Reducing over- and under-estimation of the \textit{a priori} SNR in speech enhancement techniques

Mohamed Djendi\textsuperscript{a,}\textsuperscript{*}, Pascal Scalart\textsuperscript{b}

\textsuperscript{a} Blida University—LATSI Laboratory, 09000, Blida, Algeria \\
\textsuperscript{b} University of Rennes—IRISA/ENSSAT, 22305 Lannion, France

\textbf{Abstract}

Most speech enhancement methods based on short-time spectral modification are generally expressed as a spectral gain depending on the estimate of the local signal-to-noise ratio (SNR) on each frequency bin. Several studies have analyzed the performance of \textit{a priori} SNR estimation algorithms to improve speech quality and to reduce speech distortions. In this paper, we concentrate on the analysis of over- and under estimation of the \textit{a priori} SNR in speech enhancement and noise reduction systems. We first show that conventional approaches such as the decision-directed approach proposed by Ephraïm and Malah lead to a biased estimator for the \textit{a priori} SNR. To reduce this bias, our strategy relies on the introduction of a correction term in the \textit{a priori} SNR estimate depending on the current state of both the available \textit{a posteriori} SNR and the estimated \textit{a priori} one. The proposed solution leads to a bias-compensated \textit{a priori} SNR estimate, and allows to finely estimating the output speech signal to be very close to the original one on each frequency bin. Such refinement procedure in the \textit{a priori} SNR estimate can be inserted in any type of spectral gain function to improve the output speech quality. Objective tests under various environments in terms of the Normalized Covariance Metric (NCM) criterion, the Coherence Speech Intelligibility Index (CSII) criterion, the segmental SNR criterion and the Perceptual Evaluation of Speech Quality (PESQ) measure are presented showing the superiority of the proposed method compared to competitive algorithms.

\section{Introduction}

The problem of enhancing speech degraded by uncorrelated additive noise has been widely studied in the past and is still an active field of research. Although background and applications differ largely, a specific class of these methods is essentially based on the SNR estimation that is used with a gain function to correct the corrupted speech signal. In practice, the deficiency of speech acquisition and transmission system results a speech signal commonly corrupted by noise. The contamination of the speech not only affects its audio quality but also reduces the performance of high-level speech processing applications such as voice communication and automatic speech recognition where efficient noise reduction techniques are required. Depending of the application, the goal of the speech enhancement techniques is to reduce the noise to make the speech intelligible, to decrease annoyance or to improve the overall sound quality. The most challenging case is the single-channel speech enhancement where only a single noisy speech recording is available for recovering the clean speech. Many algorithms have been proposed to solve this problem, such as spectral subtraction (SS) [1], minimum-mean square error (MMSE) estimator [2,3], and Wiener filter based algorithms [4,5]. Such noise reduction techniques rely mainly on the estimation of a short-time spectral gain which is a function of the \textit{a priori} SNR and/or the \textit{a posteriori} SNR computed on each frequency component. In [6], the authors highlighted the interest of estimating the \textit{a priori} SNR thanks to the “decision-directed” (DD) approach initially proposed in [2]. This DD approach has then received a lot of attention due to its low computational complexity and good performance in various noise reduction applications. Unfortunately, the DD approach suffers from inherent errors (or bias) in estimating the spectral SNRs. Such errors affect directly the estimation of the short-time spectral gain, and inaccuracies in estimating the gain function often lead to spectral attenuation (i.e., enhanced spectral components are smaller in magnitude than corresponding clean spectral components) and/or spectral amplification (i.e., enhanced spectral components are larger in magnitude than corresponding clean spectral components).
component). Consequently, the output speech intelligibility may be impacted by the presence of these two kinds of distortions.

A first source of bias in the \textit{a priori} SNR estimate is due to the one-frame delay bias introduced by the decision-directed estimator. This behavior has been analyzed in [5] where the authors have demonstrated that the \textit{a priori} SNR estimate follows the shape of the \textit{a posteriori} SNR with a one frame delay in regions of high \textit{a priori} SNR. Consequently, since the spectral gain depends on the \textit{a priori} SNR, it does not match the current frame and thus the performance of the noise suppression system is degraded especially at transient periods (speech to non-speech or non-speech to speech). To reduce this effect, a Two-Step Noise Reduction algorithm (TS-NR) has been proposed in [7] that refines the estimation of the \textit{a priori} SNR. By suppressing the frame delay bias, the TS-NR algorithm removes some of the drawbacks of the DD approach while maintaining its advantage, i.e. a highly reduced musical noise level. Such algorithm has been selected by ITU-T in 2008 as the optional post-filter in the standardized G.711.1 (multi-rate wide-band extension for the well-known ITU-T G.711) to reduce the lower band quantization noise at the decoder [8].

Recently, a second source of bias in the \textit{a priori} SNR estimate has been identified in [9]. To assess the impact of SNR and gain over- and under-estimation on speech intelligibility, the authors in [9] conducted listening tests with different biased spectral gain function. More precisely, by assuming perfect \textit{a priori} knowledge of the short-time versions of the \textit{a priori} SNR, such a bias was introduced by including a bias in the \textit{a priori} SNR estimation. Listening tests indicated that SNR and gain-function over-estimation errors in frequency bins with negative SNR are particularly harmful to speech intelligibility. Furthermore, the authors suggested that better methods are needed to estimate the spectral SNR from noisy observations, particularly at low input SNR levels. Such methods hold promise for improving speech intelligibility.

In this work, we focus on the problem of over- and under-estimation of the spectral SNR from noisy observations. The proposed correction technique is applied to DD- and TS-SNR Wiener filters but it could also be applied to any recent MMSE based techniques relying on more than the \textit{a priori} and \textit{a posteriori} SNR estimates [10–13]. Starting with the same experiments as those presented in [9], we consider a baseline noise reduction filter which is implemented as a short-time Wiener filter combined with a DD approach for the estimation of the \textit{a priori} SNR. By assuming perfect \textit{a priori} knowledge of the evolution of the short-time SNR, we have analyzed the over- and under-estimation errors of the \textit{a priori} SNR. However, as oppose to [9], we do not introduce a known bias in the \textit{a priori} SNR estimation. Instead we use a well-known technique such as the DD approach for the estimation of the \textit{a priori} SNR and we analyze the SNR bias by comparing this SNR estimate with the perfect \textit{a priori} knowledge of the evolution of the short-time SNR. By processing several noisy speech sentences from a training database and analyzing the SNR bias for all frequency bins within all short-time frames, we thus obtain an empirical estimate of the 2-dimensional distribution of the SNR bias as a function of two main parameters: the \textit{a priori} and the \textit{a posteriori} SNRs. Once this 2-D bias distribution is available, we applied a k-means clustering method (Lloyd’s algorithm) to segment the 2-D plane (\textit{a priori} SNR and \textit{a posteriori} SNR parameters) into m clusters in which each observation (i.e. the bias) belongs to the cluster with the nearest mean. This results into a partitioning of the \textit{a priori-a posteriori} SNR space in m Voronoi cells. The mean estimation error in each cluster is then used as bias correction terms of the \textit{a priori} SNR for all features belonging to that cluster in the 2-D-space. Moreover, it might be worth mentioning that each frequency bin of the estimated SNR is corrected independently according to the estimated bias by the new method that we propose in this paper.

The remainder of this paper is organized as follows. Section 2 presents a brief description of speech enhancement techniques that operate in the frequency domain. Then, we focus our analysis on the considered approach which corresponds to a combination of the DD approach and the two step noise reduction (TS-NR) algorithm. The bias effect in the DD \textit{a priori} SNR estimator is then studied in Section 3. The probability density function of this bias is analyzed in a 2-D space driven by the \textit{a posteriori} and \textit{a priori} SNR parameters. Also in this same Section 3, we deal with the partition of the 2-D plane into m clusters in order to retain the main properties of the bias information. We then proposed in Section 4 a refinement of the noise reduction technique in which the DD and TS-NR \textit{a priori} SNR estimates are modified according to the mean value (i.e. bias) of the error observed in the corresponding cluster. Simulations are carried out in Section 5 to assess the performance of the proposed technique. Objective evaluation in various environmental conditions show that the proposed modification approach is advantageous, particularly for low input SNRs. Excellent noise reduction can be achieved even in the most adverse noise conditions, while avoiding musical residual noise and the attenuation of weak speech components.

2. Frequency domain speech enhancement

In the additive noise model, the clean speech \( s(n) \) is corrupted by an independent noise signal \( d(n) \) with zero-mean. The resulting noisy speech \( y(n) \) is given in the time domain by

\[
y(n) = s(n) + d(n)
\]

where \( n \) is the discrete time index. To obtain the clean speech from the noisy one, conventional noise reduction techniques used a noise reduction filter whose impulse response \( g(n) \) is designed to produce an enhanced speech \( \hat{s}(n) \) corresponding to a trade-off between noise reduction and speech distortion. At the output of this filter, the enhanced speech is expressed as

\[
\hat{s}(n) = g(n) * y(n)
\]

where \( * \) is the convolution operator. The residual noise signal \( e(n) \) is defined as the difference between clean and enhanced speech. Incorporating (1) and (2), the residual noise signal is defined as:

\[
e(n) = \hat{s}(n) - s(n) = (g(n) - \delta(n)) * s(n) + g(n) * d(n),
\]

where \( \delta(n) \) is the Kronecker delta function. Considering that the observed signal \( y(n) \) is divided into overlapping frames by the application of a window function, and analyzed using the short-time Fourier transform (STFT), the time domain residual noise can be equivalently expressed in the frequency domain as:

\[
e(p,k) = \frac{G(p,k)}{G(p,k) - 1} \bigg( S(p,k) + D(p,k) \bigg) e_{SD}(p,k)
\]

where \( k \) represents the frequency bin index, and \( p \) the frame index. \( S(p,k), D(p,k) \) and \( G(p,k) \) represent the kth spectral component of the short-time frame \( p \) of the speech signal \( s(n) \), noise \( d(n) \), and noise reduction filter \( g(n) \), respectively. The two quantities \( e_{S}(p,k) \) and \( e_{D}(p,k) \) represent respectively, the kth spectral components of the short-time frame \( p \) of the speech and the noise distortions amounts. From (4), it can be seen that the spectral gain function \( G(p,k) \) has a significant impact on the speech signal and the noise components energies. This relation (4) does well highlight the trade-off between speech and noise reduction. This is because \( G(p,k) \) affects such speech components and noise equally. The higher the degree of noise suppression, the higher is the amount of speech distortion \( e_{S}(p,k) \). The choice of the distortion measure (which can be modeled by the \textit{a priori} SNR estimation

methods) determines the gain behavior, i.e. the trade-off between noise reduction factor and speech distortion. Such a trade-off is found in the so-called spectral subtraction technique [1] that produces low distortion in the enhanced speech signal at the expense of a high level of musical noise leading to a very poor global quality from a subjective point of view.

However, most noise reduction techniques based on short-time spectral modifications of the noisy speech spectra require the evaluation of two main parameters: the *a priori* and the *a posteriori* SNRs, respectively defined by:

\[ y(p, k) = \frac{|Y(p, k)|^2}{\hat{\lambda}_D(p, k)} \]

and

\[ \xi(p, k) = \frac{|S(p, k)|^2}{\hat{\lambda}_D(p, k)} \]

where \( \hat{\lambda}_D(p, k) = \mathbb{E}[D(p, k)^2] \) stands for the noise power spectral density (PSD), and the notation \( \mathbb{E}[\bullet] \) denotes the expectation operator. As usually assumed when considering spectral modification based noise reduction techniques, the DFT coefficients are considered as independent across time and frequency. In practical implementations of speech enhancement systems, the speech and noise PSDs are unknown since only the noisy speech spectrum \( Y(p, k) \) is available. Thus, both the *a posteriori* and the *a priori* SNRs have to be estimated. The estimation of the noise PSD is beyond the scope of this paper. It can be practically estimated during speech pauses using a conventional recursive estimator [1] or continuously using either the Minimum Statistics [14] technique or the Minima Controlled Recursive Averaging approach [15] to get a more accurate estimate in case of noise level fluctuations. In the following we will assume that such noise PSD is available for each short-time frame and that it will be denoted \( \hat{\lambda}_D(p, k) \). Once a noise power spectral density estimate is available, the spectral gain \( G(p, k) \) is obtained in most noise reduction techniques by a function

\[ G(p, k) = f(\hat{\xi}(p, k), \hat{\gamma}(p, k)) \]

depending on the chosen distortion measure (which can be modeled by the *a priori* SNR estimation). The function \( f(\bullet) \) can be selected among the different spectral gain functions proposed in the literature (e.g. amplitude or power spectral subtraction [1], MMSE STSA [2], MMSE LSA [3], Wiener filtering [4], OM LSA [16]). The resulting speech spectrum is then estimated by applying the spectral gain to the noisy spectrum

\[ \hat{S}(p, k) = G(p, k)Y(p, k). \]

It should be noted that conventional noise reduction techniques hold some shortcomings. It has been shown that the spectral gain \( G(p, k) \) will reach the asymptotic values of 0 or 1 when the SNR is 0 or infinity. However, it is extremely unlikely that the gain will ever reach the value of 1. Therefore, for a given frame, if noise is not present in a frequency component, this component will still be attenuated. This is especially true for speech components that are well below the noise level. They are suppressed along with the noise by conventional noise reduction techniques.

In the rest of the paper, we will focus our analysis on a specific noise reduction scheme combining the “decision directed estimator” (DD) and the two step noise reduction (TS-NR) algorithm. The main properties of these two techniques will be briefly recalled in the following subsections.

### 2.1. Decision-directed *a priori* SNR estimation

Assuming an estimate \( \hat{\lambda}_D(p, k) \) of the noise PSD is available at each short-time frame, we define the *a posteriori* SNR estimate as follows:

\[
\hat{\gamma}(p, k) = \frac{|Y(p, k)|^2}{\hat{\lambda}_D(p, k)}.
\]

This relation to corresponds to a local SNR computed from the data in the current short-time frame. It can be noted that given the noise variance, the *a posteriori* SNR is not estimated but known at each frame. The so-called *a priori* SNR estimator represents the information on the unknown spectrum magnitude gathered from previous frames and is estimated in the “decision-directed” approach [2] by

\[
\hat{\xi}^{(DD)}(p, k) = \max\left\{ \beta \frac{|\hat{S}(p-1, k)|^2}{\hat{\lambda}_D(p, k)} + (1 - \beta)P\left(\hat{\gamma}(p, k) - 1\right), \xi_{\text{min}} \right\}.
\]

As \( \hat{\gamma}(p, k) \) defined by (9), is not necessarily greater than one, the operator \( P[\bullet] \), in (10), is a function which guarantees that \( \hat{\xi}^{(DD)}(p, k) \) is always non-negative. In the first term of (10), \( \hat{S}(p-1, k) \) corresponds to the noiseless signal spectrum value as estimated in the previous frame from relation (8). The *a priori* SNR estimator (10) represents the information on the unknown spectrum magnitude gathered from previous frames. \( \xi_{\text{min}} \) is a threshold that limits the *a priori* SNR of (10) to \(-20\, \text{dB}\). Without loss of generality, in the following, the spectral gain function \( f(\bullet) \) in (7) is selected as the Wiener filter, and then

\[
G^{(DD)}(p, k) = \frac{\hat{\xi}^{(DD)}(p, k)}{1 + \hat{\xi}^{(DD)}(p, k)}.
\]

The approach defined by (9), (10) and (11) is called the DD estimator algorithm. It has been demonstrated in [5] that the DD estimator has interesting properties (the musical noise phenomenon is greatly reduced) but suffers from a one-frame delay. This delay inherent to the DD algorithm is a drawback especially in the speech transients, e.g., speech onsets and offsets. It introduces a bias in gain estimation which limits noise reduction performance and generates either speech distortions (for onsets) or an annoying reverberation effect (for offsets) because the level of the enhanced speech in the previous frame is artificially maintained in the current one due to this one-frame delay.

### 2.2. Two-step noise reduction technique

To enhance the noise reduction process, it is proposed in [5] to refine the *a priori* SNR estimate in a two-step procedure. This technique is based on the observation that the DD algorithm introduces a one frame delay when the parameter \( \beta \) is close to one. Consequently, the spectral gain computed at current frame \( p \) matches the previous frame \( p - 1 \). Based on this fact, it is proposed in [7] to compute the spectral gain for the next frame \( p + 1 \) using the DD approach and to apply it to the current frame because of the frame delay. This leads to a two step procedure algorithm. The first step consists exactly of the decision directed algorithm and a spectral gain \( G^{(DD)}(p, k) \) is computed as described in (11). In the second step, this gain is re-used to refine the estimate of the *a priori* SNR as:

\[
\hat{\xi}(p, k) = \frac{|G^{(DD)}(p, k)Y(p, k)|^2}{\hat{\lambda}_D(p, k)}.
\]

We can notice that the TS-NR approach uses the DD estimator as an intermediate step to compute the output spectral gain.
(12) that is used in the refinement a priori SNR estimate. Finally, another spectral gain is computed:

\[ G^{(TS-NR)}(p, k) = h(\hat{\xi}(p, k), \hat{\gamma}(p, k)), \quad (13) \]

which is used to enhance the noisy speech

\[ \hat{S}(p, k) = G^{(TS-NR)}(p, k)Y(p, k). \quad (14) \]

By taking the inverse Fourier transform of (14), the enhanced speech signal is then reconstructed in the time-domain signal using the overlap-add method. Note that function \( h(\bullet, \bullet) \) may be different from the function \( f(\bullet) \) defined in (7). However, without loss of generality, in the following, this gain will be selected as a Wiener filter, and then

\[ G^{(TS-NR)}(p, k) = \frac{\hat{\xi}(p, k)}{1 + \hat{\xi}(p, k)}. \quad (15) \]

This algorithm in two steps defined by (9), (10), (12), and (15), is called the TS-NR algorithm. This technique improves the noise reduction performance since the gain matches to the current frame whatever the SNR. The main advantages of this approach are the ability to preserve speech onsets and offsets, and to successfully remove the annoying reverberation effect typical of the DD approach [2,3,5,16,19]. Note that in practice this reverberation effect can be reduced by increasing the overlap between successive frames but cannot be suppressed whereas the TS-NR approach [7] makes it possible with a typical overlap of 50%. Such solution has been selected by ITU-T [8] in 2008 as the optional non-parametric post-filter in the standardized G.711.

### 3. Properties of SNR estimators

In this section, we concentrate our analysis on several important properties of the a priori SNR estimate as this parameter plays a central role in spectral noise reduction techniques. As described previously, estimating the a priori SNR through the “decision-directed” approach is widely used (instead of the a posteriori SNR) because it reduces the musical noise to an acceptable level. However, as we shall see, this a priori SNR estimate often leads to an under- or over-estimation of the SNR on each frequency component thus reducing the output speech quality. From a subjective point of view, this SNR’s over- or under-estimation is at similar to the same effect as in the onset/offset distortions that are created by the one-frame delayed SNR estimation. It should be noted that the a posteriori SNR is a known quantity (that is, given the noise statistics), and it appears to be simply used in this work as an indicator on how to adjust the a priori SNR estimate - which is unknown.

#### 3.1. A priori and a posteriori SNR estimation

To better understand the impact of errors in the estimation of these SNRs, we show in Figs. 1 and 2 scatter plots of the a posteriori and the estimated a priori SNRs versus the ‘true’ short-time values of the SNR defined as:

\[ \xi(p, k) = \frac{|S(p, k)|^2}{\lambda_d(p, k)}. \quad (16) \]

To provide this parameter available for the analysis, we assumed knowledge of the magnitude spectrum of the clean speech signal. It should be stated that the true ideal SNR is unobtainable due to the non-stationarity of the speech signal. In this paper, due to the unavailability of this true SNR, we define the ground truth as given by (16), which is appropriate because any noise reduction filter that uses (16) will provide an enhanced speech signal with high noise reduction level and minimum spectral distortions. As a result, let us consider that the a posteriori SNR (Fig. 1) is obtained from (9), and the a priori SNR estimate (Fig. 2) is obtained by relations (10), (11) and (12) using the TS-NR approach (with \( \beta = 0.98 \)). In these figures, for each frame and each frequency, we have represented features (or 2-tuples) considered as a point in the two-dimensional space defined by the two SNRs, namely the SNR (estimated a priori or known a posteriori) and ‘true’ SNR defined by (16) which is considered as a reference. Moreover, to focus the analysis on the behavior of the SNR estimators for speech components, only features (about 900 x 257 features) corresponding to speech activity periods are represented on this figure. In this experiment, the frame length is \( L = 256 \) for a sampling frequency of 8 kHz. Each of the noisy signal frames are Hanning windowed with a 50% overlap and transformed in the frequency domain with a 2L-point FFT to avoid distortions resulting from the circular convolution effect [27]. The overall input SNR is 13 dB and the background noise is babble. To evaluate the performance of each estimator, the bold line corresponds to a perfect SNR esti-
mator (i.e. zero estimation error) that can be used as a reference. The upper triangular part of this figure corresponds to an over-
estimation of the true SNR (respectively under-estimation in the
lower-triangular part).

From Figs. 1 and 2, it is obvious that the features corresponding
to the \textit{a posteriori} SNR parameter are closer to the reference
bold line and are less dispersed than the \textit{a priori} SNR estimator
ones. This is especially true for high signal-to-noise ratio values.
Furthermore, to provide mean behavior analysis, we have also rep-
presented on Fig. 1 and Fig. 2 the conditional expectations of the
two estimated SNRs given the knowledge of the true SNR \( \xi(p,k) \)
which are obtained as follows:

\[
E(\hat{\xi}|\xi) = \int_{\mathbb{R}} \hat{\xi} f_{\hat{\xi}|\xi}(\hat{\xi}|\xi)d\hat{\xi}
\]
\[
E(\hat{\gamma}|\xi) = \int_{\mathbb{R}} \hat{\gamma} f_{\hat{\gamma}|\xi}(\hat{\gamma}|\xi)d\hat{\gamma}
\]

where \( f_{\hat{\xi}|\xi} \) and \( f_{\hat{\gamma}|\xi} \) are the conditional probability density func-
tions of the estimated SNR knowing the true SNR. We can observe
that the conditional mean of the estimated SNR is closer to the
perfect estimator for the \textit{a posteriori} SNR parameter than for the
\textit{a priori} SNR estimator. The available \textit{a posteriori} SNR values can
be viewed as slightly underestimated for high SNR whereas for the
\textit{a priori} SNR the underestimation is large for SNR greater than
\(-17\) dB. However, since the dispersion is high for the \textit{a priori}
SNR features, even if the mean is largely underestimated, the case
where SNR features are overestimated exists. Furthermore, for SNR
smaller than \(-17\) dB, the \textit{a priori} SNR is overestimated.

Let us now examine the behavior of the bias in the \textit{a priori}
SNR estimate as a function of the value of the parameter \( \beta \) in
the decision-directed loop (10). To do so, Fig. 3 compares for speech
activity periods the conditional mean (17) of the \textit{a priori} SNR esti-
mator for \( \beta \) ranging from 0.3 to 0.98.

For frequency bins with low SNR values, we see that the bias
increases with decreasing \( \beta \). The \( \hat{\xi} \) value is over-estimated by as
much as \( 20-40 \) dB at extremely low (<40 dB) SNR levels. A value
of \( \beta \) near one provides the highest noise reduction, while avoiding
the musical noise. However, it comes at the expense of a reduction
in intelligibility, because important speech transitions are exces-
sively smoothed [6]. On the other end, as \( \beta \) tends to zero, the
behavior of the conditional mean curve of the \textit{a priori} SNR estimate
tends to mimic the one of the \textit{a posteriori} SNR presented in Fig. 1
but this is obtained at the expense of a musical noise which is in-
trduced in the enhanced speech. On the contrary, for frequency
bins with high SNR (greater than 0 dB) the bias in the \textit{a priori} SNR
estimate is decreasing with increasing values of \( \beta \). This bias analy-
sis provides a more comprehensive understanding of the important
role of the \textit{a priori} SNR estimate in noise reduction problems and
is in agreement with other experimental observations [5,17].

The results confirm that the \textit{a posteriori} SNR parameter is more
reliable than the \textit{a priori} SNR estimator for high SNR frequency bins
(i.e. speech components) but this behavior is obtained at the ex-
 pense of the insertion of the musical noise in the enhanced speech
signal. From these curves, we can predict that the estimation of
the spectral gain function (15) will be particularly affected by the
perturbations in the \textit{a priori} SNR values especially in the negative
region of the SNR as inaccuracies in the estimation of \textit{a priori} SNR
values (such as under-estimation) directly cause a bias in the gain
function (a lower value), thus generating distortion in the output
speech.

3.2. Analysis of the probability density function

To go further insight the analysis of the role of \textit{a priori} SNR es-
timation within the TS-NR algorithm, we now concentrate on the
probability density function of this SNR estimate through its cor-
responding empirical 2-D histogram which is obtained by counting
in Fig. 2 the number of features (or 2-tuples) that fall within each
3 dB × 3 dB area. The obtained normalized histograms are repre-
sented in Fig. 4 as a function of the \textit{a priori} SNR estimate and true
SNR values.

One can see in Fig. 4 that the 2-D-histogram of the \textit{a priori} SNR
estimate exhibits two distinct kernels or clusters (denoted by the
letters A and B) that are defined as maximal sets of connected
dense units in the 2-D plane.

In this sense, the empirical histogram corresponds to a mix-
ture density that can be interpreted as being derived in a sim-
ple way from an underlying set of other random variables. The
connection between these clusters and the inherent properties
of the corresponding time-frequency components \( Y(p,k) \) is made
clearer by plotting in Fig. 5 the spectrogram corresponding to the
same speech segment as the one considered in Fig. 4. On this
figure, all frequency components corresponding to cluster B (i.e.
\(-30 \) dB < \( \hat{\xi}(p,k) < -10 \) dB) are represented in ‘red’. Note from
(15) that such low values of the \textit{a priori} signal-to-noise ratio may
result in a mean value of \(-20 \) dB for the short-time spectral

\[
G^{TS-NR}(p,k) \]

Thus, we can assume that the corresponding frequency bins will be suppressed within the enhanced speech
spectrum. From Fig. 4, we can consider that the remaining fea-
tures belong to cluster A. However, we can conclude from Fig. 5
that the features corresponding to kernel A are mainly related to
speech harmonics (rather than noise components) that should be
preserved by the spectral noise reduction filter. From the previous
analysis of Figs. 4 and 5, we can expect that the performance of the
amplitude estimator (14) will greatly depend on the inherent prop-
erties of the \textit{a priori} SNR estimator used. In the TS-NR approach
different behaviors may appear when we estimate the \textit{a priori}
SNR from the nonlinear recursive relation (10) in combination with
(11) and (12). When \( \hat{\gamma}(p,k) \) stays below 0 dB, the \textit{a priori} SNR
corresponds to a highly smoothed version of the \textit{a posteriori} SNR
over successive short-time frames. The spectral gain \( G^{TS-NR}(p,k) \)
in such conditions is lower than \(-24 \) dB (as seen for cluster B) and
does not exhibit large variations over successive short-time
frames. As a consequence, the musical noise phenomenon is re-
duced. On the contrary, when \( \hat{\gamma}(p,k) \) is much larger than 0 dB, the
\textit{a priori} SNR is no longer a smoothed SNR estimate. Its short-time
evolution follows the one of the \textit{a posteriori} SNR (cluster A). Such

![Fig. 3](image-url)

Fig. 3. Conditional mean of the \textit{a priori} SNR estimator \( \hat{\xi} \) for different values of pa-
rameter \( \beta \) in the TS-NR approach. Speech activity periods. Input global SNR 13 dB
and babble noise.

This is in agreement with other experimental observations [5,17].

Role of the \textit{a posteriori} SNR parameter provides a more comprehen-
sive understanding of the important role of the \textit{a priori} SNR estimate in
noise reduction problems and is in agreement with other experimental observations [5,17].
studies propose new approaches for the proposal of alternative analytical suppression rules \cite{20,21}. Other short-time spectral gain at data-driven approach \cite{18,19}, or to modify noise reduction techniques. Some of them have tried to modify the short-time spectral gain in a data-driven approach \cite{18,19}, or to propose alternative analytical suppression rules \cite{20,21}. Other studies propose new approaches for the a priori SNR estimate such as robust estimators \cite{22,23}, or noncausal estimators \cite{24}. However, as seen previously, the a priori SNR estimation represents a crucial parameter and suffers from over- or under-estimation errors thus leading to a local sub-optimal solution for the noise reduction filter. In this section, we aim to design a more accurate a priori SNR estimator that solves the drawbacks of the conventional estimator in order to reduce speech distortion while retaining a high level of noise attenuation in speech absence. In this section, we describe our strategy that relies on the optimization of the mean square error (MSE) in the a priori SNR estimate.

4. Bias compensation of the a priori SNR estimator

In this section, we address the problem of improving the SNR estimation. Several attempts have been made in the past to modify noise reduction techniques. Some of them have tried to modify the short-time spectral gain in a data-driven approach \cite{18,19}, or to propose alternative analytical suppression rules \cite{20,21}. Other studies propose new approaches for the a priori SNR estimate such as robust estimators \cite{22,23}, or noncausal estimators \cite{24}. However, as seen previously, the a priori SNR estimation represents a crucial parameter and suffers from over- or under-estimation errors thus leading to a local sub-optimal solution for the noise reduction filter. In this section, we aim to design a more accurate a priori SNR estimator that solves the drawbacks of the conventional estimator in order to reduce speech distortion while retaining a high level of noise attenuation in speech absence. In this section, we describe our strategy that relies on the optimization of the mean square error (MSE) in the a priori SNR estimate.

4.1. Proposed technique for bias compensation

As the two contributing SNR components of the conventional decision-directed approach are two direct observations of the true SNR $\xi(p, k)$, we can symbolically write

$$\xi(p, k) = T\{\hat{\xi}(p, k), \gamma(p, k)\}.$$  \hspace{1cm} (18)

where the analytical expression of function $T(\bullet)$ is unknown. However, its main properties can be recovered from the analysis of the correspondence between the experimental time-frequency series $z(p, k) = (\hat{\xi}(p, k), \gamma(p, k))$ and the true a priori SNR $\xi(p, k)$. Assuming a priori knowledge of the short-time versions of $\hat{\xi}(p, k)$ in a training phase, we can thus consider in the rest of the analysis that function $T(\bullet)$ is known from an experimental point of view. To investigate the impact of the estimation errors, we consider the error in the a priori SNR estimate:

$$e(p, k) = \xi(p, k) - \hat{\xi}(p, k) = T(z(p, k)) - \hat{\xi}(p, k) = H(z(p, k)).$$  \hspace{1cm} (19)

which is expressed as a function $H(\bullet)$ of the available a posteriori and the estimated a priori SNRs parameters. For the same reason as previously discussed for $T(\bullet)$, the function $H(\bullet)$ can be assumed known from an experimental point of view. The optimization process can be realized by minimizing the mean-squared error (MSE), defined as $J = \mathbb{E}[e^2(p, k)] = \iint d\sigma$ where $d\sigma$ stands for the infinitesimal MSE defined as:

$$d\sigma = c^2(z)f_Z(z)d\sigma ,$$  \hspace{1cm} (20)

where $d\sigma$ is the area differential and $f_Z(z)$ denotes a probability density function (that will be estimated by its corresponding normalized 2-D-histogram). By integrating locally the infinitesimal MSE (20) on small surfaces (of $3 \text{ dB} \times 3 \text{ dB}$), we obtain in Fig. 6 the behavior of the parameter $d\sigma$ in the $z = (\hat{\xi}, \gamma)$ plane. Notice, that the infinitesimal MSE is represented in decibel in this figure. We can see that the random variable $e(p, k)$ is dominated by two main behaviors: over-estimation errors are introduced for negative values of $d\sigma$ (expressed in dB) in spectral components with negative a priori and a posteriori SNR. In contrast, for spectral components with $\hat{\xi}(p, k) > 10 \text{ dB}$, the decision-directed and/or TS-NR approaches cause over-estimation errors (positive values in dB).
These results are in agreement with the conditional mean analysis presented previously in Fig. 1. On the high-left and low-right corners in Fig. 6, we note that some combinations of the SNRs are highly unlikely and may not often enough have occurred during the training (for clarity of the representation, these points have been assigned a 0 dB value in this figure).

To describe our strategy, our goal is to propose a new a priori estimator \( \hat{\xi}(p, k) = \xi(p, k) - \Delta(p, k) \) which reduces over- and under-estimation errors \( J = E[\epsilon^2(p, k)] \) where \( \epsilon(p, k) = \xi(p, k) - \hat{\xi}(p, k) \). The correction term \( \Delta(p, k) \) can be viewed as a bias compensation mechanism for the conventional a priori SNR estimate \( \xi(p, k) \). However, for each frequency bin within each frame, deriving an analytical optimal expression for the compensation factor \( \Delta(p, k) \) is difficult. As pointed in [21], correcting fully for the bias is difficult because of the non-linear feedback loop in the decision-directed approach. Instead, our strategy consists in searching in the \( z = (\xi, \hat{\gamma}) \) plane for the centroidal Voronoi tessellation \( (\Omega^p_{\xi})_{m=1}^{\infty} \) (with generators \( z_i \)) and bias compensation factors \( \Delta_i \) that constitute the set \( (\{\Delta_1, \xi_i\}) \in \mathbb{R} \times \mathbb{R}^2, i = 1, \ldots, m \) such that

\[
J' = \sum_{i=1}^{m} \int_{\Omega^p_i} \epsilon^2(z) f_z(z) \, d\sigma \\
= \sum_{i=1}^{m} \int_{\Omega^p_i} (\epsilon(z) + \Delta_i)^2 f_z(z) \, d\sigma 
\tag{21}
\]

is minimized.

A sub-optimal solution to this problem can be obtained by using an iterative procedure such as the Lloyd–Max algorithm [25]. Starting, at time \( t = 0 \), with an initial set of generators \( z_i^{(0)} \), \( i = 1, \ldots, m \), we iterate the following steps:

**Step 1:** Construct the Voronoi constellation \( \{\Omega^p_i\}_{i=1, \ldots, m} \) from previous generators \( z_i^{(t-1)} \), \( i = 1, \ldots, m \).

\[
\Omega^p_i = \{ z(p, k) \in \mathbb{R}^2 \mid \| z - z_i^{(j)} \| < \| z - z_i^{(j')} \|, j \neq i \} 
\tag{22}
\]

**Step 2:** Compute the new bias compensation factors such that \( \partial J / \partial \Delta_i = 0 \). The optimal solution is obtained as the mean value of the error \( \epsilon(p, k) \):

\[
\Delta_i(t) = E[\epsilon(z) \mid z(p, k) \in \Omega^p_i] 
\tag{23}
\]

**Step 3:** Take the mass centroids of \( \{\Omega^p_i\}_{i=1, \ldots, m} \) as the new set of generators:

\[
z_i(t + 1) = E[z(p, k) \mid z(p, k) \in \Omega^p_i] \\
= \int_{\Omega^p_i} z f_z(z) \, d\sigma / \int_{\Omega^p_i} f_z(z) \, d\sigma. 
\tag{24}
\]

This procedure is continued until global convergence is reached, i.e. until some stopping test criterion is fulfilled. In this work, the algorithm is stopped when further iteration no longer produce any changes in the mean-squared error \( J = E[\epsilon^2(p, k)] \) or when changes are below a suitable threshold fixed by the user. In this case, an extra-computation time can be observed as the run time is depending on the selected number of clusters. As for all Lloyd–Max algorithms, the initialization of generator’s values plays an important role for both the convergence speed and the asymptotic precision and many solutions have been proposed in the literature to initialize the Lloyd–Max’s algorithm. As our main interest in this work is not related to the optimization of the convergence speed of the training phase, we used a simple uniform 2-D-rectangular grid (from \(-30 \text{ to } +30 \text{ dB} \) on each dimension, \( \xi \) and \( \hat{\gamma} \)) for the initial set of generators at time \( t = 0 \).

It should be noted that the considered approach is expected to improve the MMSE estimates as far as the SNR values that have been used as reference informations are perceptually better than the estimated values. In addition, as there is a bias between the estimated and the reference SNRs, we attempt to correct for this bias so that we get closer to these desirable reference values, whether or not they are the true/optimal values.

### 4.2. Including uncertainty of speech observations

In the previous sections (even in Section 4.2), we have studied the estimation inaccuracies in the a priori SNR by assuming that speech is almost always present in the observation signal. As a consequence, all the previous figures were obtained from noisy speech samples corresponding to speech activity periods (we used a handmade ideal speech detector). However, in reality, speech pauses are always present in the observation signal and this uncertainty should be taken into account in the proposed bias compensation approach. Given two hypotheses, \( H_0(p, k) \) and \( H_1(p, k) \) which indicate respectively speech absence and presence in the \( k \)th frequency bin of the \( p \)th frame, we have:

\[
\begin{align*}
H_0(p, k) : Y(p, k) &= D(p, k) \\
H_1(p, k) : Y(p, k) &= S(p, k) + D(p, k)
\end{align*}
\]

The problem we face now is how to incorporate this information into the bias-corrected noise reduction algorithm that has been described in the previous Section 4.1. This problem can be easily solved by considering two different sets of parameters \( \{z_i, \Delta_i_{H_0} \}_{i=1, \ldots, m} \) and \( \{z_i, \Delta_i_{H_1} \}_{i=1, \ldots, m} \) during the training procedure, each being optimized during speech absence (\( H_0 \)) or presence (\( H_1 \)) in the observed frequency bins. A pseudocode program of the complete training algorithm including uncertainty of speech observation is shown in Algorithm 1. In this algorithm, we use the following definition:

\[
\bar{x}(t) = \frac{1}{|A|} \sum_{t \in A} x(t)
\]

to express the arithmetic mean operating on a given set of indices (\( |A| \) denotes the cardinality of \( A \)).

### 4.3. Learning phase: methods and procedure

As previously described, a learning procedure is proposed to estimate the bias compensation parameters \( \{z_i, \Delta_i\}_{i=1, \ldots, m} \). For training, we used a speech material which consists of 30 phonetically-balanced sentences recorded for the NOIZEUS database and freely
available from the Website of the University of Texas at Dallas. These sentences (15 males and 15 females) were originally sampled at 25 kHz and downsampled to 8 kHz. Noisy speech files were generated by artificially adding noise to the original clean speech. Noticing that in the previous Sections 4.1 and 4.2, the time-varying Fourier coefficients of each process (noise, clean and noisy speech) were considered as (zero-mean) statistically independent random variables (as this is usually the case for a majority of speech-enhancement algorithms operating in the frequency domain), the previous assumptions are actually equivalent to the assumption that the Fourier coefficients are uncorrelated (in practice they will be correlated to a certain degree). Consequently, in the learning phase, only stationary white noise was added to the clean speech data to generate noisy speech at overall SNRs ranging from -15 dB to +25 dB, in steps of 5 dB.

Algorithm 1 Training phase.

Initialization: \( i := 0, p = 0 \). Initialize \( \xi^{H_0}(0) \) and \( z^{H_0}(0) \), \( i = 1, \ldots, m \).

Main loop. For each frame \( p \) within the training database do:

Step 1: compute the N-point FFT of the noisy speech, noise and clean speech as \( \{Y(p,k)\}_{k=1}^{N/2}, \{N(p,k)\}_{k=1}^{N/2} \) and \( \{S(p,k)\}_{k=1}^{N/2} \).

Step 2: compute the ‘true’ SNR from (16), \( \xi(p,k) \). Compute the a priori SNR estimate \( \hat{\xi}(p,k) \) using relation (9). Compute the a priori SNR estimate \( \hat{\xi}(p,k) \) using both (10) and (11) for the DD approach (Proposition I) or using (10), (11) and (12) for the TS-NR approach (Proposition II). Form the pair \( z = (\hat{\xi}(p,k), \hat{\gamma}(p,k)) \).

Step 3: For each frequency index \( k \), if \( H_0(p,k) = \text{‘true’} \) then

- form the set: \( \mathcal{H}_0 = \{ (p,k) \in \{0, \ldots, t \} \times \{1, \ldots, 1 + \frac{N}{2} \} / H_0(p,k) = \text{‘true’} \} \)
- update from (22) the Voronoi constellation \( \Omega_i^{H_0}(t) \)

- form the set: \( \Omega_i^{H_0}(t) = \{ z(p,k) / i = \arg \min_{j=1,\ldots,m} |z(p,k) - z^{H_0}_j| \} / \Omega_i^{H_0}(t) \)
- update the bias compensation factors:

Step 4: For each frequency index \( k \), if \( H_0(p,k) = \text{‘true’} \) then

- obtain the mass centroids by using:

\[ z_i^{H_0}(t+1) = 2z(p,k) / |z(p,k) \in \mathcal{H}_0,p,k| \]

Step 5: same as Step 4 but when replacing \( H_0 \) by \( H_1 \).

iterate loop \( t \leftarrow t+1 \)

Once the training database is available, the previously described training procedure consisting of the combination of (22), (23) and (24) is run on the entire noisy speech database in order to estimate the parameters \( \{z_i, \Delta_i\}_{i=1,\ldots,m} \). The block diagram of this procedure is depicted in Fig. 7.

An illustration of the resulting Voronoi partition available at the end of training phase is depicted in Fig. 6 for \( m = 100 \) clusters. To obtain this figure, we used the following experimental parameters: a Hamming window with a 50% overlap for spectral analysis, and a rectangular window for synthesis for the DFT-based noise reduction algorithm. One notes in Fig. 6 that the previous training procedure generates dense areas of cells that occur when the a posteriori SNR approaches 0 dB.

4.4. Proposed bias-compensated noise reduction approach

Once the parameters \( \{z_i, \Delta_i\}_{i=1,\ldots,m} \) and \( \{z_i, \Delta_i\}_{i=1,\ldots,m}^{H_1} \) are known at the end of the ‘training’ phase, our aimed is to propose a bias-compensated noise reduction approach that can be implemented as an enhancement in the DD noise reduction (Proposition I) or in the TS-NR noise reduction algorithm (Proposition II). As shown in the top-level block diagram of Fig. 8, the proposed system is built on a two-stream procedure [18].

It consists in running two algorithms in parallel. From relation (10), (9) or (12), a conventional DD (Proposition I) or TS-NR (Proposition II) algorithm is first run in the background to obtain a pair of estimated SNRs \( z = (\hat{\xi}(p,k), \hat{\gamma}(p,k)) \) in each frame and in each frequency bin. Then, a second a priori estimate \( \hat{\xi}(p,k) = \hat{\xi}(p,k) - \Delta(p,k) \) is computed leading to a bias-corrected SNR estimate. For each frequency bin \( k \) in each short-time frame \( p \), the value of the bias compensation factor \( \Delta(p,k) \) is selected according to the current value \( z(p,k) \) of the estimated SNRs in the z-plane but also on the absence/presence of speech in the considered frequency bin. A pseudocode for the computation of \( \Delta(p,k) \) is given by.

If \( H_0(p,k) = \text{‘true’} \) then

\[ i = \arg \min_{j=1,\ldots,m} |z(p,k) - z_j^{H_0}| \quad \text{and} \quad \Delta(p,k) = \Delta_i^{H_0} \]

Else

\[ i = \arg \min_{j=1,\ldots,m} |z(p,k) - z_j^{H_1}| \quad \text{and} \quad \Delta(p,k) = \Delta_i^{H_1} \]

End

Once the bias correction factor \( \Delta(p,k) \) is known, the final spectral noise reduction filter and the output speech reconstruction is another stream that uses \( \hat{\xi}(p,k) \) in (15) in place of \( \hat{\xi}(p,k) \). This procedure guarantees an improvement in terms of the J’ error criterion (21), because the first stream which performs the DD or...
TS-NR a priori SNR estimate is left untouched (free-running algorithm). Without loss of generality, we present hereafter results obtained for the Wiener spectral gain. A pseudocode program of the complete testing algorithm including uncertainty of speech observation is described in Algorithm 2. All computations are embedded into loops over all frequency indices \( k \) and all frame indices \( p \). During our first quality test experiments under various noise conditions, we note that, in non-stationary babble noise, the SNR estimator with bias compensation factors (23) was sensitive to noise bursts that are not tracked by the noise power estimation algorithm. To obtain full potential of the proposed method, we replace (23) by the following rule which tracks the minimum value of the error estimate within each cluster:

\[
\Delta_i(t) = \min_{z \in \mathcal{D}_i^{(0)}} H(z). \quad (25)
\]

Using these modified compensation factors, informal listening tests demonstrate that the enhanced signals had much less speech distortion. It should also be noted that the specific properties of the spectral gain are automatically taken into account during the ‘training/learning’ phase of the proposed bias-compensation approach (see Section 4.3). The degree of compensation of the a-priori SNR bias is indeed adjusted according to the experimental differences observed between the true a-priori SNR values and the a-priori SNR estimates obtained by the considered speech enhancement estimator. In that sense, the proposed bias-compensation mechanism can be considered as a global optimization method that takes into account the intrinsic properties of the speech enhancement estimator, i.e., the gain function (Wiener gain, ... etc).

**Algorithm 2 Testing phase.**

Initialization: \( \ell = 0, p = 0 \). Initialize \( \hat{z}^{(0)}(p) \) and \( \hat{z}^{(1)}(p) \), \( i = 1, \ldots, m \).

Main loop. For each frame \( p \) within the testing database do:

Step 1: Compute the FFT of the noisy speech \( Y(p,k) \) with \( k=1,\ldots,N/2 \).

Step 2: Compute the a posteriori SNR \( \hat{\gamma}(p,k) \) using relation (9). Free-running estimation of the a priori SNR \( \hat{\xi}(p,k) \) using both (10) and (11) for the DD approach (Proposition I) or using (10), (11) and (12) for the TS-NR approach (Proposition II).

Form the pair \( z = (\hat{z}(p,k), \hat{\gamma}(p,k)) \).

Step 3: If the test hypothesis \( H_0(p,k) = \text{true} \) then

- for each \( z(p,k) \), obtain the Voronoi cell \( \mathcal{D}_i \) by using:

\[
i = \arg \min_{j=1,\ldots,\text{m}} |z(p,k) - \hat{z}^{(j)}(p,k)|
\]

- then, apply the corresponding bias correction factor:

\[
\hat{\xi}'(p,k) = \hat{\xi}(p,k) - \Delta(p,k) \quad \text{with} \quad \Delta(p,k) = \Delta_{i}(\hat{\gamma})
\]

Step 4: same as Step 3 but when replacing \( H_0 \) by \( H_1 \).

Step 5: compute and apply the corresponding Wiener filter

\[
G(p,k) = \frac{\hat{\xi}'(p,k)}{1 + \hat{\xi}'(p,k)}
\]

and reconstruct the output by applying overlap-add technique. 

iterative loop \( \ell \leftarrow \ell + 1 \)

5. Experimental results

In this section, the performance of the two proposed approaches is compared to that of the ‘conventional’ DD as well as TS-NR approaches. The bias-compensated approach is applied in two configurations, namely the ‘conventional’ DD and the TS-NR a priori SNR estimators previously described in Section 2. These two techniques will be hereafter called Propositions I and II corresponding to the combination of the bias-compensated approach together with respectively (10) or (12). For testing, the speech material consists of a different set of speech sentences in comparison with the learning database. We have used different noise/speech samples in the learning/testing parts to show that the proposed method can work even in the worst case when different databases are used in the learning (Section 4.3) and testing phases. Doing so, general conclusions could be drawn in this section (without bias due to a specific choice of the learning/validation databases).

For testing, the sentences were taken from the AURORA database [26] (different from the training one). The performance of the two proposed approaches (Propositions I and II) is evaluated in different noise environments with noise signals (white, babble, street and F16-aircraft noise) taken from the Signal Processing Information Base (SPIB) and freely available on the Rice University homepage. These noise files were originally sampled at 19.98 kHz and recorded from 1/2" B&K condenser microphone onto digital audio tape (DAT). Combining the clean speech and the noise test signals, noisy speech utterances in each environment are obtained for testing. The noisy speech signals were sampled at 8 kHz and all the test sets are prepared for 0, 5, 10, 15, 20 dB SNR conditions, respectively. In this section, the parameters of the noise reduction system are kept unchanged (256 samples per frame, overlap between frames of 50%, Hanning windowing, FFT length of 512 points and overlap-add technique for reconstruction). For all experiments, the decision-directed a priori SNR estimate (10) is computed with a parameter \( \beta \) set to 0.98.

We recall here that the training procedure that it will be used to correct the conventional DD and TS-NR algorithms (Proposed I and Proposed II) corresponds to the different steps that have been previously described in Section 4.3 (Also, see block diagram in Fig. 7 for more understanding of the training phase).

5.1. Influence of the number of cells

To highlight the influence of the number of cells on the behavior of the proposed noise reduction approach, we show in Figs. 9 and 10 scatter plots of the proposed a priori SNR estimate vs. true a priori SNR values for two different values of the global input SNR, respectively 0 dB and 10 dB.

For the sake of brevity, results are given for the bias compensated TS-NR approach (Proposition II) but similar results have been observed when combining the proposed bias compensation approach with the ‘conventional’ DD a priori estimator (Proposition I). On these figures, the gray points represent the features \( u = (\xi', \hat{\gamma}) \) corresponding to 850 frames of speech activity periods. Only frames corresponding to speech activity periods are represented. As discussed earlier, the proposed bias-compensated a priori SNR given by \( \xi'(p,k) = \hat{\xi}(p,k) - \Delta_i \) is obtained by introducing the two sets of compensation factors \( \{\xi_i, \Delta_i^{(H)}\}_{i=1,\ldots,m} \) into the ‘conventional’ a priori SNR estimators (DD or TS-NR) as previously described in Sections 2.1 and 2.2 respectively, (for the training phase, refer to Section 4.2).

The diagonal bold solid line corresponds to a perfect SNR estimator that can be used as a reference and the thin lines represent the conditional mean of the proposed bias-compensated a priori SNR estimates obtained through (17). These curves are given for different number of cells, i.e. 5, 10, 50, 150, 350 and 500 cells.

We may note in Fig. 9 that increasing the number of cells in the \( (\xi', \hat{\gamma}) \) plane provides a significant improvement of the a priori SNR estimation. By comparing Figs. 9 and 10 with the conditional mean obtained previously with the “conventional” TS-NR method (see Fig. 2), it is clear that increasing the number of cells drastically reduces the bias in the a priori SNR estimate. For example, in the case of 150 cells, one notes that the proposed a priori SNR estimate is getting closer to the perfect SNR estimator (bold line). However, the estimated values are slightly over-estimated for true SNR values in the range \(-40 \text{ dB} \) to \(-12 \text{ dB} \), whereas they are
slightly under-estimated for SNR values greater than 20 dB. Referring to Figs. 9 and 10, we see that this behavior still holds for lower values of the input global SNR. It is clear that the obtained conditional mean behaviors of the \textit{a priori} SNR are closer to the perfect estimator line as the number of cells is large (i.e. case of 150–500 classes). In the following, the total number of cells will be set to \( m = 150 \) as this value realize an efficient trade-off between performance (i.e. low bias) and complexity.

5.2. Comparisons with conventional \textit{a priori} SNR estimates

We now compare the proposed bias-compensated \textit{a priori} SNR estimates (Propositions I and II) computed as described in Section 4 with conventional approaches such as the decision-directed (DD) [2] and TS-NR techniques [7]. We first consider in Fig. 11 an experiment with a USASI (United States of America Standards Institute) noise that has been added to the speech signal at a SNR of 0 dB. On this figure, we have plotted the conditional mean obtained for the proposed bias-compensated \textit{a priori} SNR estimate (Proposition II) for different number of cells. The input global SNR was 10 dB and the background noise was USASI noise.

We now concentrate on the influence of \textit{a priori} SNR estimation accuracy in the output speech quality of the corresponding noise reduction systems. To do so, we have conducted intensive experiments to compare the proposed bias-compensated approach with local bias in each cell to be more accurate and thus the new \textit{a priori} SNR to be very close to the true value (i.e. the error shift values tend to zero) for instantaneous SNR ranging from –20 dB to +10 dB. Moreover, as far as the local SNR on a frequency bin is greater than –15 dB, the proposed bias-compensated approach yielded the best performance compared to competitive algorithms that can be considered as the state of the art in the noise reduction field (i.e. the conventional DD estimator and the TS-NR algorithm).

In a second experiment (see Fig. 12), we consider a high SNR condition. The scaling factor between the clean speech and noise signals was changed to reach a global input SNR of 15 dB. Here again we observe the same behavior as the one described previously for the case of low input SNR (Fig. 11). Similar improvements were also noted with other input SNR levels (i.e. input SNR = 0 dB, 3 dB, 10 dB . . . , etc), and with different types of noise (white noise, USASI noise, F16 aircraft noise, street noise, and babble noise). We can thus conclude on the superiority of the proposed bias-compensated technique in comparison with the two other conventional techniques, i.e. DD and TS-NR. This superiority will be confirmed in Section 5.3 by through the evaluation of four objective criteria of the proposed approach (i.e. Proposed I and Proposed II).

5.3. Measures of quality

We now concentrate on the influence of \textit{a priori} SNR estimation accuracy in the output speech quality of the corresponding noise reduction systems. To do so, we have conducted intensive experiments to compare the proposed bias-compensated approach with
the ‘conventional’ TS-NR [7] and DD algorithms [2]. To compare these algorithms, several objective criteria are considered: the Normalized Covariance Metric (NCM) criterion, the Coherence Speech Intelligibility Index (CSI) criterion, the Signal-to-Noise Ratio loss (SNRloss) [30], and the Perceptual Evaluation of Speech Quality (PESQ) measure [28,29], which has been found to correlate highly with speech quality in the context of noise-corrupted speech processed via noise suppression algorithms. To evaluate these four criteria, we have used different segments of randomly selected speech sentences from a wide real corpus measured in France Telecom audio research laboratory (testing database) which is commonly used for evaluating the output speech quality of speech coders for ITU-T tests. This speech database consists in phonetically balanced double sentences (input level 0 dB) with relatively low word-context predictability. Clean speech utterance sequences were composed of 6 utterances from 4 speakers (2 male and 2 female) totaling 24 utterances. The speech segments were originally sampled at 16 kHz and then downsampled to 8 kHz with silence periods at the beginning (to estimate the noise PSD) and at the end of each sentence. These sentences were corrupted by the previously described noises, namely, white noise, USASI noise, F16 aircraft noise, babble noise, and street noise. All these noises were downsampled at 8 kHz and quantized at 16 bits. The signal to noise ratio was adjusted by means of the speech voltmeter (ITU-T Recomill.yahoo.fr mendation P56). All the parameters of the noise reduction systems (frame length, overlap, FFT size, ...) are kept unchanged (see Section 5).

Table 1 compares a modified version of the normalized covariance metric (NCM) measure which corresponds to a speech-based transmission index (STI) measure that has been specifically modified to predict the intelligibility of nonlinearly-processed speech via hearing-aid algorithms or noise-suppression. The NCM in each frequency band is computed from the estimate of the normalized covariance determined from the envelopes of the noise-free and enhanced signals (output of the speech enhancement system) in each frequency band. We select the modified implementation of the NCM measure proposed in [29] which uses signal-dependent weighting functions and performed substantially better than the conventional NCM measure which uses fixed (for all speech stimuli) weights.

Table 1 provides the modified NCM criterion values obtained with the four noise reduction approaches (DD and TS-NR with/ without bias compensation) and for three levels of the global input SNR, i.e. 0 dB, 5 dB, and 10 dB. The column indicated as “UN” shows the scores obtained with noise-corrupted (unprocessed) stimuli, the columns indicated as “DD” and “TS-NR” show the baseline scores obtained with the ‘conventional’ enhancement algorithms, while the columns indicated as “PropI” and “PropII” show the results obtained with the proposed bias compensated approaches, respectively DD compensated and TS-NR compensated. We observe a large gap between the NCM measures obtained with the two conventional techniques and the NCM measures obtained with the proposed algorithms. Moreover, in comparison with the ‘conventional’ algorithms, we see that the two proposed bias-compensated noise reduction algorithms (Propositions I and II) produce an enhanced speech signal with the highest speech intelligibility for all five types of noise tested and for all SNR levels proposed. These results are in agreement with the fact that the enhanced signals sounded much clearer and with much less speech distortion than output speech signals provided by conventional techniques. According to a listening test, the output speech signal quality, obtained with the two proposed algorithms (Proposed I and II), is more harmonic with high frequency content and more vowels sounds.

The previous conclusions, obtained from the NCM criterion evaluation (see Table 1), are also confirmed when examining the results obtained with the center-clipping intelligibility index (CSI) [29]. The CSI measure is a spectral-based speech intelligibility measure, which is computed by using the magnitude squared coherence between the noise-free and enhanced signals (output of the speech enhancement system) in each frequency band. The results are presented in Table 2 for the considered noise reduction algorithms and for global input SNRs of 0 dB, 5 dB and 10 dB. As for the previous NCM measure, the proposed bias-compensated noise reduction algorithms (Prop and PropII) outperform the conventional DD and TS-NR approaches. As the modified CSI and NCM measures incorporating signal-specific weighting information have been found to perform the best in terms of predicting speech intelligibility in noise, the obtained speech signals with the proposed methods, i.e. Proposed I and Proposed II, have been highly improved in terms of intelligibility. This is one of the good characteristics of the proposed methods which provide enhanced speech signals very correlated with the original one.

In Table 3, we have collected the results for the average SNR loss criterion obtained for the previous noise reduction algorithms. The recently proposed SNR loss in [30], corresponds to the difference (expressed in dB) between the excitation spectrum of the clean signal (noise-free speech) in a given frequency band before

Table 1
Normalized covariance metric measure (ncm) comparisons: PropI: DD + Proposed technique, PropII: TS-NR + Proposed Technique.

<table>
<thead>
<tr>
<th>Noise type</th>
<th>SNR (dB)</th>
<th>NCM</th>
<th>UN</th>
<th>DD</th>
<th>TSNR</th>
<th>PropI</th>
<th>PropII</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>0</td>
<td>0.700</td>
<td>0.703</td>
<td>0.715</td>
<td>0.704</td>
<td>0.740</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.836</td>
<td>0.792</td>
<td>0.812</td>
<td>0.814</td>
<td>0.856</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.936</td>
<td>0.867</td>
<td>0.895</td>
<td>0.913</td>
<td>0.949</td>
<td></td>
</tr>
<tr>
<td>USASI</td>
<td>0</td>
<td>0.849</td>
<td>0.763</td>
<td>0.813</td>
<td>0.828</td>
<td>0.884</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.962</td>
<td>0.891</td>
<td>0.928</td>
<td>0.952</td>
<td>0.975</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.995</td>
<td>0.966</td>
<td>0.986</td>
<td>0.994</td>
<td>0.998</td>
<td></td>
</tr>
<tr>
<td>F16 aircraft</td>
<td>0</td>
<td>0.919</td>
<td>0.866</td>
<td>0.907</td>
<td>0.900</td>
<td>0.939</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.973</td>
<td>0.950</td>
<td>0.975</td>
<td>0.968</td>
<td>0.978</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.993</td>
<td>0.989</td>
<td>0.995</td>
<td>0.990</td>
<td>0.996</td>
<td></td>
</tr>
<tr>
<td>Babble</td>
<td>0</td>
<td>0.529</td>
<td>0.454</td>
<td>0.488</td>
<td>0.588</td>
<td>0.587</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.731</td>
<td>0.683</td>
<td>0.709</td>
<td>0.760</td>
<td>0.768</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.885</td>
<td>0.839</td>
<td>0.867</td>
<td>0.903</td>
<td>0.912</td>
<td></td>
</tr>
<tr>
<td>Street</td>
<td>0</td>
<td>0.586</td>
<td>0.570</td>
<td>0.599</td>
<td>0.626</td>
<td>0.649</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.777</td>
<td>0.744</td>
<td>0.772</td>
<td>0.788</td>
<td>0.808</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.895</td>
<td>0.848</td>
<td>0.873</td>
<td>0.897</td>
<td>0.917</td>
<td></td>
</tr>
</tbody>
</table>

Table 2
Center-clipping intelligibility index (csii) comparisons: PropI: DD + Proposed technique, PropII: TS-NR + Proposed Technique.

<table>
<thead>
<tr>
<th>Noise type</th>
<th>SNR (dB)</th>
<th>CSI</th>
<th>UN</th>
<th>DD</th>
<th>TSNR</th>
<th>PropI</th>
<th>PropII</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>0</td>
<td>0.625</td>
<td>0.653</td>
<td>0.680</td>
<td>0.684</td>
<td>0.699</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.733</td>
<td>0.747</td>
<td>0.763</td>
<td>0.735</td>
<td>0.763</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.831</td>
<td>0.819</td>
<td>0.841</td>
<td>0.833</td>
<td>0.854</td>
<td></td>
</tr>
<tr>
<td>USASI</td>
<td>0</td>
<td>0.715</td>
<td>0.692</td>
<td>0.727</td>
<td>0.718</td>
<td>0.758</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.865</td>
<td>0.831</td>
<td>0.857</td>
<td>0.866</td>
<td>0.891</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.968</td>
<td>0.923</td>
<td>0.946</td>
<td>0.956</td>
<td>0.897</td>
<td></td>
</tr>
<tr>
<td>F16 aircraft</td>
<td>0</td>
<td>0.775</td>
<td>0.793</td>
<td>0.832</td>
<td>0.842</td>
<td>0.852</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.893</td>
<td>0.898</td>
<td>0.921</td>
<td>0.895</td>
<td>0.923</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.969</td>
<td>0.969</td>
<td>0.982</td>
<td>0.970</td>
<td>0.983</td>
<td></td>
</tr>
<tr>
<td>Babble</td>
<td>0</td>
<td>0.648</td>
<td>0.603</td>
<td>0.624</td>
<td>0.649</td>
<td>0.671</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.799</td>
<td>0.739</td>
<td>0.759</td>
<td>0.800</td>
<td>0.818</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.910</td>
<td>0.855</td>
<td>0.879</td>
<td>0.911</td>
<td>0.921</td>
<td></td>
</tr>
<tr>
<td>Street</td>
<td>0</td>
<td>0.594</td>
<td>0.579</td>
<td>0.607</td>
<td>0.596</td>
<td>0.626</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.745</td>
<td>0.737</td>
<td>0.765</td>
<td>0.746</td>
<td>0.775</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.848</td>
<td>0.826</td>
<td>0.847</td>
<td>0.850</td>
<td>0.873</td>
<td></td>
</tr>
</tbody>
</table>
processing to the excitation of the processed signal (enhanced) in the same critical band [30]. The average SNRloss is finally computed by averaging the critical-band spectral representation of the SNRloss values over all frames and across all bands. One should also note that it is reasonable to expect that as the averaged SNR level increases, the estimated spectrum \( \hat{S}(p,k) \) approaches the clean spectrum \( S(p,k) \) and consequently the average SNR loss tends towards zero. These values are given in Table 3, for three conditions of global input SNR, i.e. 0 dB, 5 dB, and 10 dB. The noisy speech condition has not been evaluated since there is no SNRloss that is introduced by noise suppression algorithm in this case (or equivalently the SNRloss is not meaningful). We can see that the SNRloss characteristics are directly proportional to the level of the input SNR, i.e. the SNRloss decreases as a function of the input SNR. According to these results, it is clear that the proposed bias-compensated algorithms (PropI and PropII) lead to lowest SNRloss at the processing output for all five types of noise tested and for all SNR levels proposed. From the definition of the SNRloss measure, the previous results provide evidence that the estimated spectrum available at the output of the proposed bias-compensated noise reduction technique is much closer to the original speech spectrum (clean speech).

Finally, we consider the perceptual evaluation of speech quality (PESQ) objective measure which has been recommended by ITU-T (Recommendation P. 862) for speech quality assessment of narrow-band handset telephony and speech codecs [31]. Recently, the PESQ measure has also been shown in [28,29], to correlate well with subjective ratings of speech distortion introduced by noise-suppression algorithms. The PESQ is computed as follows. The original signal (clean) and the speech component available at the output of the noise reduction system are first level equalized to a standard listening level and the signals are time aligned to correct for time delays, and then processed through an auditory transform to obtain the loudness spectra.

Table 4 presents the results of the PESQ measures obtained with the five considered noise reduction algorithms. This experiment was carried out with three levels of the global input SNR, i.e. 0 dB, 5 dB, and 10 dB. From these results, we can easily see that the two proposed bias-compensated noise reduction algorithms perform better than the other ones. We can say that the results obtained with the PESQ criterion confirm the accuracy and the efficiency of the proposed technique to positively enhance the estimation of the a priori SNR and hence improve the overall quality of the output speech signal.

To conclude this section dedicated to the characterization of the inherent properties of the enhanced speech available at the output of the considered noise reduction algorithms, it is clear that the proposed bias compensated approaches performed the best in terms of speech quality and intelligibility assessments. For all types of noise tested, for all SNR levels proposed and for all objective measures considered in this study, these algorithms performed substantially better than the considered conventional approaches. Such result confirms the interest of the proposed a priori SNR estimation technique which is based on the compensation of the bias over several cells in the two-dimensional \( z = (\hat{e}, \hat{y}) \) plane. Informal listening tests confirmed that the proposed bias-compensated noise reduction algorithm yielded the highest quality, consistent with the different objective results presented in this section.

### Table 3

<table>
<thead>
<tr>
<th>Noise type</th>
<th>SNR (dB)</th>
<th>SNRloss</th>
<th>UN</th>
<th>DD</th>
<th>TS-NR</th>
<th>PropI</th>
<th>PropII</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>0</td>
<td>0.975</td>
<td>0.981</td>
<td>0.924</td>
<td>0.927</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.981</td>
<td>0.982</td>
<td>0.928</td>
<td>0.927</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.982</td>
<td>0.981</td>
<td>0.921</td>
<td>0.923</td>
<td></td>
<td></td>
</tr>
<tr>
<td>USASI</td>
<td>0</td>
<td>0.976</td>
<td>0.979</td>
<td>0.957</td>
<td>0.925</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.977</td>
<td>0.979</td>
<td>0.917</td>
<td>0.934</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.980</td>
<td>0.981</td>
<td>0.931</td>
<td>0.948</td>
<td></td>
<td></td>
</tr>
<tr>
<td>F16 aircraft</td>
<td>0</td>
<td>0.978</td>
<td>0.979</td>
<td>0.917</td>
<td>0.927</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.978</td>
<td>0.981</td>
<td>0.924</td>
<td>0.938</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.980</td>
<td>0.980</td>
<td>0.937</td>
<td>0.948</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Babble</td>
<td>0</td>
<td>0.969</td>
<td>0.969</td>
<td>0.919</td>
<td>0.944</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.969</td>
<td>0.970</td>
<td>0.928</td>
<td>0.947</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.973</td>
<td>0.973</td>
<td>0.937</td>
<td>0.950</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Street</td>
<td>0</td>
<td>0.973</td>
<td>0.976</td>
<td>0.928</td>
<td>0.932</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.977</td>
<td>0.977</td>
<td>0.920</td>
<td>0.933</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.980</td>
<td>0.979</td>
<td>0.924</td>
<td>0.939</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 4

<table>
<thead>
<tr>
<th>Noise type</th>
<th>SNR (dB)</th>
<th>PESQ</th>
<th>UN</th>
<th>DD</th>
<th>TS-NR</th>
<th>PropI</th>
<th>PropII</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>0</td>
<td>1.27</td>
<td>2.51</td>
<td>2.29</td>
<td>3.02</td>
<td>2.94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1.61</td>
<td>2.55</td>
<td>2.31</td>
<td>3.11</td>
<td>3.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>1.96</td>
<td>2.64</td>
<td>2.47</td>
<td>3.20</td>
<td>3.11</td>
<td></td>
</tr>
<tr>
<td>USASI</td>
<td>0</td>
<td>1.68</td>
<td>2.26</td>
<td>2.08</td>
<td>3.03</td>
<td>2.86</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2.05</td>
<td>2.52</td>
<td>2.41</td>
<td>3.21</td>
<td>3.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>2.39</td>
<td>2.76</td>
<td>2.68</td>
<td>3.45</td>
<td>3.20</td>
<td></td>
</tr>
<tr>
<td>F16 aircraft</td>
<td>0</td>
<td>1.87</td>
<td>2.49</td>
<td>2.36</td>
<td>3.08</td>
<td>2.99</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>2.20</td>
<td>2.73</td>
<td>2.63</td>
<td>3.34</td>
<td>3.17</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>2.50</td>
<td>2.93</td>
<td>2.88</td>
<td>3.50</td>
<td>3.27</td>
<td></td>
</tr>
<tr>
<td>Babble</td>
<td>0</td>
<td>1.60</td>
<td>2.22</td>
<td>2.06</td>
<td>2.94</td>
<td>2.74</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1.91</td>
<td>2.39</td>
<td>2.25</td>
<td>3.12</td>
<td>2.94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>2.28</td>
<td>2.60</td>
<td>2.52</td>
<td>3.31</td>
<td>3.08</td>
<td></td>
</tr>
<tr>
<td>Street</td>
<td>0</td>
<td>1.39</td>
<td>2.31</td>
<td>2.15</td>
<td>2.99</td>
<td>2.87</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>1.68</td>
<td>2.44</td>
<td>2.26</td>
<td>3.09</td>
<td>2.99</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>2.02</td>
<td>2.53</td>
<td>2.39</td>
<td>3.18</td>
<td>3.11</td>
<td></td>
</tr>
</tbody>
</table>
the a priori SNR estimate by optimizing within each cell the trade-off between the bias term and the variance of the estimator.

References


Mohamed Djendi received his first Ph.D. degree in Electronics-Signal and Telecommunications from the High National Polytechnic School (ENSPT) of Algiers, Algeria, in 2006. He received, in January 2010, his second Ph.D. degree in Signal Processing and Telecommunications from the University of Rennes—IRISA/ENSSAT, France. He holds a Postdoctoral position at High National School of Technologies and Sciences (ENSSAT, France), engaging in research on Signal Processing and Communication Systems. In this Postdoctoral position periods, he was supervised by Professor Pascal Scalart. From 2001 to 2011 and from 2012 to present, he has been a full Professor and researcher at Blida University and in LATSI research Laboratory. His current research activities include and not limited to adaptive filtering algorithms, noise reduction and speech enhancement algorithms, blind source separation, and digital communication.

Pascal Scalart received the M.S. (1989), and the Ph.D. degree (1992) in Signal Processing from the University of Rennes (France). In 1993, he held a postdoctoral position (Laval University, Canada) engaging in research on digital communications. From 1994 to 2001, he was with the R&D center of France Telecom working on speech signal processing techniques (noise reduction, acoustic echo cancellation, acoustic antennas, post-filters ...). He has authored over 100 papers and holds several patents in key audio technologies such as ITU-T G.711 speech coder, or noise reduction for W-CDMA terminals. In 2001, he joined the engineering school ENSSAT, University of Rennes where he is currently a Professor, head of the Electronic and Computer Engineering Department.