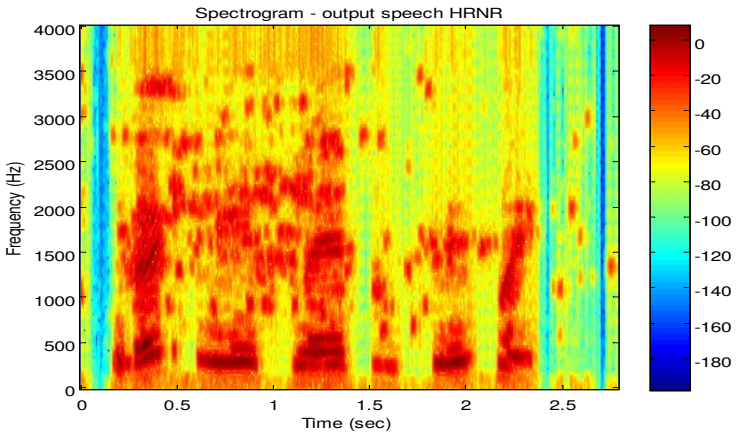


Fig. 4. The spectrogram, clean speech, noisy speech and enhancment speech of **(a)** signal1 clean_S_01_01 **(b)** signal2 clean_S_01_01_16KHz with DD approach



(a)



(b)

Fig. 5. Spectrograms **(a)** TSNR and **(b)** HRNR

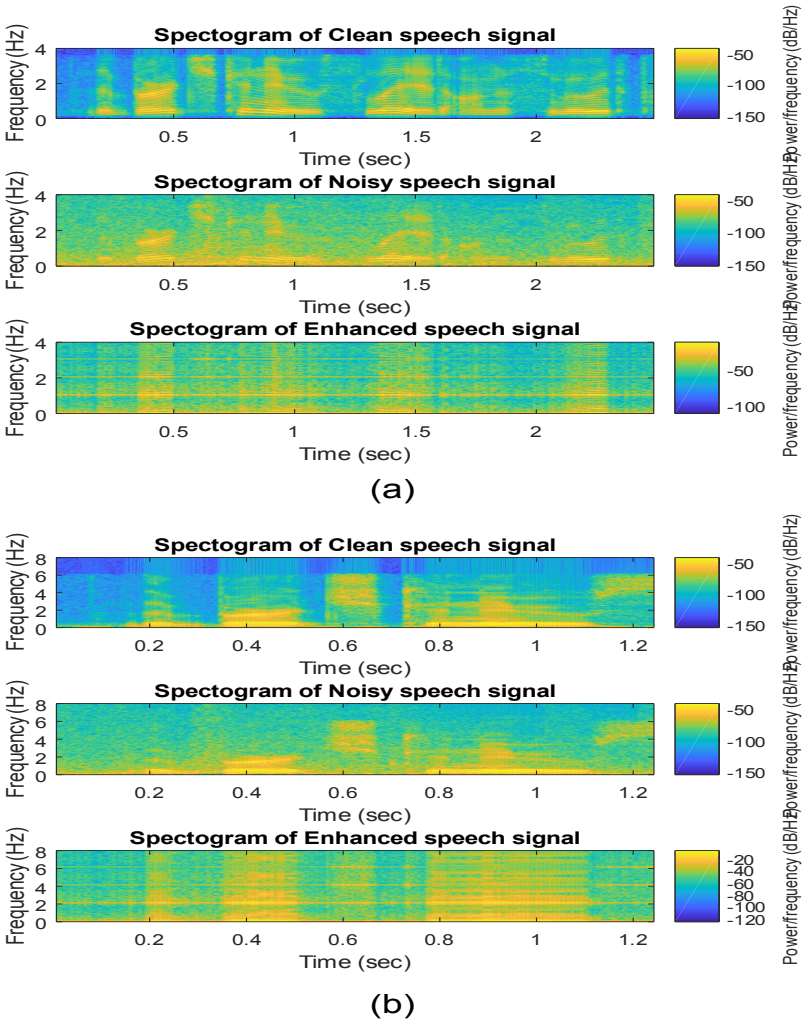


Fig. 6. Spectrogram of the **(a)** Signal1 "clean_S_01_01" and noisy speech and enhanced speech, **(b)** signal2 "clean_S_01_01.16KHz" and noisy speech and Enhanced speech with Wiener method

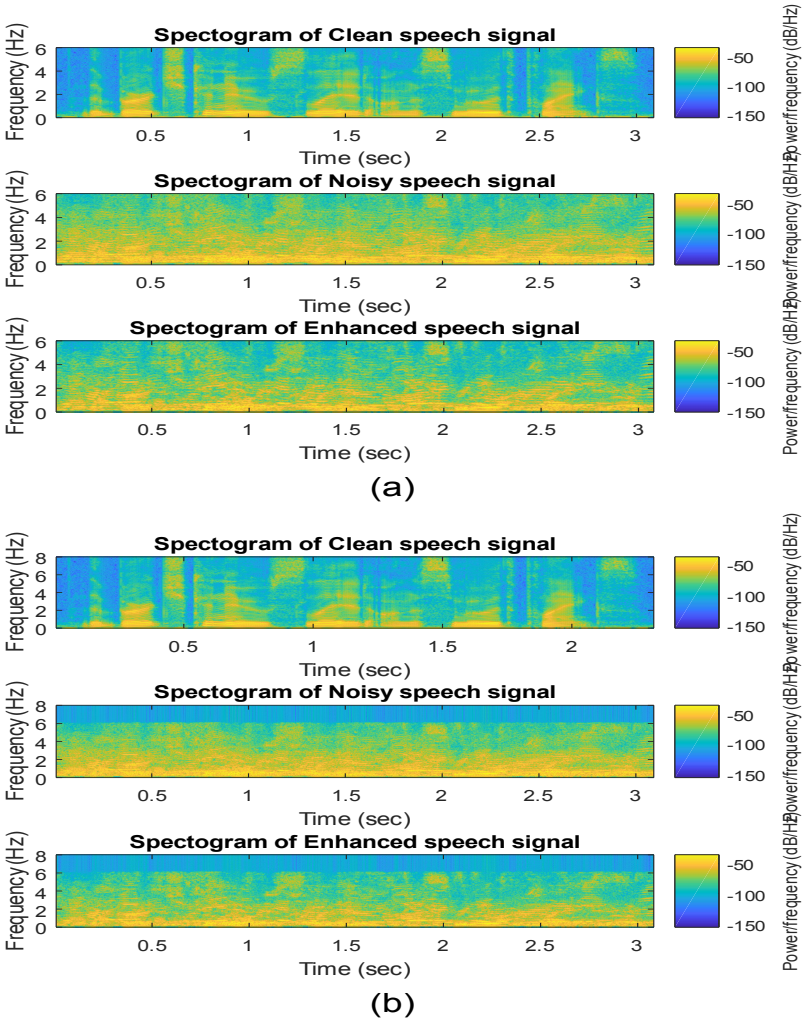


Fig. 7. Spectrogram of the **(a)** Signal1 "clean_S_01_01.wav" and noisy speech and enhanced speech, **(b)** signal2 "clean_S_01_01.16KHwav" and noisy speech and Enhanced speech with logMMSE method

The approaches essentially seek to further preserve the component of the speech signal while improving the quality of the restored signal, according to the programming results and the evaluation parameters, such as mean squared error (MSE), mean absolute error (MAE), signal to noise ratio (SNR), peak signal to noise ratio (PSNR), cross correlation (cross_core) give in Table 1, we notice that the DD approach gives better results for all parameters by the two signals (signal1, signal2) except the SNR, the Wiener method which gives better results.

Note that the lower the PSNR and mse values, the smaller the error, as shown by the results of the DD approach. And for the results of the cross core is better for the DD approach because it is very close to 1 (the more Cross_Core is close to 1 the closer the information on the signals), and the lower the signal to noise ratio the lower the component speech signal is greater than that of noise as shown by Wiener's result.

Table 1. Evaluation with five objective parameters

Evaluation	**Signal**	**MSE**	**MAE**	**SNR (db)**	**PSNR (db)**	**Cross_Core**
LogMMSE	Signal1	0.000718	0.060008	-4.674853	12.085624	0.827319
	Signal2	0.000759	0.060342	-4.700813	12.058158	0.826014
Wiener	Signal1	0.000290	0.009199	0.018727	17.898187	0.522226
	Signal2	0.002357	0.009758	0.096314	16.633584	0.767327
DD	Signal1	0.000120	0.009564	1.5436945	10.526052	0.999506
	Signal2	0.000235	0.009589	1.4815496	9.0929756	0.999554

4 Conclusion

We presented an analysis of the DD approach, and evaluated its performance by a comparison with Weiner and logMMSE methods.

Its algorithm consists of two steps, the first step ensures the reduction of the musical noise while the second step ensures the suppression of the frame delay but this approach however has a major defect. Harmonic distortion due to noise PSD estimation errors, the TSNR and HRNR approaches make it possible to limit the defects of the DD approach, thus making it possible to suppress the reverberation effect and to further reduce the level of noise. of musical noise.

The TSNR approach is chosen because of its simplicity, its low complexity and of course its effectiveness as a basis for the HRNR approach which makes it possible to overcome the limitations related to noise and phase PSD estimation problems. These essentially results in the distortion of the harmonics of the speech signal. The HRNR approach allows to regenerate the harmonics destroyed by conventional approaches using nonlinear processing of the distorted signal. In practice, the limitation of the harmonic distortion of the speech signal makes it possible to suppress more noise than with a conventional technique.

The results, in terms of spectrographic analysis, objective and subjective tests, are provided to evaluate the performance of various techniques. All the results show that the TSNR DD approachs followed by the HRNR technique has the best performance among the others analyzed in terms of objective and subjective tests.

References

1. Ephraim, Y., Mallah, D.: Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. IEEE Transactions on Acoustics, Speech, and Signal Processing, **32**(6), 1109-1121 (1984)
2. Shekar, S., Ravi, D.-J.: Denoising of a Speech Signal using Wiener Filter. In: Third International Conference on Current Trends in Engineering Science and Technology, (2017)
3. Loizou, P.-C., Kim, G.: Reasons why current speech-enhancement algorithms do not improve speech intelligibility and suggested solutions. IEEE Transactions on Audio, Speech, and Language Processing, **19**(1), 47-56 (2011)
4. Kim, G.: Binary Mask Criteria Based on Distortion Constraints Induced by a Gain Function for Speech Enhancement. IEIE Transactions on Smart Processing and Computing, **2**(4), 197-202 (2013)
5. Scalart, P., Filho, J.-V.: Speech enhancement based on a priori signal to noise estimation. In: IEEE International Conference on Acoustics, Speech and Signal Processing, **2**,629632 (1996)
6. Zhao, Y., Wang, Z.-Q., Wang, D.: A Two-Stage Algorithm for Noisy and Reverberant Speech Enhancement. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 5580-5584 (2017)
7. Plapous, C., Marro, C., Scalart P.: Speech enhancement using harmonic regeneration. In: IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'05), Philadelphie, **1**, I-157 (2005)
8. Wang, J., Yang, C., Yan, L., Huang, M., Sang, J.: Speech Enhancement Algorithm of Binary Mask Estimation Based on a Priori SNR Constraints. In: 2018 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC), IEEE, 937-943 (2018)
9. Plapous, C., Marro, C., Mauuary, L., Scalart, P.: A two-step noise reduction technique. In: 2004 IEEE international conference on acoustics, speech, and signal processing, **1**, I-289 (2004)
10. Plapous, C.: Traitements pour la réduction de bruit. Application à la communication parlée. PhD dissertation, University of Rennes 1 (2005)
11. Scalart, P., Lepauloux, L.: On the convergence behavior of recursive adaptive noise cancellation structure in the presence of crosstalk. In: Sensor Signal Processing for Defence Conference (SSPD2010), London, IET, 1-5 (2010).
12. Crochiere, R.-E., Rabiner, L.-R.: Multirate Digital Signal Processing. Prentice-Hall Signal Processing Series: Advanced monographs, Prentice-Hall, First edition (1983)
13. ITU-T Recommandation, Telephone Transmission Quality Objective Measuring Apparatus, pp. 56, (1996)
    `https://www.itu.int/ITU-T/recommendations/rec.aspx?rec=11830&lang=en`