NOISY SPEECH DEREVERBERATION AS A SIMO SYSTEM IDENTIFICATION ISSUE

W. Bobillet[1], E. Grivel[1], R. Guidorzi[2] and M. Najim[1]
[1] Equipe Signal et Image, UMR 51 31 LAPS,
351 Cours de la libération, 33405 Talence Cedex, France
[2] Dipartimento di Elettronica, Informatica e Sistemistica, Università di Bologna,
Viale del Risorgimento 2, 40136 Bologna, Italy
email: {bobillet, eric.grivel, najim}@laps.u-bordeaux1.fr, rguidorzi@deis.unibo.it

ABSTRACT

This paper deals with the speech dereverberation issue based on a Single Input Multiple Output (SIMO) system, when the reverberations are modeled by Finite Impulse Response (FIR) filters.

In most of the existing methods, the authors assume either that the white noises have the same variance or that the noise statistics are available. Here, we investigate the blind speech deconvolution using two microphones, when the white noise variances are not equal. For this purpose, we present a modified version of an identification approach previously developed in the framework of control and based on the properties of the definiteness and the positiveness of the autocorrelation matrices of the reverberated versions of the speech and the observations. This makes it possible to estimate both the variances of the additive noises and the FIR. Then, the speech signal is retrieved in the Least Square (LS) or Minimum Variance (MV) sense.

Index Terms — speech enhancement, unbalanced additive noises, blind dereverberation, FIR estimation.

1. INTRODUCTION

Noise reduction algorithms are required in numerous applications such as hands-free communication, speech recognition and teleconferencing systems.

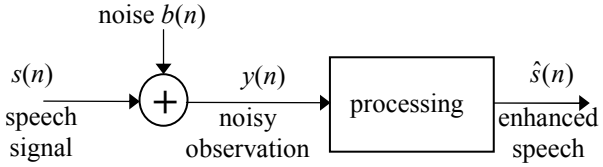

FIG 1: enhancing a speech signal contaminated by an additive noise

When a single microphone records a speech signal corrupted by an additive background noise, several approaches have been developed to enhance the speech. Among them, non-parametric methods using short time spectral attenuation [6] provide significant results. An alternative consists in carrying out a parametric solution based on an *a priori* model of the speech. Thus, an AutoRegressive (AR) model leads to Kalman filtering based methods [9] [10], whereas the complex exponential models yield various subspace techniques [7] [16].

However, the speech enhancement model depicted in FIG 1 does not take into account the acoustic features of the speaker environment, such as the reverberations. For this reason, a Finite Impulse Response (FIR) filter is introduced to represent the multipath propagation effects, leading to spectral distortions into the observations.

If a single microphone is used, only a minimum phase filter can be estimated when taking into account only the Second Order Statistics (SOS) [2]. For this reason, various High Order Statistics (HOS) based methods have been proposed [20] [21]. Nevertheless, a large number of noisy observation samples is necessary in that case. In addition, the HOS methods may fail because the statistical features of the speech signal are strongly non stationary.

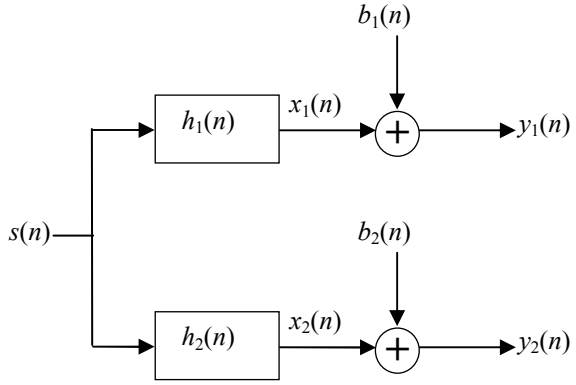

FIG 2: reverberated speech contaminated by additive white noises

Therefore, a multimicrophone based device is often required to complete the so-called equalization of the signal $s(n)$. For this purpose, blind[1] channel identification approaches using the SOS of the observations $y_i(n)$ can be considered, such as the Cross-Relations[2] (CR) [17] or the Two Step Maximum Likelihood [12] algorithms, providing the Signal to Noise Ratio (SNR) is high. An alternative consists in using the subspace approaches [19] [22]. Their purpose is to estimate the FIR from the null space of the autocorrelation matrices of the reverberated versions $x_i(n)$ of the signal $s(n)$. The so-called EigenVector-based Algorithm for Multichannel (EVAM) blind deconvolution proposed in [11] has the advantage of estimating both the channel transfer functions and their orders. To make the channel identification computationally

[1] *i.e.* only the observations are available, without assuming any knowledge on the signal $s(n)$.

[2] $(x_1 * h_2)(n) = (x_2 * h_1)(n) = (s * h_1 * h_2)(n)$.

less intensive, Huang et *al.* [14] have suggested using adaptive approaches such as the LMS algorithm. For this purpose, they aim at minimizing the least mean square error obtained from the Cross-Relations between the reverberated versions of the signal when additive noises are present. Nevertheless, since the algorithm convergence is too slow, the authors have extended their studies into the frequency domain [15].

More recently, Gannot et *al.* [8] have developed a new approach for the multimicrophone speech dereverberation, when the additive noise is coloured. The purpose is to construct the null subspace of the observations matrix by using a Generalized Singular Value Decomposition (GSVD). However, this method can be carried out providing the noise statistics are estimated during speech-free frames or the variances of the white noises are equal.

In the following, we will focus our attention on the unbalanced noise case, *i.e.* when the variances of the additive noises are no longer equal. For this purpose, we investigate the relevance of the blind identification method previously developed by Guidorzi et *al.* [5] in the framework of control. Although its computational cost may be high, this method has the advantage of providing both the FIR and the variances of the additive noises. More particularly, we present here a new algorithm of the blind speech dereverberation we proposed in [3]. Firstly, as an alternative to the criteria presented in [4] and [5], we investigate modified versions. Secondly, instead of estimating the reverberated versions of the signal from noisy observations to retrieve the speech signal, we directly take advantage of the Cross-Relations.

The remainder of the paper is organized as follows: in section 2, the two-microphone based device and the identification method proposed by Guidorzi et *al.* [3] are presented. Then, the new criteria are introduced. In section 3, the original speech estimation is given. Simulation results and conclusions are respectively presented in sections 4 and 5.

2. THE TWO-MICROPHONE SIMO SYSTEM IDENTIFICATION

2.1. System model

The two-microphone based system is depicted in FIG 2. Each observation $y_{\mathrm{i}}(n)$ recorded by the i^{th} microphone is the sum of the reverberated version $x_{\mathrm{i}}(n)$ of the speech and the additive white noise $b_{\mathrm{i}}(n)$ with variance σ_{i}^2:

$$y_{\mathrm{i}}(n) = x_{\mathrm{i}}(n) + b_{\mathrm{i}}(n), \text{ for } \mathrm{i} = 1,2 . \quad (1)$$

Since the reverberations between the speech signal and each microphone are modeled by an L^{th} order Finite Impulse Response (FIR) filter, one has:

$$x_{\mathrm{i}}(n) = (s * h_{\mathrm{i}})(n) = \sum_{k=0}^{L} h_{\mathrm{i}}(k)\, s(n-k) \quad (2)$$

where $h_{\mathrm{i}}(n)$ denotes the L^{th} order acoustic room FIR for the i^{th} microphone. In the sequel, let us introduce the following sample vectors, with $l > 0$:

$$\underline{\mathrm{s}}(n) = \left[s(n-L) \;\; s(n-L+1) \;\; \cdots \;\; s(n+l) \right]^{\mathrm{T}} \quad (3)$$

$$\underline{\mathrm{x}}(n) = \left[x_1(n) \;\; \cdots \;\; x_1(n+l) \;\; x_2(n) \;\; \cdots \;\; x_2(n+l) \right]^{\mathrm{T}} \quad (4)$$

$$\underline{\mathrm{y}}(n) = \left[y_1(n) \;\; \cdots \;\; y_1(n+l) \;\; y_2(n) \;\; \cdots \;\; y_2(n+l) \right]^{\mathrm{T}} \quad (5)$$

$$\underline{\mathrm{b}}(n) = \left[b_1(n) \;\; \cdots \;\; b_1(n+l) \;\; b_2(n) \;\; \cdots \;\; b_2(n+l) \right]^{\mathrm{T}} \quad (6)$$

$$\underline{\mathrm{h}} = \left[h_1(0) \;\; \cdots \;\; h_1(L) \;\; h_2(0) \;\; \cdots \;\; h_2(L) \right]^{\mathrm{T}} \quad (7)$$

and the generalized Sylvester matrix defined as follows:

$$\mathrm{H}_l = \begin{bmatrix} \mathrm{H}_l^1 \\ \mathrm{H}_l^2 \end{bmatrix} \text{ where:} \quad (8)$$

$$\mathrm{H}_l^{\mathrm{i}} = \begin{bmatrix} h_{\mathrm{i}}(L) & h_{\mathrm{i}}(L-1) & \cdots & h_{\mathrm{i}}(0) & 0 \\ & \ddots & \ddots & & \ddots \\ 0 & h_{\mathrm{i}}(L) & h_{\mathrm{i}}(L-1) & \cdots & h_{\mathrm{i}}(0) \end{bmatrix}_{(l+1)\times(l+\mathrm{L}+1)} . \quad (9)$$

Shaping the relations (1) and (2) in a matrix form leads to the following system:

$$\underline{\mathrm{y}}(n) = \underline{\mathrm{x}}(n) + \underline{\mathrm{b}}(n) \quad (10)$$

$$\underline{\mathrm{x}}(n) = \mathrm{H}_l \underline{\mathrm{s}}(n) . \quad (11)$$

Mathematically speaking, the blind multichannel identification issue consists in estimating the FIR $h_{\mathrm{i}}(n)$ by exploiting the inherent structure of the SIMO system. Indeed, under the identifiability conditions given in [13], the $(l-L+1)$ dimensional null space of the sample correlation matrix $\mathrm{R}_x^l = \frac{1}{N-l} \sum_{k=0}^{N-l-1} \underline{\mathrm{x}}(k) \underline{\mathrm{x}}^{\mathrm{T}}(k)$ is given by:

$$\mathrm{Null}\left(\mathrm{R}_x^l\right) = \mathrm{Span}\left(\mathrm{M}_l^{\mathrm{T}}\right) \quad (12)$$

where M_l is defined by:

$$\mathrm{M}_l = \left[\mathrm{H}_l^2 \;\; -\mathrm{H}_l^1 \right] . \quad (13)$$

However, R_x^l is unknown and hence must be estimated.

2.2. How to estimate R_x^l and subsequently the FIR?

If the additive noises are assumed stationary, mutually uncorrelated and uncorrelated with the speech, one has:

$$\mathrm{R}_y^l = \mathrm{R}_x^l + \mathrm{R}_b^l \quad (14)$$

providing a large number of samples N is available. Note that R_y^l and R_b^l are the sample autocorrelation matrices of respectively $\underline{\mathrm{y}}(n)$ and $\underline{\mathrm{b}}(n)$.

In the balanced noise case, one has:

$$\mathrm{R}_b^l = \sigma^2 \mathrm{I}_{2(l+1)} \tag{15}$$

where $\sigma^2 = \sigma_1^2 = \sigma_2^2$ denotes the variance of the additive noises. Thus, the matrices R_x^l and R_y^l have the same eigenvectors. Moreover, their corresponding eigenvalues λ_x^k and λ_y^k satisfy:

$$\lambda_y^k = \lambda_x^k + \sigma^2 . \tag{16}$$

The null space of R_x^l is therefore the eigenspace of R_y^l associated to the smallest eigenvalue σ^2.

However, the previous remark is no longer true in the unbalanced noise case. Thus, the blind estimation of the FIR will be based on the following properties:

i) R_x^l *is positive and singular, i.e.* $\mathrm{R}_x^l \geq 0$.

ii) R_y^l *is positive definite, i.e.* $\mathrm{R}_y^l > 0$

and the theorem initially presented in [1] which states that:

The pair $P^* = \left(\sigma_1^2, \sigma_2^2\right)$ *is the only solution of the following equation system*:

$$\mathrm{R}_x^l(\alpha,\beta) = \mathrm{R}_y^l - \mathrm{R}_b^l(\alpha,\beta) \geq 0 \ \textit{for}\ l \geq L \tag{17}$$

and where:

$$\mathrm{R}_b^l(\alpha,\beta) = \mathrm{diag}\left(\alpha \mathrm{I}_{l+1}, \beta \mathrm{I}_{l+1}\right) = \begin{bmatrix} \alpha \mathrm{I}_{l+1} & 0 \\ 0 & \beta \mathrm{I}_{l+1} \end{bmatrix}. \tag{18}$$

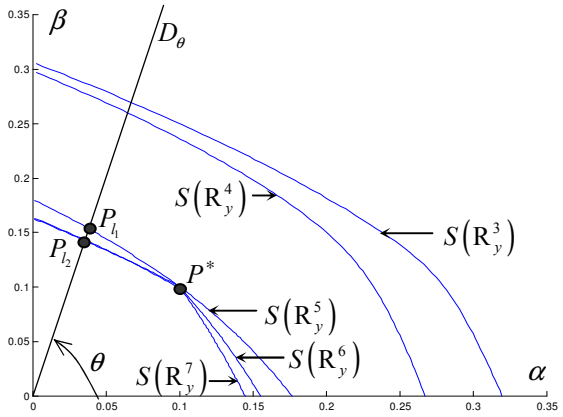

FIG 3: convex curves $S\left(\mathrm{R}_y^l\right)$ with $L = 5$ and $l = 3,..,7$.

The set of solutions $S\left(\mathrm{R}_y^l\right)$ of each equation (17) corresponds to a convex curve in the plane (α,β).

In the following, we will denote P_l the intersection between $S\left(\mathrm{R}_y^l\right)$ and the straight line D_θ passing through the origin and whose slope is $\tan(\theta)$ (See FIG 3). For the sake of simplicity, the resulting matrix $\mathrm{R}_x^l(\alpha,\beta)$ defined in relation (17) will be denoted $\mathrm{R}_x^l(\theta)$.

In theory, estimating the variances consists in searching the common point P^* belonging to the curves $S\left(\mathrm{R}_y^l\right)$ with $l \geq L$. However, in real cases, P^* does not exist. For this reason, Guidorzi et *al.* have proposed three criteria to estimate $\left(\sigma_1^2, \sigma_2^2\right)$ and subsequently the FIR, $h_1(n)$ and $h_2(n)$. More precisely, the so-called Input Covariance Matching (ICM) [4], the Shift Relations (SR) [5] and the Error Covariance Matching (ECM) [5] criteria aim at minimizing a criterion, denoted $J_{criterion}(\theta)$. This procedure requires the estimation $\hat{\underline{\mathrm{h}}}(\theta)$ of $\underline{\mathrm{h}}$, deduced from the one-dimensional null space of the matrix $\mathrm{R}_x^L(\theta)$. Indeed, for $l = L$ the relation (12) becomes:

$$\mathrm{Null}\left(\mathrm{R}_x^L\right) = \mathrm{Span}\left(\left[h_2(L)\cdots h_2(0) - h_1(L)\cdots - h_1(0)\right]^{\mathrm{T}}\right). \tag{19}$$

However, in the framework of speech dereverberation, it is difficult to extract the null space. Indeed, the eigenvalues of R_x^L may be too close to separate the noise eigenvalue from the smallest signal eigenvalues, due to the high sizes of the matrices and the speech signal features. Thus, we propose two improvements, which consist in:

i) estimating the FIR by taking advantage of the inherent structure of the null space of higher size autocorrelation matrices, given by the relation (13). It should be noted that this structure only appears in the two channel case.

ii) generalizing the above criteria to higher size autocorrelation matrices R_x^l, with $l > L$.

2.3. Alternative estimation of the FIRs

Given (12), we theoretically have:

$$\mathrm{R}_x^l \mathrm{M}_l^{\mathrm{T}} = 0 \tag{20}$$

Then, the estimated vector $\hat{\underline{\mathrm{h}}}$ satisfies:

$$\hat{\underline{\mathrm{h}}}(\theta) = \underset{\|\underline{\mathrm{u}}\|_2 = 1}{\arg\min}\left\{C(\underline{\mathrm{u}},\theta)\right\} \tag{21}$$

where $C(\underline{\mathrm{u}},\theta)$ is the following quadratic function of components of any $(2L+2)$ vector $\underline{\mathrm{u}}$:

$$C(\underline{\mathrm{u}},\theta) = \mathrm{Trace}\left(\mathrm{M}_l(\underline{\mathrm{u}})\mathrm{R}_x^l(\theta)\mathrm{M}_l^{\mathrm{T}}(\underline{\mathrm{u}})\right) \tag{22}$$

which is minimized under the constraint $\|\underline{\mathrm{u}}\|_2 = 1$.

This hence leads to the following algorithm.

Algorithm

i) Start from a angle θ and compute the point P_{l_1} of $S\left(\mathrm{R}_y^{l_1}\right)$ and if necessary the point P_{l_2} of $\mathrm{S}\left(\mathrm{R}_y^{l_2}\right)$, with $l_2 > l_1 \geq L$. See FIG 3.

ii) Compute the matrix $\mathrm{R}_x^{l_1}(\theta)$.

iii) Estimate the corresponding FIR $\hat{\underline{\mathrm{h}}}(\theta)$ using (21).

iv) Evaluate the criterion $J_{criterion}(\theta)$ for $\theta \in \left[0; \frac{\pi}{2}\right]$.

Then, the minimum value provides the estimations $\hat{\underline{\mathrm{h}}}$ and $\left(\hat{\sigma}_1^2, \hat{\sigma}_2^2\right)$ of $\underline{\mathrm{h}}$ and $\left(\sigma_1^2, \sigma_2^2\right)$ respectively.

Let us now focus on the modified criteria $J_{criterion}(\theta)$.

2.4. Modified criteria to be tested

2.4.1. How to Improve the ICM criterion?

The idea is to compare the matrices $\mathrm{R}_x^{l_1}(\theta)$ and $\mathrm{R}_x^{l_2}(\theta)$ by taking into account the structure of the two microphone system. For this purpose, let us introduce the $(2l+1)\times(2l+1)$ speech sample autocorrelation matrices:

$$\mathrm{R}_s^l = \frac{1}{N-l}\sum_{k=0}^{N-l-1} \underline{\mathrm{s}}(k)\underline{\mathrm{s}}^{\mathrm{T}}(k), \text{ for } l \in \{l_1, l_2\}. \tag{23}$$

According to (11), the matrices R_x^l and R_s^l satisfy the following relationship:

$$\mathrm{R}_x^l = \mathrm{H}_l \mathrm{R}_s^l \mathrm{H}_l^{\mathrm{T}}. \tag{24}$$

Since the identifiability conditions are assumed to be fulfilled, H_l is full column rank and we have:

$$\mathrm{R}_s^l = \mathrm{H}_l^{+} \mathrm{R}_x^l \mathrm{H}_l^{\mathrm{T}+} \tag{25}$$

where the upper script $+$ denotes the Moore-Penrose pseudo-inverse. Moreover, R_s^l is a block matrix that can be expressed as follows:

$$\mathrm{R}_s^l = \begin{bmatrix} \mathrm{R}_s^L & \mathrm{D}_l \\ \mathrm{D}_l^{\mathrm{T}} & \mathrm{C}_l \end{bmatrix}. \tag{26}$$

The upper left $(2L+1)\times(2L+1)$ matrix R_s^L does not depend on l, and hence can be estimated both from $\mathrm{R}_x^{l_1}$ and $\mathrm{R}_x^{l_2}$. Thus, the modified ICM criterion aims at minimizing the difference between the two estimations of R_s^L, denoted $\mathrm{R}_s^L(\theta, l_1)$ and $\mathrm{R}_s^L(\theta, l_2)$, obtained from $\mathrm{R}_x^{l_1}(\theta)$ and $\mathrm{R}_x^{l_2}(\theta)$ respectively.

$$J_{ICM}(\theta) = \left\| \mathrm{R}_s^L(\theta, l_1) - \mathrm{R}_s^L(\theta, l_2) \right\|_{Fro} \tag{27}$$

where $\|.\|_{Fro}$ denotes the Frobenius norm.

2.4.2. How to improve the SR criterion?

The Shifted Relations criterion is based on the following idea. Since the relation (12) holds for every $l > L$, the equation:

$$\mathrm{R}_x^{l_2}(\theta)\mathrm{M}_{l_2}^{\mathrm{T}}(\theta) = 0 \tag{28}$$

has a single solution for $\theta = \theta^*$, *i.e.* $P_{l_1} = P_{l_2} = P^*$. The modified SR criterion is then defined as follows:

$$J_{SR}(\theta) = \left\| \mathrm{R}_s^{l_2}(\theta)\mathrm{M}_{l_2}^{\mathrm{T}}(\theta) \right\|_{Fro}. \tag{29}$$

2.4.3. How to improve the ECM criterion?

The purpose of the Error Covariance Matching criterion is to minimize the norm of the autocorrelation matrix of the Cross Relation error defined as follows:

$$\begin{aligned} e(n) &= \left(h_1(\theta) * y_2\right)(n) - \left(h_2(\theta) * y_1\right)(n) \\ &= e_x(n) + e_b(n) \end{aligned} \tag{30}$$

where:

$$\begin{aligned} e_x(n) &= \left(h_1(\theta) * x_2\right)(n) - \left(h_2(\theta) * x_1\right)(n) \\ e_b(n) &= \left(h_1(\theta) * b_2\right)(n) - \left(h_2(\theta) * b_1\right)(n) \end{aligned} \tag{31}$$

As the additive noises $b_1(n)$ and $b_2(n)$ are zero-mean processes uncorrelated with the speech, $e_x(n)$ and $e_b(n)$ are mutually uncorrelated and hence their autocovariance matrices satisfy:

$$\mathrm{R}_e(\theta) = \mathrm{R}_{e_x}(\theta) + \mathrm{R}_{e_b}(\theta) \tag{32}$$

with obvious notations. Moreover, since $e_b(n)$ is the sum of two Moving Average processes driven by white noises input $b_1(n)$, according to (31), it can be proved that:

$$\mathrm{R}_{e_b}(\theta) = \alpha \mathrm{H}_L^1(\theta)\mathrm{H}_L^{1\mathrm{T}}(\theta) + \beta \mathrm{H}_L^2(\theta)\mathrm{H}_L^{2\mathrm{T}}(\theta). \tag{33}$$

When $\theta = \theta^*$, $e_x(n) = 0$, which implies that:

$$\mathrm{R}_{e_x}(\theta^*) = 0. \tag{34}$$

Thus, the new criterion is defined as follows:

$$J_{ECM}(\theta) = \left\| \mathrm{R}_{e_x}(\theta) \right\|_{Fro} = \left\| \mathrm{R}_e(\theta) - \mathrm{R}_{e_b}(\theta) \right\|_{Fro}. \tag{35}$$

3. ORIGINAL SPEECH ESTIMATION

Once the FIR are estimated, one can retrieve the speech by using the relations (10) and (11). Since $l = N-1$, and by omitting the indices for the sake of simplicity, the original speech satisfies the Cross-Relations that are here viewed as linear constraints:

$$\underline{\mathrm{x}} \in \mathrm{Span}(\mathrm{H}) \tag{36}$$

or equivalently:

$$\mathrm{M}\underline{\mathrm{x}} = 0. \tag{37}$$

The Least Square (LS) estimation $\underline{\mathrm{x}}_{LS}$ and the Minimum Variance (MV) estimation $\underline{\mathrm{x}}_{MV}$, when assuming that the additive noises $b_1(n)$ are Gaussian, are given by:

$$\underline{\mathrm{x}}_{LS} = \left[\mathrm{I}_{2N} - \mathrm{M}^{\mathrm{T}}\left(\mathrm{M}\mathrm{M}^{\mathrm{T}}\right)^{-1}\mathrm{M}\right]\underline{\mathrm{y}} = \mathrm{P}_{\perp}\underline{\mathrm{y}} \tag{38}$$

$$\underline{x}_{MV} = \left[I_{2N} - R_b M^T \left(M R_b M^T \right)^{-1} M \right] \underline{y} = P_{\angle} \underline{y} \quad (39)$$

where the matrices $P_{\perp}$ and $P_{\angle}$ respectively correspond to the orthogonal and the oblique projections on the null space of M. Since we have:

$$\mathrm{Im}\left(M^T\right) = \mathrm{Im}\left(H\right)^{\perp}, \quad (40)$$

where the upper script $\perp$ denotes the orthogonal space, $P_{\perp}$ and $P_{\angle}$ can be rewritten using H as follows:

$$P_{\perp} = H\left(H^T H\right)^{-1} H^T \quad (41)$$

$$P_{\angle} = H\left(H^T R_b^{-1} H\right)^{-1} H^T R_b^{-1}. \quad (42)$$

Hence, the estimated original speech is the only solution of (11), which is given by:

$$\underline{s}_{LS} = H^{+} \underline{x}_{LS} = H^{+} P_{\perp} \underline{y} = \left(H^T H\right)^{-1} H^T \underline{y} \quad (43)$$

$$\underline{s}_{MV} = H^{+} \underline{x}_{MV} = H^{+} P_{\angle} \underline{y} = \left(H^T R_b^{-1} H\right)^{-1} H^T R_b^{-1} \underline{y}. \quad (44)$$

In real case however, we will use the respective estimates $\hat{M}$ and $\hat{H}$ of the matrices M and H, built with the estimated impulse responses $\hat{h}_i(n)$.

4. SIMULATION RESULTS

4.1. Protocols

In this section, we carry out a comparative study. We first evaluate the original versions of the criteria proposed by Guidorzi et *al.* [4] [5] for FIR identification. Then, we investigate the modified versions proposed in this paper.

Evaluating the estimated impulse response vector $\hat{\underline{h}}$ is a challenging issue since the estimates are only known up to a constant scalar. Since $\hat{\underline{h}}$ and $\underline{h}$ define direction in a $(2L+2)$ dimension space, we consider the following direction error, initially given in [18]:

$$\xi\left(\underline{h}, \hat{\underline{h}}\right) = \sin^2(\phi) = \min_{\gamma} \frac{\left\| \underline{h} - \gamma \hat{\underline{h}} \right\|^2}{\left\| \underline{h} \right\|^2} = 1 - \left(\frac{\underline{h}^T \hat{\underline{h}}}{\left\| \underline{h} \right\| \left\| \hat{\underline{h}} \right\|} \right)^2$$

where $\phi \in \left[0; \frac{\pi}{2}\right]$ is the angle between the two directions given by $\underline{h}$ and $\hat{\underline{h}}$. We can now introduce the Normalized Root Mean Square Error (NRMSE) defined by:

$$\mathrm{NRMSE} = \frac{1}{R} \sum_{k=1}^{R} \xi\left(\underline{h}, \hat{\underline{h}}\right)$$

where $R = 100$ is the number of runs.

The proposed approach is exercised on a speech signal, sampled at 8 kHz, and filtered by 100^{th} order synthetic FIR. 5000 samples have been used for the FIR identification. The reverberated versions of the speech $x_i(n)$ are contaminated by Gaussian additive white noises, corresponding to a Signal to Noise Ration (SNR) varying from 20 dB to 40 dB. The variances of the unbalanced additive noises satisfy $\tan(\theta) = \frac{\sigma_1^2}{\sigma_2^2} = 1.82$.

The simulation results are presented in Table 1-4, for various values of l_1 and l_2. The original version criteria are evaluated in Table 1, where $l_1 = L$ and $l_2 = L+1$.

4.2. Results and comments

It should be first noted that the FIR estimation algorithm may fail to retrieve the impulse responses, especially when the SNR is less than 20 dB. The algorithm is considered as convergent if the norm of the projection of $\hat{\underline{h}}$ on $\mathrm{span}\left(\underline{h}\right)$ is greater than the norm of the projection of $\hat{\underline{h}}$ on $\mathrm{span}\left(\underline{h}\right)^{\perp}$, *i.e.* $\hat{\underline{h}}$ is "closer" to $\underline{h}$ than to any orthogonal direction. Equivalently, the algorithm is convergent if one has $\xi\left(\underline{h}, \hat{\underline{h}}\right) < \frac{1}{2}$.

In the Table 1-4, we provide respectively the convergence rate for the convergent runs, the NRMSE, and its standard deviation.

	SNR=20 dB	SNR=30 dB	SNR=40 dB
Original ICM	100% 0.0964 ± 0.0490	100% 0.0321 ± 0.0209	100% 0.0078 ± 0.0041
Original SR	84% 0.1187 ± 0.1105	96% 0.0368 ± 0.0228	100% 0.0079 ± 0.0041
Original ECM	88% 0.1162 ± 0.0959	98% 0.0335 ± 0.0232	100% 0.0079 ± 0.0041

Table 1: convergence rate NRMSE vs SNR. l_1=L, and l_2=L+1.

	SNR=20 dB	SNR=30dB	SNR=40 dB
New ICM	100% 0.0917 ± 0.0480	100% 0.0302 ± 0.0194	100% 0.0079 ± 0.0042
New SR	94% 0.1086 ± 0.0600	100% 0.0314 ± 0.0209	100% 0.0079 ± 0.0041
New ECM	94% 0.1091 ± 0.0686	100% 0.0311 ± 0.0210	100% 0.0079 ± 0.0040

Table 2: convergence rate NRMSE vs SNR. l_1=L+3, and l_2=L+4.

	SNR=20 dB	SNR=30 dB	SNR=40 dB
New ICM	100% 0.0768 ± 0.0480	100% 0.0311 ± 0.0201	100% 0.0079 ± 0.0043
New SR	96% 0.1091 ± 0.0500	100% 0.0309 ± 0.0197	100% 0.0080 ± 0.0043
New ECM	96% 0.1109 ± 0.0574	100% 0.0313 ± 0.0191	100% 0.0079 ± 0.0042

Table 3: convergence rate NRMSE vs SNR. l_1=L+6, and l_2=L+7.

	SNR = 20 dB	SNR = 30 dB	SNR = 40 dB
New ICM	100% 0.0735 ± 0.0387	100% 0.0303 ± 0.0191	100% 0.0078 ± 0.0042
New SR	96% 0.1114 ± 0.0539	100% 0.303 ± 0.0191	100% 0.0079 ± 0.0042
New ECM	96% 0.1162 ± 0.0600	100% 0.0320 ± 0.0194	100% 0.0078 ± 0.0043

Table 4: NRMSE vs SNR, with $l_1=L+9$, and $l_2=L+10$.

When the SNR is between 20 dB and 30 dB, the new criteria make it possible to improve the FIR estimation. Indeed, the convergence rate of the algorithm for the original SR and ECM criteria is under 88% when the SNR is equal to 20 dB (See Table 1), whereas it increases up to 96% when using the modified criteria (See Table 3-4).
Moreover, the quality of the FIR estimations is improved for the three criteria when using the modified ones. Indeed, when the SNR is equal to 20 dB, the NRMSE and the standard deviation of the NRMSE are lower. This makes the FIR estimations and then the speech dereverberation more accurate.
It should be noted that for high SNR (40 dB) the original and modified criteria provide similar results.

5. CONCLUSIONS

In this paper, the blind dereverberation of a speech signal is carried out, when dealing with two microphones and assuming that the additive white noises are unbalanced. The proposed method aims at estimating both the FIR and the variances of the noises, exploiting the properties of positiveness and definiteness of the sample autocorrelation matrices R_x^l and R_y^l.

Our contribution consists in improving the FIR estimation. Although they are computationally intensive, the modified criteria we propose take into account the structure of the null space of R_x^l and improve the estimation of the FIR and subsequently the dereverberation process.

REFERENCES

[1] S. Beghelli, R.P.Guidorzi U.Soverini, The Frisch Scheme in Dynamic System Identification, Automatica 26, 1990.

[2] J. Benesty, S. Makino, J.Chen (eds.), Speech Enhancement, Springer, chap. 11-12, 2005.

[3] W. Bobillet, E. Grivel, R. Guidorzi and M. Najim, Cancelling convolutive and additive coloured noises for speech enhancement, Proc. IEEE-ICASSP 2004, Montreal, Canada, 2004.

[4] P. Castaldi, R. Diversi, R.P. Guidorzi, U. Soverini, Blind Estimation and Deconvolution of Communication Channels with Unbalanced Noise, Proc. of the 12th IFAC Symposium on System Identification, Santa Barbara, CA, June 2000.

[5] R. Diversi, R. Guidorzi, U. Soverini, Blind Identification and Equalization of Two-Channel FIR Systems in Unbalanced Noise Environments, Signal Processing, vol. 85, no. 1, January 2005.

[6] Y. Ephraim and D. Malah, Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator, IEEE Trans. on ASSP, vol. 32, no. 6, 1984, pp. 1109-1121.

[7] Y. Ephraim and H. L. Van Trees, A signal Subspace Approach for Speech Enhancement, IEEE Trans. on SAP, vol. 3, no. 4, pp 251-266, July 1995.

[8] S. Gannot, M. Moonen, Subspace Method for Multimicrophone Speech Dereverberation, EURASIP Journal on Applied Signal Processing, Special Issue on Signal Processing for Acoustic Communication Systems, vol. 2003, no. 11, pp. 1074-1090, October 2003.

[9] J. D. Gibson, B. Koo and S. D. Gray, Filtering of Colored Noise for Speech Enhancement and Coding, IEEE Trans. on SP, vol. 39, no. 8, pp. 1732-1742, August 1991.

[10] E. Grivel, M. Gabrea and M. Najim, Speech Enhancement as a Realization Issue, Signal Processing, vol. 82, no. 12, pp. 963-978, December 2002.

[11] M.İ. Gürelli, C.L. Nikias, EVAM: An Eigenvector-Based Algorithm for Multichannel Blind Deconvolution of Input Colored Signals, IEEE Trans. on SP, vol. 43, no. 1, February 1995.

[12] Y. Hua, Fast Maximum Likelihood for Blind Identification of Multiple FIR Channels, IEEE Trans. on SP, vol. 44, no. 3, March 1996.

[13] Y. Hua and M. Wax, Strict Identifiability of Multiple FIR Channels Driven by an Unknown Arbitrary Sequence, IEEE Trans. on SP, vol. 44, no. 3, March 1996.

[14] Y.A. Huang, J. Benesty, Adaptive Multi-Channel Least Mean Square and Newton Algorithms for Blind Channel Identification, Signal Processing, vol. 82, no.8, pp. 1127-1138, August 2002.

[15] Y.A. Huang, J. Benesty, Class of Frequency-Domain Adaptive Approaches to Blind Multichannel Identification, IEEE Trans. on SP, vol. 51, no. 1, January 2003.

[16] S. H. Jensen, P. C. Hansen, S. D. Hansen and J. Sorensen, Reduction of Broad Band Noise in Speech by Truncated QSVD, IEEE Trans. on SAP, vol. 3, no. 6, pp. 439-448, November 1995.

[17] H. Liu, G. Xu, and L. Tong, A Deterministic Approach to Blind Identification of Multi-Channel FIR Systems, Proc. IEEE ICASSP '94, pp. 581-584, April 1994.

[18] D.R. Morgan, J. Benesty, M. Sondhi, On the Evaluation of Estimated Impulse Responses, IEEE Signal Processing Letters, vol. 5, no. 7, March 1998.

[19] E. Moulines, P. Duhamel, J. Cardoso and S. Mayrargue, Subspace Methods for the Blind Identification of Multichannel FIR Filters, IEEE Trans. on SP, vol. 43, no. 2, February 1995.

[20] O. Shalvi and E. Weinstein, New Criteria for Blind Deconvolution of Nonminimum Phase Systems (Channels), Trans. on IT, vol. 36, no. 2, March 1990.

[21] J.K Tugnait, Identification of Linear Stochastic System via Second and Fourth-Order Cumulant Matching, , IEEE Trans. on IT, vol. 33, no. 3, May 1987.

[22] G. Xu, H. Liu, L. Tong, T. Kailath, , A Least Squares Approach to Blind Channel Identification, IEEE Trans. on SP, vol. 43, no. 12, December 1995.