

Subspace Methods for Multi-Microphone Speech Dereverberation

Sharon Gannot

*Department of Electrical Engineering, Technion — Israel Institute of Technology,
Technion City, Haifa 32000, Israel
E-mail: gannot@siglab.technion.ac.il;
Tel.: +972 4 8294756; Fax: +972 4 8323041.*

Marc Moonen

*Department of Electrical Engineering, ESAT-SISTA
K.U.Leuven,
Kasteelpark Arenberg 10,
3001 Leuven-Heverlee, Belgium
E-mail: marc.moonen@esat.kuleuven.ac.be*

Abstract

A novel approach for multi-microphone speech dereverberation is presented. The method is based on the construction of the null subspace of the data matrix in the presence of colored noise, using the *generalized singular value decomposition* (GSVD) technique, or the *generalized eigenvalue decomposition* (GEVD) of the respective correlation matrices. The special Sylvester structure of the filtering matrix, related to this subspace, is exploited for deriving a *total least squares* (TLS) estimate for the *acoustical transfer functions* (ATFs). Other, less robust but computationally more efficient methods are derived based on the same structure and on the QR decomposition (QRD). A preliminary study of the incorporation of the subspace method into a

This research work was partly carried out at the ESAT laboratory of the Katholieke Universiteit Leuven, in the frame of the Interuniversity Attraction Pole IUAP P4-02, *Modeling, Identification, Simulation and Control of Complex Systems*, the Concerted Research Action *Mathematical Engineering Techniques for Information and Communication Systems* (GOA-MEFISTO-666) of the Flemish Government and the IT-project *Multi-microphone Signal Enhancement Techniques for handsfree telephony and voice controlled systems (MUSETTE-2)* of the I.W.T., and was partially sponsored by Philips-ITCL.

subband framework proves to be efficient, although some problems remain open. Speech reconstruction is achieved by virtue of the *matched filter beamformer* (MFBF). An experimental study supports the potential of the proposed methods.

Keywords

Speech Dereverberation , Microphone Arrays, Subspace Methods, Subband Structures.

I. INTRODUCTION

In many speech communication applications, the recorded speech signal is subject to reflections on the room walls and other objects on its way from the source to the microphones. The resulting speech signal is then called reverberated. The quality of the speech signal might deteriorate severely and this can even cause a degradation in intelligibility. Subsequent processing of the speech signal, such as speech coding or automatic speech recognition might be rendered useless in the presence of reverberated speech. Although single-microphone dereverberation techniques do exist, the most successful methods for dereverberation are based on multi-microphone measurements.

Spatio-temporal methods, which are directly applied to the received signals, have been presented by Liu *et al.* [1] and by Sánchez-Bote *et al.* [2]. They consist of a spatial averaging of the minimum-phase component of the speech signal and cepstrum domain processing for manipulating the all-pass component of the speech signal. Other methods use the linear prediction residual signal to dereverberate the speech signal [3],[4].

Beamforming methods [5],[6] which use an estimate of the related ATFs can reduce the amount of reverberation, especially if some a priori knowledge of the acoustical transfer is given. The average ATFs of all the microphones proves to be efficient and quite robust to small speaker movements. However, if this information is not available, these methods can not eliminate the reverberation completely. Hence, we will avoid using the small movement assumption in this work, as it is not valid in many important applications.

Subspace methods appear to be the most promising methods for dereverberation. These methods consist of estimating the null subspace of the data matrix. These null subspace vectors are used to extract the ATFs (e.g. [7] and [8]). Of special interest is the EVAM algorithm presented by Gürelli and Nikias [9]. As the null subspace vectors are shown to be filtered versions of the actual ATFs, extraneous zeros should be eliminated. This is done by the “fractal” method which is essentially a

recursive method for successively eliminating these zeros, yielding the correct filters.

The methods presented in this contribution are also based on null subspace estimation. The special Sylvester structure of the filtering matrix is taken into account to derive some algorithms.

The general dereverberation problem is presented in Section II. The proposed method is outlined in III. We start by deriving a method for constructing the null subspace in the presence of colored noise. Then, the special structure of the filtering matrix is exploited to derive a TLS approach for *acoustical transfer function* (ATF) estimation. Suboptimal procedures, based on the QRD, are derived in Section IV. The use of decimated subbands for reducing the complexity of the algorithm and increasing its robustness, is explored in Section V. A reconstruction procedure, based on the ATFs' matched filter and incorporated into an extension of the *generalized sidelobe canceller* (GSC) is proposed in Section VI. The derivation of the algorithms is followed by an experimental study presented in Section VII.

II. PROBLEM FORMULATION

Assume a speech signal is received by M microphones in a noisy and reverberating environment. The microphones receive a speech signal which is subject to propagation through a set of ATFs and contaminated by additive noise. The M received signals are given by,

$$z_m(t) = y_m(t) + n_m(t) = a_m(t) * s(t) + n_m(t) = \sum_{k=0}^{n_a} a_m(k)s(t-k) + n_m(t) \quad (1)$$

where $m = 1, \dots, M$; $t = 0, 1, \dots, T$. $z_m(t)$ is the m -th received signal, $y_m(t)$ is the corresponding desired signal part, $n_m(t)$ is the noise signal received in the m -th microphone, $s(t)$ is the desired speech signal and $T + 1$ is the number of samples observed. The convolution operation is denoted by $*$. We further assume that the ATFs relating the speech source and each of the M microphones can be modelled as an FIR filter of order n_a , with taps given by

$$\mathbf{a}_m^T = [a_m(0), a_m(1), \dots, a_m(n_a)]; \quad m = 1, 2, \dots, M.$$

Define also the Z -transform of each of the M filters as,

$$A_m(z) = \sum_{k=0}^{n_a} a_m(k)z^{-k}; \quad m = 1, 2, \dots, M.$$

All the involved signals and ATFs are depicted in Fig. 1. The goal of the dereverberation problem

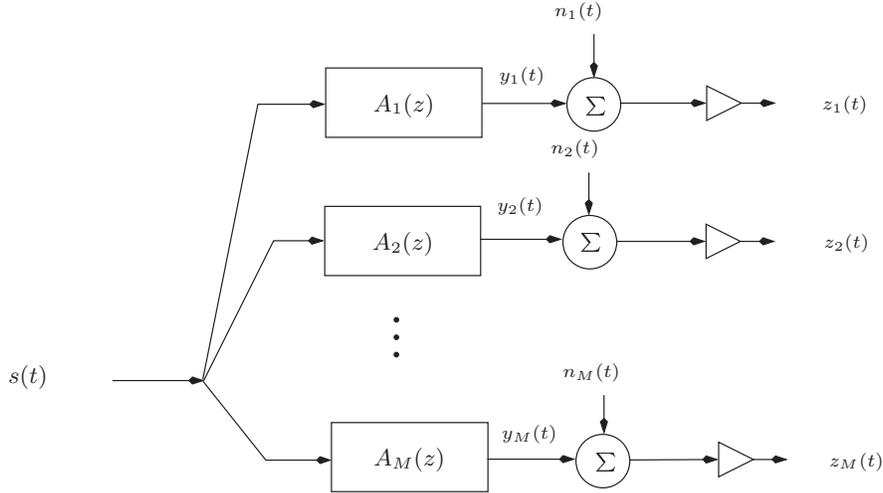


Fig. 1. The general dereverberation problem.

is to reconstruct the speech signal $s(t)$ from the noisy observations $z_m(t)$, $m = 1, 2, \dots, M$. In this contribution we will try to achieve this goal by first estimating the ATFs, \mathbf{a}_m , followed by a signal reconstruction scheme based on these ATFs estimate. Schematically, an “ATF Estimation” procedure, depicted in Fig. 2 is searched for.

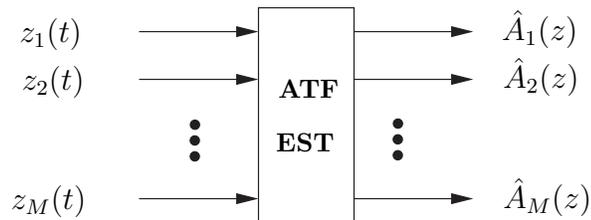


Fig. 2. ATFs’ estimation.

III. ATF ESTIMATION - ALGORITHM DERIVATION

In this section the proposed algorithm is derived in several stages. First, it is shown that the desired ATFs are embedded in a data matrix null subspace. Then, the special structure of the null subspace is exploited to derive several estimation methods. We start our discussion with the special case of the problem, namely, the *two microphones noiseless* case. We proceed through the *two microphones contaminated by colored noise* case. We end with the general *multi-microphone colored noise* case.

A. Two Microphone Noiseless Case - Preliminaries

In this section we lay the foundations of the algorithm by showing that the desired ATFs are embedded in the null subspace of a signal data matrix. This proof is merely a repetition of previously established results (e.g. [9]), but in a more intuitive way of presentation.

The two microphone noiseless case is depicted in Fig. 3. The noiseless signals, $y_m(t)$, are given in Eq. 2, as can be seen from the left-hand side of the figure.

$$\begin{aligned} y_1(t) &= a_1(t) * s(t) \\ y_2(t) &= a_2(t) * s(t). \end{aligned} \quad (2)$$

Clearly, as depicted in the right-hand side of Fig. 3, the identity in Eq. 3 holds.

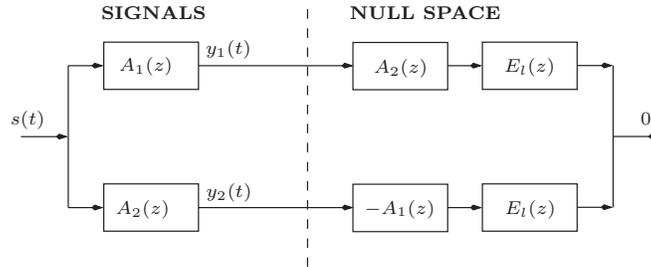


Fig. 3. Null subspace in the two microphone noiseless case.

$$[y_2(t) * a_1(t) - y_1(t) * a_2(t)] * e_l(t) = 0 \quad (3)$$

where, $e_l(t), l = 0, 1, 2, \dots$ are arbitrary and unknown filters, the number of which and their order will be discussed in the sequel. It is evident that filtered version of the desired ATFs, subject to the constraint that the arbitrary filters, $e_l(t)$ are common to all the microphone, might result in zero output. This observation was previously shown in [7],[9],[8].

Define the $(\hat{n}_a + 1) \times (T + \hat{n}_a + 1)$ single channel data matrix \mathcal{Y}_m , given in Eq. 4. Note, that as the ATFs order, n_a , is unknown, we use instead an (over-) estimated value, \hat{n}_a . An estimate of the

correct order would be a product of the proposed algorithm.

$$\mathcal{Y}_m = \begin{bmatrix} y_m(0) & y_m(1) & \cdots & y_m(\hat{n}_a) & y_m(\hat{n}_a + 1) & \cdots & y_m(T) & 0 & \cdots & 0 \\ 0 & y_m(0) & y_m(1) & \cdots & \vdots & \vdots & \cdots & y_m(T) & 0 & 0 \\ \vdots & 0 & \ddots & \ddots & & & & & \ddots & \vdots \\ 0 & & & \ddots & \vdots & & \ddots & 0 & \ddots & \\ 0 & & \cdots & 0 & y_m(0) & y_m(1) & \cdots & y_m(\hat{n}_a) & \cdots & y_m(T) \end{bmatrix} \quad (4)$$

We assume that the inequality $\hat{n}_a \geq n_a$ holds, i.e., the ATFs order is always overestimated. Define also the two-channel data matrix,

$$\mathcal{Y} = \begin{bmatrix} \mathcal{Y}_2 \\ -\mathcal{Y}_1 \end{bmatrix}.$$

The $2(\hat{n}_a + 1) \times 2(\hat{n}_a + 1)$ correlation matrix of the data is thus given by $\hat{R}_y = \frac{\mathcal{Y}\mathcal{Y}^T}{T+1}$.

Now, following [9] and [7], the null subspace of the correlation matrix can be calculated by virtue of the eigenvalue decomposition. Let λ_l ; $l = 0, 1, \dots, 2\hat{n}_a + 1$ be the eigenvalues of the correlation matrix \hat{R}_y , then by sorting them in ascending order we have,

$$\begin{aligned} \lambda_l &= 0 \quad l = 0, 1, \dots, \hat{n}_a - n_a \\ \lambda_l &> 0 \quad \text{otherwise} \end{aligned} \quad (5)$$

Thus, as proven by Gürelli and Nikias [9], the rank of the null subspace of the correlation matrix is $\hat{n}_a - n_a + 1$. This rank is useful for determining the correct ATFs order, n_a . We note that the singular value decomposition (SVD) of the data matrix, \mathcal{Y} , might be used instead of the eigenvalue decomposition for determining the null subspace. The SVD is generally regarded as a more robust method.

Denote the null subspace vectors (eigenvectors corresponding to zero eigenvalues or singular values) by \mathbf{g}_l for $l = 0, 1, 2, \dots, \hat{n}_a - n_a + 1$. Then, splitting each null subspace vector into two parts of equal length $\hat{n}_a + 1$ we obtain,

$$\mathcal{G} = \begin{bmatrix} \mathbf{g}_0 & \mathbf{g}_1 & \cdots & \mathbf{g}_{\hat{n}_a - n_a} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{a}}_{1,0} & \tilde{\mathbf{a}}_{1,1} & \cdots & \tilde{\mathbf{a}}_{1,\hat{n}_a - n_a} \\ \tilde{\mathbf{a}}_{2,0} & \tilde{\mathbf{a}}_{2,1} & \cdots & \tilde{\mathbf{a}}_{2,\hat{n}_a - n_a} \end{bmatrix}.$$

Each of the vectors $\tilde{\mathbf{a}}_{m,l}$ represents a null subspace filter of order \hat{n}_a .

$$\tilde{A}_{ml}(z) = \sum_{k=0}^{\hat{n}_a} \tilde{a}_{ml}(k)z^{-k}; \quad l = 0, 1, \dots, \hat{n}_a - n_a, \quad m = 1, 2.$$

From the above discussion, these null subspace filters may be presented in the following product,

$$\tilde{A}_{ml}(z) = A_m(z)E_l(z); \quad l = 0, 1, \dots, \hat{n}_a - n_a, \quad m = 1, 2. \quad (6)$$

Thus, the zeros of the filters $\tilde{A}_{ml}(z)$ extracted from the null subspace of the data, contain the roots of the desired filters as well as some extraneous zeros. This observation was proven by Gürelli and Nikias [9] as the basis of their EVAM algorithm. It can be stated in the following lemma (for the general M channel case):

Lemma 1: Let $\tilde{\mathbf{a}}_{ml}$ be the partitions of the null subspace eigenvectors into M vectors of length $\hat{n}_a + 1$, with $\tilde{A}_{ml}(z)$ their equivalent filters. Then, all the filters $\tilde{A}_{ml}(z)$ for $l = 0, \dots, \hat{n}_a - n_a$ have n_a common roots, which constitute the desired ATF $A_m(z)$, and $\hat{n}_a - n_a$ different extraneous roots, which constitute $E_l(z)$. These extraneous roots are common for all partitions of the same vector, i.e., $\tilde{A}_{ml}(z)$ for $m = 1, \dots, M$. ■

Under several regularity conditions (stated, for example by Moulines *et al.* [7]), the filters $A_m(z)$ can be found. Of special interest is the observation that common roots of the filters $A_m(z)$ can not be extracted by the algorithm, as they are treated as the extraneous roots which constitute $E_l(z)$. Although a drawback of the method, we will take benefit of it, while constructing a subband structure in Section V.

In matrix form, Eq. 6 may be written in the following manner. Define the $(\hat{n}_a + 1) \times (\hat{n}_a - n_a + 1)$ Sylvester filtering matrix (recall we assume $\hat{n}_a \geq n_a$),

$$\mathcal{A}_m = \underbrace{\begin{bmatrix} a_m(0) & 0 & 0 & \cdots & 0 \\ a_m(1) & a_m(0) & 0 & \cdots & 0 \\ \vdots & a_m(1) & \ddots & & \vdots \\ a_m(n_a) & \vdots & \ddots & \ddots & 0 \\ 0 & a_m(n_a) & & \ddots & a_m(0) \\ \vdots & 0 & & & a_m(1) \\ & \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & a_m(n_a) \end{bmatrix}}_{\hat{n}_a - n_a + 1}. \quad (7)$$

Then,

$$\tilde{\mathbf{a}}_{ml} = \mathcal{A}_m \mathbf{e}_l, \quad (8)$$

where, $\mathbf{e}_l^T = [e_l(0) \ e_l(1) \ \dots \ e_l(\hat{n}_a - n_a)]$ are vectors of the coefficients of the arbitrary unknown filters $E_l(z)$. Thus, the number of different filters (as shown in Eq. 6) is $\hat{n}_a - n_a + 1$ and their order is $\hat{n}_a - n_a$. Using Fig 3 and Eq. 3 and denoting,

$$\mathcal{E} = [\mathbf{e}_0 \ \mathbf{e}_1 \ \dots \ \mathbf{e}_{\hat{n}_a - n_a}],$$

we conclude

$$\mathcal{G} = \begin{bmatrix} \mathcal{A}_1 \\ \mathcal{A}_2 \end{bmatrix} \mathcal{E} \triangleq \mathcal{A}\mathcal{E}. \quad (9)$$

\mathcal{E} is an unknown $(\hat{n}_a - n_a + 1) \times (\hat{n}_a - n_a + 1)$ matrix. We note, that in the special case when the ATFs' order is known, i.e. $\hat{n}_a = n_a$, there is only one vector in the null subspace and its partitions $\tilde{\mathbf{a}}_{m0}$; $m = 1, \dots, M$ are equal to the desired filters \mathbf{a}_m up to a (common) scaling factor ambiguity. In the case where $\hat{n}_a > n_a$, the actual ATFs $A_m(z)$ are embedded in $\tilde{A}_{ml}(z)$; $l = 0, 1, \dots, \hat{n}_a - n_a$. The case $\hat{n}_a < n_a$ could not be treated properly by the proposed method.

The special structure depicted in Eq. 9 and Eq. 7 forms the basis of our suggested algorithm.

B. Two Microphone Noiseless Case - Algorithm

Based on the special structure of Eq. 9 and in particular on the Sylvester structure of \mathcal{A}_1 and \mathcal{A}_2 , found in Eq. 7, we derive now an algorithm for finding the ATFs $A_m(z)$.

Note that \mathcal{E} in Eq. 9 is a square and arbitrary matrix, implying that its inverse usually exists. Denote this inverse by $\mathcal{E}^i = \text{inv}(\mathcal{E})$. Then,

$$\mathcal{G}\mathcal{E}^i = \mathcal{A} \quad (10)$$

Denote the columns of \mathcal{E}^i by, $\mathcal{E}^i = [\mathbf{e}_0^i \ \mathbf{e}_1^i \ \dots \ \mathbf{e}_{\hat{n}_a - n_a}^i]$. Eq. 10 can be then rewritten as,

$$\tilde{\mathcal{G}}\mathbf{x} = \mathbf{0} \quad (11)$$

where, $\tilde{\mathcal{G}}$ is defined as,

$$\tilde{\mathcal{G}} = \begin{bmatrix} \mathcal{G} & \mathcal{O} & \dots & \dots & \dots & \mathcal{O} & -\mathcal{I}^{(0)} \\ \mathcal{O} & \mathcal{G} & \mathcal{O} & \dots & \dots & \mathcal{O} & -\mathcal{I}^{(1)} \\ \vdots & \mathcal{O} & \ddots & & & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & \mathcal{O} & \vdots \\ \mathcal{O} & \mathcal{O} & \dots & \dots & \mathcal{O} & \mathcal{G} & -\mathcal{I}^{\hat{n}_a - n_a} \end{bmatrix} \quad (12)$$

and the vector of unknowns is defined as,

$$\mathbf{x}^T = \left[\mathbf{e}_0^{iT} \quad \mathbf{e}_1^{iT} \quad \dots \quad \mathbf{e}_{\hat{n}_a - n_a}^{iT} \quad \mathbf{a}_1^T \quad \mathbf{a}_2^T \right]$$

where $\mathbf{0}$ is a vector of zeros: $\mathbf{0}^T = [0 \ 0 \ \dots \ 0]$. We used the following expressions: \mathcal{O} is a $2(\hat{n}_a + 1) \times (\hat{n}_a - n_a + 1)$ all-zeros matrix and $\mathcal{I}^{(l)}$; $l = 0, 1, \dots, \hat{n}_a - n_a$ is a fixed shifting matrix given by,

$$\mathcal{I}^{(l)} = \left[\begin{array}{c|c} \mathcal{O}_{l \times (n_a + 1)} & \\ \hline I_{(n_a + 1) \times (n_a + 1)} & \mathcal{O}_{(\hat{n}_a + 1) \times (n_a + 1)} \\ \mathcal{O}_{(\hat{n}_a - n_a - l) \times (n_a + 1)} & \\ \hline & \mathcal{O}_{l \times (n_a + 1)} \\ \mathcal{O}_{(\hat{n}_a + 1) \times (n_a + 1)} & I_{(n_a + 1) \times (n_a + 1)} \\ & \mathcal{O}_{(\hat{n}_a - n_a - l) \times (n_a + 1)} \end{array} \right].$$

$I_{(n_a + 1) \times (n_a + 1)}$ is the $(n_a + 1) \times (n_a + 1)$ Identity matrix. A non-trivial (and exact) solution for the homogenous set of equations (11) may be obtained by finding the eigenvector of the matrix $\tilde{\mathcal{G}}$ corresponding to its zero eigenvalue. The ATF coefficients are given by the last $2(n_a + 1)$ terms of this eigenvector. The beginning of the eigenvector contains the nuisance parameters \mathbf{e}_l^i ; $l = 0, 1, \dots, \hat{n}_a - n_a$. In the presence of noise, the somewhat non-straight-forward procedure will prove to be useful.

C. Two Microphone Noisy Case

Recall that \mathcal{G} is a matrix containing the eigenvectors corresponding to zero eigenvalues of the noiseless data matrix. In the presence of additive noise, the noisy observations $z_m(t)$, given in Eq. 1, can be stacked into a data matrix fulfilling

$$\mathcal{Z} = \mathcal{Y} + \mathcal{N},$$

where, \mathcal{Z} and \mathcal{N} are noisy signal and noise-only signal data matrices, respectively, given by 13 and 14.

$$\mathcal{Z}_m = \begin{bmatrix} z_m(0) & z_m(1) & \cdots & z_m(\hat{n}_a) & z_m(\hat{n}_a + 1) & \cdots & z_m(T) & 0 & \cdots & 0 \\ 0 & z_m(0) & z_m(1) & \cdots & \vdots & \vdots & \cdots & z_m(T) & 0 & 0 \\ \vdots & 0 & \ddots & \ddots & & & & & \ddots & \vdots \\ 0 & & & \ddots & \vdots & & \ddots & 0 & & \ddots \\ 0 & & \cdots & 0 & z_m(0) & z_m(1) & \cdots & z_m(\hat{n}_a) & \cdots & z_m(T) \end{bmatrix} \quad (13)$$

and,

$$\mathcal{N}_m = \begin{bmatrix} n_m(0) & n_m(1) & \cdots & n_m(\hat{n}_a) & n_m(\hat{n}_a + 1) & \cdots & n_m(T) & 0 & \cdots & 0 \\ 0 & n_m(0) & n_m(1) & \cdots & \vdots & \vdots & \cdots & n_m(T) & 0 & 0 \\ \vdots & 0 & \ddots & \ddots & & & & & \ddots & \vdots \\ 0 & & & \ddots & \vdots & & \ddots & 0 & & \ddots \\ 0 & & \cdots & 0 & n_m(0) & n_m(1) & \cdots & n_m(\hat{n}_a) & \cdots & n_m(T) \end{bmatrix}. \quad (14)$$

Now, for a long observation time the following approximation holds,

$$\hat{R}_z \approx \hat{R}_y + \hat{R}_n$$

where, $\hat{R}_z = \frac{\mathcal{Z}\mathcal{Z}^T}{T+1}$ and $\hat{R}_n = \frac{\mathcal{N}\mathcal{N}^T}{T+1}$ are the noisy signal and noise-only signal correlation matrices, respectively. Now, Eq. 11 will not be accurate anymore. First, the null subspace matrix \mathcal{G} should be determined in a slightly different manner than suggested in Eq. 5. The white noise and colored noises cases are treated separately in the sequel. Second, the matrix $\tilde{\mathcal{G}}$ will in general not have an eigenvalue of value 0. A reasonable approximation for the solution, although not exact, would be to transform Eq. 11 into the following problem,

$$\tilde{\mathcal{G}}\mathbf{x} = \boldsymbol{\mu}. \quad (15)$$

where $\boldsymbol{\mu}$ is an error term, which should be minimized. To obtain this minimization, the eigenvector corresponding to the smallest eigenvalue is chosen, and the desired ATFs are obtained from the last part of the vector (as in the noiseless case). Note, that this is exactly the *total least squares* (TLS) approach for estimating the parameters. As the matrix $\tilde{\mathcal{G}}$ is highly structured, more efficient *structured total least squares* (STLS) methods [10] are called for. This issue will not be treated in this work anymore.

C.1 White Noise Case

In the case of spatio-temporally white noise - i.e. $\hat{R}_n \approx \sigma^2 I$, where I is the identity matrix - the first $\hat{n}_a - n_a + 1$ eigenvalues in Eq. 5 will be σ^2 instead of zero. The corresponding eigenvectors will remain intact. Thus, the algorithm remains unchanged.

C.2 Colored Noise Case

The case of non-white noise signal was addressed in [7],[9]. In contrast to the noise balancing method presented in [9] and the pre-whitening of the noise correlation matrix, presented in [7], the problem is treated here more rigourously, with the application of the *generalized eigenvalue decomposition* (GEVD) or *generalized singular value decomposition* (GSVD) techniques. These alternative methods are computationally more efficient. We suggest to use the (GEVD) of the measurement correlation matrix, R_z and the noise correlation matrix R_n (usually, the latter is estimated from speech-free data segments). The null subspace matrix \mathcal{G} is formed by choosing the generalized eigenvectors related to the generalized eigenvalues of value 1. Alternatively, the *generalized singular value decomposition* (GSVD) of the corresponding data matrices, \mathcal{Z} and \mathcal{N} , can be used. After determining the null subspace matrix, subsequent steps of the algorithm remain intact.

D. Multi Microphone Case ($M > 2$)

In the multi microphone case a reasonable extension would be based on channel pairing (see [9]). Each of the $\frac{M \times (M-1)}{2}$ pairs fulfills 16,

$$\begin{aligned}
 [y_i(t) * a_j(t) - y_j(t) * a_i(t)] * e_l(t) &= 0 \\
 i, j &= 1, 2, \dots, M; \quad l = 0, 1, \dots, \hat{n}_a - n_a.
 \end{aligned}
 \tag{16}$$

Thus, the new data matrix would be constructed as follows,

$$\mathcal{Z} = \begin{bmatrix} \mathcal{Z}_2 & \mathcal{Z}_3 & \cdots & \mathcal{Z}_M & \mathcal{O} & \cdots & \mathcal{O} & \cdots & \mathcal{O} \\ -\mathcal{Z}_1 & \mathcal{O} & \cdots & & \mathcal{Z}_3 & \cdots & \mathcal{Z}_M & & \mathcal{O} \\ \mathcal{O} & -\mathcal{Z}_1 & & & -\mathcal{Z}_2 & & \mathcal{O} & & \vdots \\ \vdots & \mathcal{O} & \ddots & & & & \vdots & & \mathcal{O} \\ & \vdots & & \ddots & & & \mathcal{O} & & \mathcal{Z}_M \\ \mathcal{O} & \mathcal{O} & \cdots & -\mathcal{Z}_1 & \cdots & & -\mathcal{Z}_2 & \cdots & -\mathcal{Z}_{M-1} \end{bmatrix} \quad (17)$$

where \mathcal{O} here is an $(\hat{n}_a + 1) \times (T + \hat{n}_a + 1)$ all-zero matrix. This data matrix, as well as the corresponding noise matrix can be used by either the GEVD or the GSVD methods to construct the null subspace. Denoting this null subspace by $\underline{\mathcal{G}}$, we can construct a new TLS equation.

$$\underline{\tilde{\mathcal{G}}}\mathbf{x} = \boldsymbol{\mu}$$

where, $\underline{\tilde{\mathcal{G}}}$ is constructed in a similar way $\tilde{\mathcal{G}}$ was constructed in Eq. 12. The vector of unknowns \mathbf{x} is given by,

$$\mathbf{x}^T = \left[\mathbf{e}_0^{iT} \quad \mathbf{e}_1^{iT} \quad \cdots \quad \mathbf{e}_{\hat{n}_a - n_a}^{iT} \quad \mathbf{a}_1^T \quad \mathbf{a}_2^T \quad \cdots \quad \mathbf{a}_M^T \right]$$

Note, that the last $M \times (n_a + 1)$ terms of \mathbf{x} are the required filter coefficients, \mathbf{a}_m ; $m = 1, 2, \dots, M$.

E. Partial Knowledge Of The Null Subspace

In the noisy case, especially when the dynamic range of the input signal $s(t)$ is high (which is the case for speech signals), determination of the null subspace might be a troublesome task. As there are no zero eigenvalues and as some of the eigenvalues are small due to the input signal, the borderline between the signal eigenvalues and the noise eigenvalues becomes vague. As the number of actual null subspace vectors is not known in advance, using only a subgroup of the eigenvectors, which are associated with the smallest eigenvalues, might increase the robustness of the method. Based on Lemma 1, it is obvious that, in the noiseless case, even two null subspace vectors are sufficient to estimate the ATFs, just by extracting their common zeros. Denote by $\bar{L} < \hat{n}_a - n_a$ the number of eigenvectors used. The matrix \mathcal{E} in Eq. 9 is than of dimensions $(\hat{n}_a - n_a + 1) \times \bar{L}$ and thus non-invertible. To overcome this problem we suggest to concatenate several shifted versions of Eq. 9 in a manner depicted in Eq. 18.

$$\bar{\mathcal{G}} = \begin{bmatrix} \mathcal{G} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{G} & \mathbf{0} & \mathbf{0} \\ \vdots & & \ddots & \mathbf{0} \\ & & & \ddots \\ \mathbf{0} & & & \mathcal{G} \end{bmatrix} = \bar{\mathcal{A}} \underbrace{\begin{bmatrix} \mathcal{E} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{E} & \mathbf{0} & \mathbf{0} \\ \vdots & & \ddots & \mathbf{0} \\ & & & \ddots \\ \mathbf{0} & & & \mathcal{E} \end{bmatrix}}_{L > \hat{n}_a - n_a + \hat{l}} = \bar{\mathcal{A}}\bar{\mathcal{E}} \quad (18)$$

The new dimensions of $\bar{\mathcal{E}}$ is $L \times (\hat{n}_a - n_a + \hat{l})$, where \hat{l} is the number of blocks added. Each block adds 1 to the row dimension and \bar{L} to the columns dimension.

The matrix $\bar{\mathcal{A}}$ has a similar structure as \mathcal{A} in Eqs. 7 and 9 but with more columns. The resulting matrix $\bar{\mathcal{E}}$ has now more columns than rows and thus can generally be pseudo-inverted

$$\mathcal{E}^{Pi} = \text{Pinv}(\bar{\mathcal{E}}) = \bar{\mathcal{E}}^T (\bar{\mathcal{E}}\bar{\mathcal{E}}^T)^{-1} \quad (19)$$

resulting into,

$$\bar{\mathcal{G}}\mathcal{E}^{Pi} = \bar{\mathcal{A}} \quad (20)$$

Now the extended matrix $\bar{\mathcal{G}}$ can be used in Eq. 15, instead of \mathcal{G} to construct $\tilde{\mathcal{G}}$ in a similar manner to Eq. 12. Subsequent stages of the algorithm remain intact.

IV. A SUBOPTIMAL METHOD - THE QR DECOMPOSITION AND ESTIMATES AVERAGING

Recall that the special structure of the filtering matrix \mathcal{A} was the basis for the TLS approach. In this section a new method is derived for the estimate of the ATFs, which is computationally more efficient although less robust. We rely again on the fact that each column of the Sylvester matrix is a delayed version of the previous one. Thus, in the noiseless case, it is enough to extract one of the columns. In the noisy case, each column may be different. Thus extracting all the columns might give several slightly different estimates. We can take the median (or average) of these estimates to increase the robustness.

A. Complete Knowledge Of The Null Subspace

Apply the transpose operation to Eq. 9:

$$\mathcal{G}^T = \mathcal{E}^T \mathcal{A}^T. \quad (21)$$

As \mathcal{E}^T is an arbitrary matrix it will usually have a QRD (11). Denote, $\mathcal{E}^T = \mathcal{Q}_\mathcal{E}\mathcal{R}_\mathcal{E}$. Then,

$$\mathcal{G}^T = \mathcal{Q}_\mathcal{E}\mathcal{R}_\mathcal{E}\mathcal{A}^T = \mathcal{Q}_\mathcal{E}\mathcal{R}_\mathcal{G} \quad (22)$$

where, $R_\mathcal{G} = R_\mathcal{E}\mathcal{A}^T$ is also an upper triangular matrix, since it consists of a multiplication of two upper triangular matrices. Since the QRD is unique Eq. 22 constitutes the QRD of \mathcal{G}^T . As $R_\mathcal{E}$ is a square and upper triangular matrix it has only 1 non-zero element in its last row. Therefore, the last row of $R_\mathcal{G}$ will be a scaled version of the last column of \mathcal{A} . This last column consists of a concatenation of the vectors \mathbf{a}_m , $m = 1, 2, \dots, M$ each preceded by $\hat{n}_a - n_a$ zeros.

For extracting the other columns of the matrix \mathcal{A} , we use rotations of the null subspace matrix \mathcal{G} . Note that the previous procedure will extract the last column of \mathcal{A} regardless of the its Silvester structure. Define, the $K \times K$ row rotation matrix,

$$J_K = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 1 & 0 & \cdots & & 0 \\ 0 & \ddots & 0 & \cdots & 0 \\ \vdots & 0 & \ddots & & \vdots \\ 0 & \cdots & & 1 & 0 \end{bmatrix}.$$

It is obvious that left multiplication of a K -row matrix by J_K^k will rotate its rows downwards k times, while right multiplication of a L -columns matrix by $(J_L^l)^T$ will rotate its columns rightwards l times. Lemma 2 can now be used to extract an estimate of the ATFs.

Lemma 2: Compute the QR decomposition of the transpose of the k -times ($k \leq \hat{n}_a - n_a + 1$) row rotated null subspace matrix \mathcal{G} . The last row of the “ R ” matrix equals the last but k column of the filtering matrix \mathcal{A} up to a scaling factor. ■

The proof of this lemma follows.

Proof: Rotate the $M(\hat{n}_a + 1) \times (\hat{n}_a - n_a + 1)$ null subspace matrix \mathcal{G} not more than $\hat{n}_a - n_a + 1$ times. Then,

$$\mathcal{G}^R = J_{M(\hat{n}_a+1)}^k \mathcal{G} = J_{M(\hat{n}_a+1)}^k \mathcal{A} \mathcal{E}.$$

Exploiting the orthogonality of the matrices J_K^k we have,

$$\mathcal{G}^R = J_{M(\hat{n}_a+1)}^k \mathcal{A} (J_{\hat{n}_a - n_a + 1}^k)^T J_{\hat{n}_a - n_a + 1}^k \mathcal{E}.$$

Then, applying the transpose operation

$$(\mathcal{G}^R)^T = (J_{\hat{n}_a - n_a + 1}^k \mathcal{E})^T (J_{M(\hat{n}_a + 1)}^k \mathcal{A} (J_{\hat{n}_a - n_a + 1}^k)^T)^T$$

Now assume a QRD for the first term (although, \mathcal{E} is not known),

$$(J_{\hat{n}_a - n_a + 1}^k \mathcal{E})^T = QR.$$

Then,

$$(\mathcal{G}^R)^T = QR (J_{M(\hat{n}_a + 1)}^k \mathcal{A} (J_{\hat{n}_a - n_a + 1}^k)^T)^T = Q\tilde{R}. \quad (23)$$

The last row of $(J_{M(\hat{n}_a + 1)}^k \mathcal{A} (J_{\hat{n}_a - n_a + 1}^k)^T)^T$ is the last but k row of \mathcal{A}^T , provided $k \leq \hat{n}_a - n_a + 1$ and it is still an upper triangular matrix. Thus, the same statements regarding the non-rotated matrices apply for the rotated matrices. ■

By rotating through all the columns of matrix \mathcal{A} several estimates of the desired filter are obtained. An average or a median of these estimated can be used to obtain a more robust estimate.

B. Partial Knowledge Of The Null Subspace

As in the TLS approach we may want to use only part of the null subspace vectors. Assume that we have only two of these null subspace vectors.

$$\check{\mathcal{G}} = \mathcal{A}\check{\mathcal{E}}$$

where, $\check{\mathcal{G}}$ is an $M(\hat{n}_a + 1) \times 2$ matrix and $\check{\mathcal{E}}$ is an $(\hat{n}_a - n_a + 1) \times 2$ matrix. Since $\check{\mathcal{E}}$ is not a square matrix the algorithm of Section IV-A is not applicable anymore.

Let,

$$\check{\mathcal{G}}^T = \begin{bmatrix} (\tilde{\mathbf{a}}_{1,0})^T & (\tilde{\mathbf{a}}_{2,0})^T & \cdots & (\tilde{\mathbf{a}}_{M,0})^T \\ (\tilde{\mathbf{a}}_{1,1})^T & (\tilde{\mathbf{a}}_{2,1})^T & \cdots & (\tilde{\mathbf{a}}_{M,1})^T \end{bmatrix}. \quad (24)$$

Each of the vectors $\tilde{\mathbf{a}}_{m,l}$ represents a null subspace filter of order \hat{n}_a . Since there are only two rows, applying the QRD to $\check{\mathcal{G}}^T$, will yield the following $R_{\check{\mathcal{G}}}$ matrix,

$$R_{\check{\mathcal{G}}} = \begin{bmatrix} \cdots & \cdots \\ [0 \ (\tilde{\mathbf{a}}'_{1,1})^T] & [0 \ (\tilde{\mathbf{a}}'_{2,1})^T] & \cdots & [0 \ (\tilde{\mathbf{a}}'_{M,1})^T] \end{bmatrix}.$$

Note, that now $\tilde{\mathbf{a}}'_{m,1}$ relate to filters that have an order which is lower than their corresponding filters $\tilde{\mathbf{a}}_{m,1}$ by 1. As the first row $R_{\check{\mathcal{G}}}$ is not important, it is not presented. To further reduce the

order by virtue of another QRD application, we need another set of filtered version of the ATFs. These set may be obtained in several ways. One possibility (although others are also applicable) is to rotate each part of $\check{\mathcal{G}}$, i.e. $\tilde{\mathbf{a}}_{m,l}$, downwards and apply the QRD again. After this two steps stage we obtain a shorter null subspace,

$$\check{\mathcal{G}}'^T = \begin{bmatrix} (\tilde{\mathbf{a}}'_{1,0})^T & (\tilde{\mathbf{a}}'_{2,0})^T & \cdots & (\tilde{\mathbf{a}}'_{M,0})^T \\ (\tilde{\mathbf{a}}'_{1,1})^T & (\tilde{\mathbf{a}}'_{2,1})^T & \cdots & (\tilde{\mathbf{a}}'_{M,1})^T \end{bmatrix}.$$

This process is repeated $\hat{n}_a - n_a$ times, until the correct order is reached and only a common scale factor ambiguity remains. This method has an appealing structure, since the extra roots are eliminated recursively, one in each stage of the algorithm. Each stage of the recursion is similar to the previous one. This property resembles the “fractal” nature of the EVAM algorithm [9].

V. SUBBAND METHOD

The proposed method although theoretically supported can have several drawbacks in real-life scenarios. First, actual ATFs in real room environments may be very long (1000–2000 taps are common in medium-sized room). In such a case the GEVD procedure is not robust enough and it is quite sensitive to small errors in the null subspace matrix [11]. Furthermore, the matrices involved become extremely large causing huge memory and computation requirements. Another problem is the speech signal wide dynamic range. This may result in erroneous estimates of the frequency response of the ATFs in the low energy parts of the input signal.

Thus, frequency domain approaches are called for. In this section we suggest to incorporate the TLS subspace method into a subband structure. The use of subbands for splitting adaptive filters, especially in the context of echo cancellation, has gained recent interest in the literature [12],[13],[14],[15]. However, the use of subbands in subspace methods is not as common. The design of the subbands is of crucial importance. Special emphasis should be given to adjusting the subband structure to the problem at hand. In this contribution we only aim at demonstrating the ability of the method, thus only a simple 8-channel subband structure was used as depicted in Fig. 4. Each of the channel filters is an FIR filter of order 150. The filters are equi-spaced along the frequency axis and are of equal bandwidth.

Now the M microphone signals are filtered by the subband structure. The subspace methods presented above can be applied on each subband signal separately. Although the resulting subband

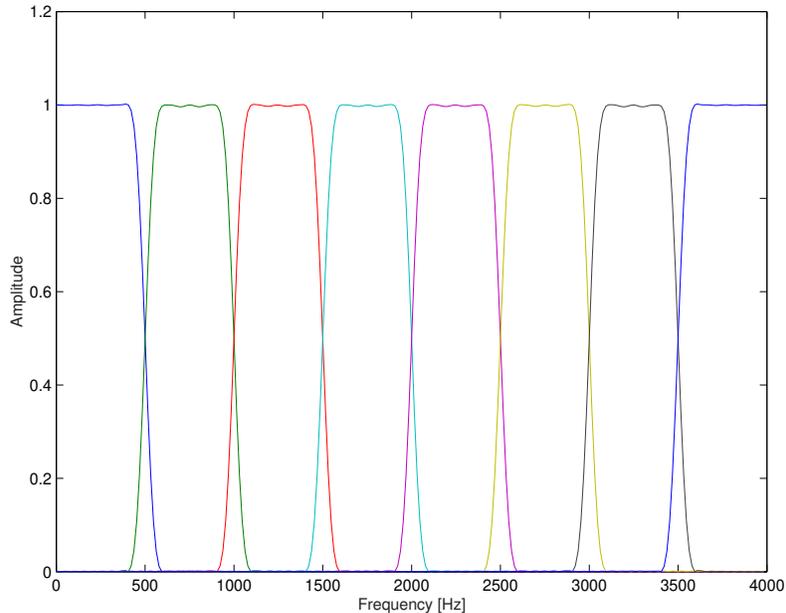


Fig. 4. Subband structure. 8 equi-spaced equi-bandwidth filters.

signal correspond to a longer filter (which is the convolution of the corresponding ATF and the subband filter), the algorithm is aimed to reconstruct the ATF alone, ignoring the filterbank roots. This is due to the fact that the subband filter is common for all channels. Recall that subspace methods are blind to common zeros, as discussed in Section III. For properly exploiting the benefits of the subband structure, each subband signal should be decimated. We took critically decimated filterbank, i.e. decimation factor equals the number of bands. By doing so, the ATF order in each band is reduced by the decimation factor, making the estimation task easier. Note that now we need only to over-estimate the reduced order of the ATFs in each subband, rather than the fullband order. Another benefit arises from the decimation. The signals in each subband are flatter in the frequency domain, making the signals processed to be whiter, and thus enabling lower dynamic range and resulting in improved performance. After estimating the decimated ATFs, they are combined together using a proper synthesis filterbank, comprised of interpolation followed by filtering with a filterbank similar to the analysis subband filters. The overall subband system is depicted in Fig. 5, where the “ATF Estimation” block is shown schematically in Fig. 2.

Gain ambiguity may be a major drawback of the subband method. Recall that all the subspace methods are estimating the ATFs only up to a common gain factor. In the fullband scheme this does

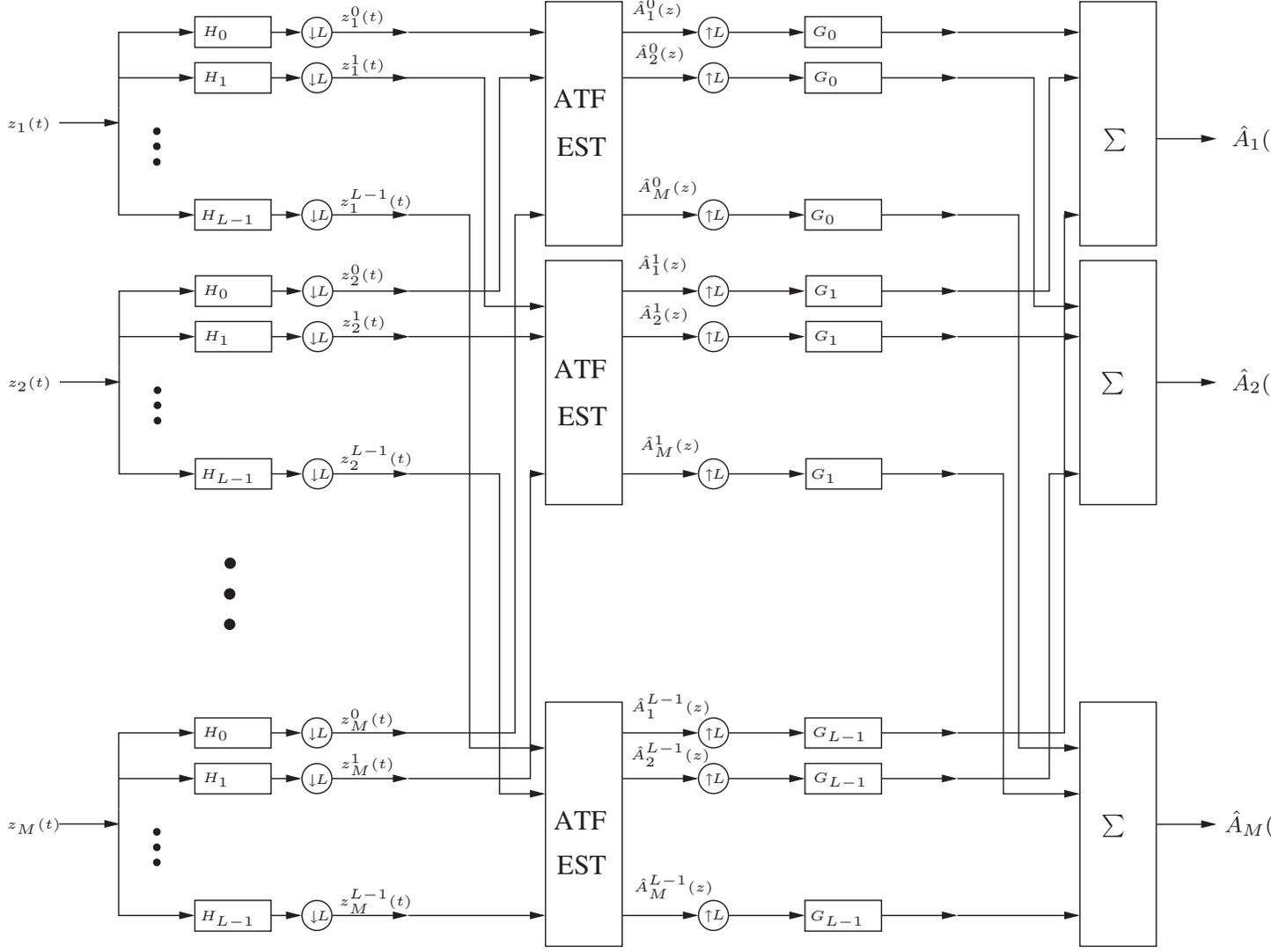


Fig. 5. Null subspace in the two microphone noiseless case.

not impose any problem since it results in overall scaling of the output. In the subband scheme, the gain factor is common for all subband signals but is generally different from band to band. Thus, the estimated ATFs (and the reconstructed signal) is actually filtered by a new arbitrary filter, which can be regarded as a new reverberation term. Several methods can be applied to overcome this gain ambiguity problem. First, the original gain of the signals in each subband may be restored an approximate gain adjustment. Another method was suggested by Rahbar *et al.* [16]. The method imposes the resulting filters to have as few taps as possible (actually, the filters are

constrained to be FIR). The order of these filters should be determined in advance. As we do not have this information we suggest to use the ATFs order estimation obtained by the subspace method. The use of this method is a topic of further research. In this contribution we will assume that the gain in each subband is known, and thus we would only demonstrate the ability of the method to estimate the frequency shaping of the method in each band.

VI. SIGNAL RECONSTRUCTION

Having the estimated ATFs, we can invert them and apply the inverse filter to the received signals to obtain the desired signal estimate. A method for inverting multi-channel FIR filters by a multi-channel set of FIR filters is presented in [17].

We use instead a frequency domain method. Rewrite Eq. 1 in time-frequency presentation, using the short time Fourier transform (STFT):

$$Z_m(t, e^{j\omega}) = A_m(t, e^{j\omega})S(t, e^{j\omega}) + N_m(t, e^{j\omega}).$$

Eliminating the reverberation term can be obtained by a matched filter beamformer:

$$\begin{aligned} \hat{S}(t, e^{j\omega}) &= \frac{1}{\sum_{m=1}^M |\hat{A}_m(t, e^{j\omega})|^2} \sum_{m=1}^M Z_m(t, e^{j\omega}) \hat{A}_m^*(t, e^{j\omega}) = \\ & S(t, e^{j\omega}) \frac{1}{\sum_{m=1}^M |\hat{A}_m(t, e^{j\omega})|^2} \sum_{m=1}^M A_m(t, e^{j\omega}) \hat{A}_m^*(t, e^{j\omega}) + \frac{1}{\sum_{m=1}^M |\hat{A}_m(t, e^{j\omega})|^2} \sum_{m=1}^M N_m(t, e^{j\omega}) \hat{A}_m^*(t, e^{j\omega}) \end{aligned} \quad (25)$$

It is easily verified that if the estimation of the ATFs is sufficiently accurate, i.e. $\hat{A}_m(t, e^{j\omega}) \simeq A_m(t, e^{j\omega})$, then the first term in Eq. 25 becomes $S(t, e^{j\omega})$ and dereverberation is obtained. The second term is a residual noise term, which can even be amplified by the procedure. To achieve a better estimation of the speech signal, when noise is present, we suggest to incorporate the procedure into the recently proposed extended GSC, derived by Gannot *et al.* [18], shown schematically in Fig. 6 and summarized in Fig. 7. This GSC based structure enables the use of general ATFs rather than delay-only filters in order to dereverberate the speech signal and to reduce the noise level. It consists of a fixed beamformer branch - which is essentially the MFBBF described in Eq. 25, a noise reference construction block - which uses the ATFs ratios (note that $U_m(t, e^{j\omega})$ contain only noise terms), and a multi-channel noise canceller branch (consisting of the filters $G_m(t, e^{j\omega})$). The use of the GSC structure is only essential when the noise level is relatively high, otherwise the MFBBF branch produces sufficiently accurate estimate.

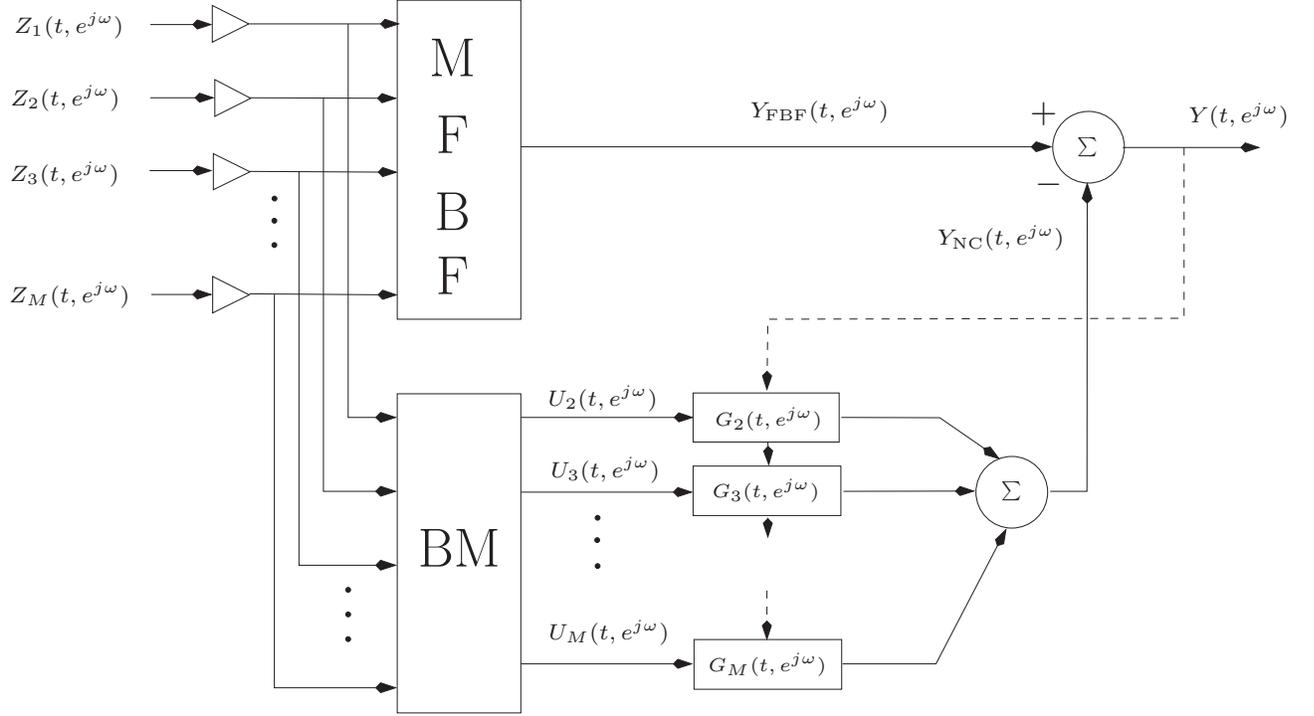


Fig. 6. Extended GSC structure for joint noise reduction and dereverberation.

1. **Estimate ATF-s:** $A_m(e^{j\omega}), m = 1, 2, \dots, M$.

$$\text{Define } \mathbf{A}(t, e^{j\omega}) = \begin{bmatrix} A_1(t, e^{j\omega}) & A_2(t, e^{j\omega}) & \dots & A_M(t, e^{j\omega}) \end{bmatrix}.$$

2. Fixed beamformer (FBF) $\mathbf{W}_0(t, e^{j\omega}) = \frac{\mathbf{A}(t, e^{j\omega})}{\|\mathbf{A}(t, e^{j\omega})\|^2}$.

$$\text{FBF output: } Y_{\text{FBF}}(t, e^{j\omega}) = \mathbf{W}_0^\dagger(e^{j\omega}) \mathbf{Z}(t, e^{j\omega}).$$

3. **Noise reference signals:**

$$U_m(t, e^{j\omega}) = A_1(e^{j\omega})Z_m(t, e^{j\omega}) - A_m(t, e^{j\omega})Z_1(t, e^{j\omega}); m = 2, \dots, M.$$

4. **Output signal:** $Y(t, e^{j\omega}) = Y_{\text{FBF}}(t, e^{j\omega}) - \mathbf{G}^\dagger(t, e^{j\omega}) \mathbf{U}(t, e^{j\omega})$.

5. **Filters update.** For $m = 1, \dots, M - 1$:

$$\tilde{G}_m(t+1, e^{j\omega}) = G_m(t, e^{j\omega}) + \mu \frac{U_m(t, e^{j\omega}) Y^*(t, e^{j\omega})}{P_{\text{est}}(t, e^{j\omega})}$$

$$G_m(t+1, e^{j\omega}) \stackrel{\text{FIR}}{\leftarrow} \tilde{G}_m(t+1, e^{j\omega})$$

where, $P_{\text{est}}(t, e^{j\omega}) = \rho P_{\text{est}}(t-1, e^{j\omega}) + (1-\rho) \sum_m |Z_m(t, e^{j\omega})|^2$.

6. Keep only non-aliased samples, according to the **overlap & save** method [19].

Fig. 7. Summary of the TF-GSC algorithm.

VII. EXPERIMENTAL STUDY

The validity of the proposed methods was tested using various input signals and randomly chosen FIR filters, and compared with the EVAM algorithm [9]. This input signal consisted of either white noise, or speech-like noise (white signal colored to have a speech-like spectrum, drawn from the NOISEX-92 database [20]), or a real speech signal comprised of a concatenation of several speech signals drawn from the TIMIT database [21]. The input signal was 32000 samples long (corresponding to 4sec for the 8KHz sampled speech signal, including silence periods). Three microphone signals were simulated, by randomly choosing the ATFs. The ATF order used was either 32 or 8. Various SNR levels were taken to test the robustness of the algorithms to additive noise. Temporally non-white but spatially white (i.e., no correlation between noise signals at the microphones) noise signals were used. The noise correlation matrix was estimated using signals drawn from the same source but at different segments. The basic fullband algorithm (using all the null subspace vectors or only part of them), as well as QRD based algorithms (again, with all the vectors or only part of them) were tested and compared with the state-of-the-art EVAM algorithm. Then, the subband based algorithm was evaluated to confirm its ability to comprehend with longer filters. The gain ambiguity problem was not addressed in this experimental study, and the gain levels in the various bands were assumed to be known a priori (see also [5]).

The frequency response of the real ATFs compared with the estimated ATFs for the fullband algorithm are depicted in Figures 8,9,10 for SNR levels of 45dB, 35dB and 25dB, respectively.

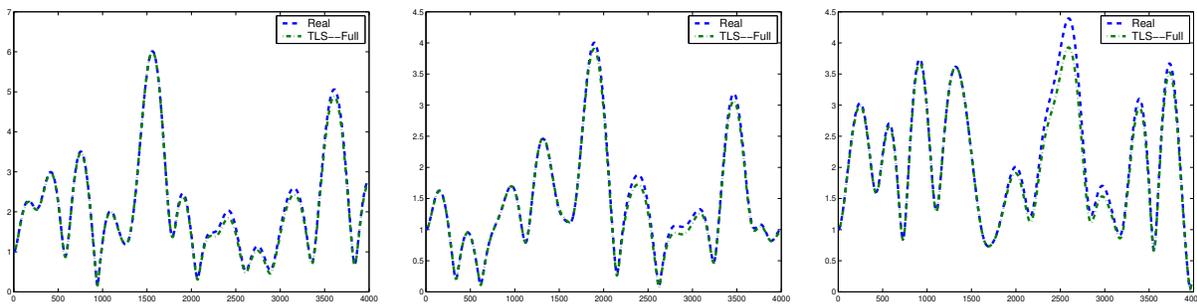


Fig. 8. Real and estimated frequency response of an arbitrary ATF of order 32 (frequency axis in Hz). Speech-like noise input. Fullband method. SNR=45dB.

The filter order was overestimated by 5, that is $\hat{n}_a - n_a = 5$ in all cases. While the estimation with speech-like noise (with a wide dynamic range) is quite good at the higher SNR level and filter order

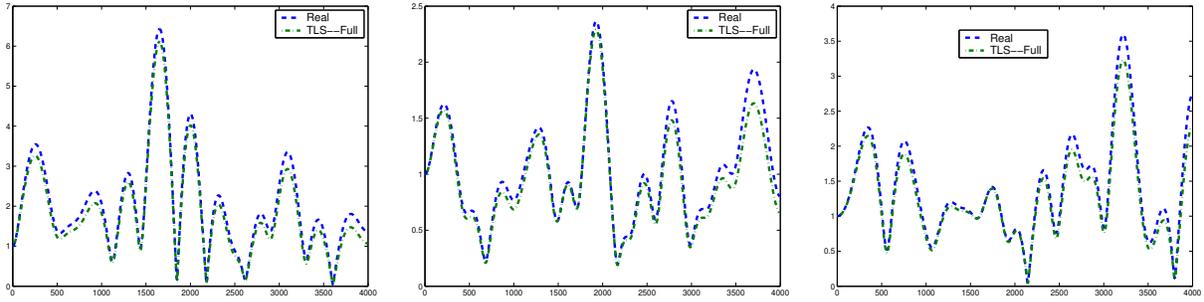


Fig. 9. Real and estimated frequency response of an arbitrary ATF of order 32 (frequency axis in Hz). White noise input. Fullband method. SNR=35dB.

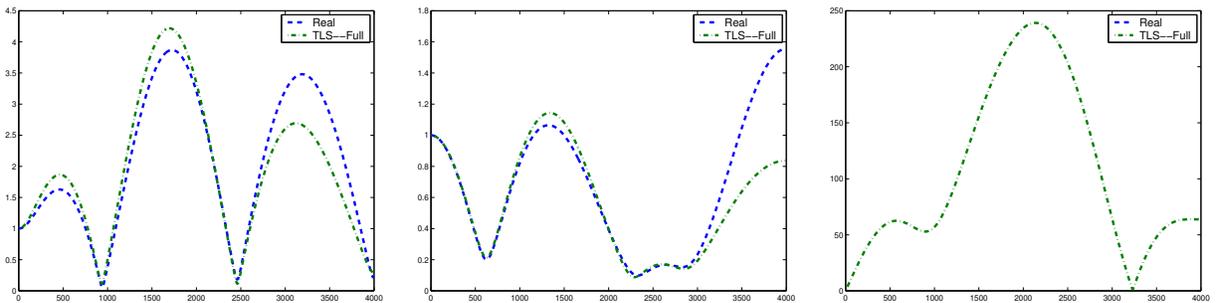


Fig. 10. Real and estimated frequency response of an arbitrary ATF of order 8 (frequency axis in Hz). White noise input. Fullband method. SNR=25dB.

32 (Fig. 8), when the SNR decreases to 35dB, the performance is maintained only with a white noise input signal (Fig. 9). For SNR level of 25dB, the algorithm fails to work with the longer filter and its order had to be reduced to only 8 (Fig. 10). The sensitivity to the noise level is thus clearly indicated.

Results for the suboptimal QRD based algorithm are depicted in Fig. 11 for the speech-like input and an SNR level of 45dB, and in Fig. 12 for a white noise input and an SNR level of 35dB. The QRD based method is more sensitive to the noise level. At 35dB good results could be obtained only with white noise input and filter order of 8.

In all cases, using only part of the null subspace vectors yielded reduced performance. Therefore, we omit results concerning these experiments from the presentation.

For comparison we used the EVAM algorithm, while successively reducing the overestimation of the filter order in their “fractal” based method as explained in [9]. Results for the speech-like input are depicted in Figures 13 and 14 for SNR levels of 45dB and 25dB respectively. It is clearly shown

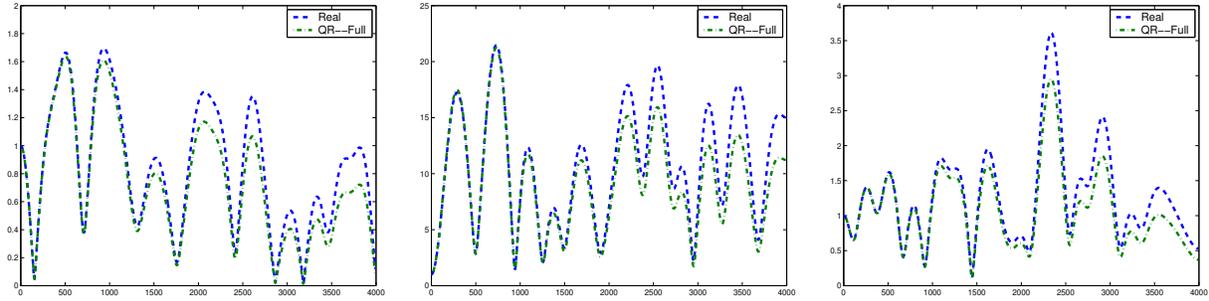


Fig. 11. Real and estimated frequency response of an arbitrary ATF of order 32 (frequency axis in Hz). Speech-like noise input. QRD method. SNR=45dB.

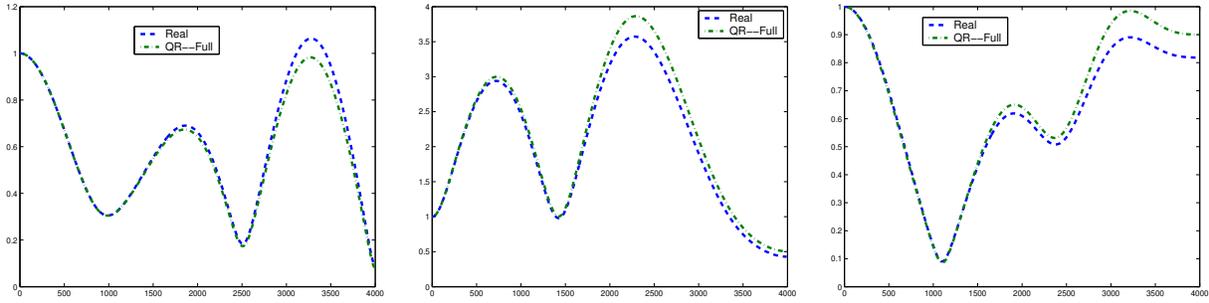


Fig. 12. Real and estimated frequency response of an arbitrary ATF of order 8 (frequency axis in Hz). White noise input. QRD method. SNR=35dB.

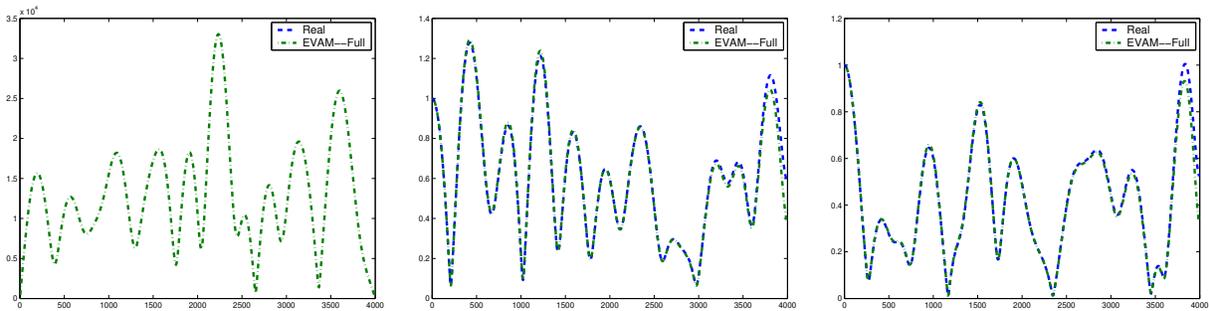


Fig. 13. Real and estimated frequency response of an arbitrary ATF of order 32 (frequency axis in Hz). Speech-like input. EVAM method. SNR=45dB.

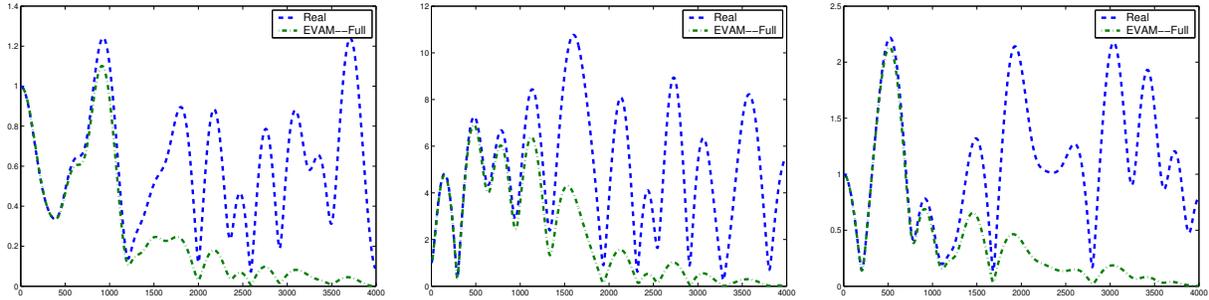


Fig. 14. Real and estimated frequency response of an arbitrary ATF of order 32 (frequency axis in Hz). Speech-like input. EVAM method. SNR=25dB.

that while high performance of the EVAM algorithm is demonstrated, degradation is encountered at the lower SNR level, especially at the high frequency range where the input signal content is low.

Finally, the incorporation of the subspace method into a subband structure is given in Figures 15 and 16. We used the 8 subband structure shown in Fig. 4. The decimation in each channel by a factor of 8 (critically decimated) allowed for a significant order reduction. In particular, the correct order of the filter in each channel was only 4. In this case we overestimated the correct order only by 2, since the null subspace determination is less robust. In Fig. 15 the subband structure is depicted, and the estimated response is given in each band separately. In Fig. 16 all the bands are combined to form the entire frequency response of the ATFs. The results demonstrate the ability

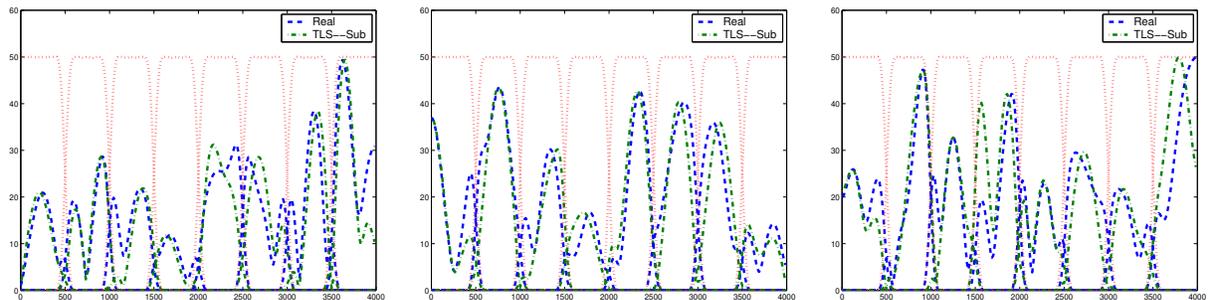


Fig. 15. Real and estimated frequency response of an arbitrary ATF of order 32 (frequency axis in Hz). Speech-like input. Subband method. SNR=25dB.

of the algorithm to work well at lower SNR levels (25dB) while the filter order is still relatively high ($n_a = 32$), even for the speech-like signal. It is worth noting that errors in the frequency response are mainly encountered in the transition regions between the frequency bands. This phenomenon

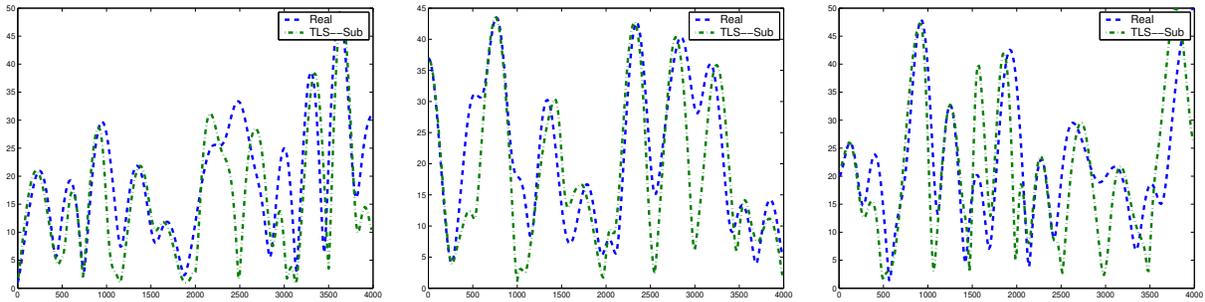


Fig. 16. Real and estimated frequency response of an arbitrary ATF of order 32 (frequency axis in Hz). Speech-like input. Subband method. SNR=25dB.

should be explored in depth, to enable a filterbank design, which is more suited to the problem at hand.

Finally, the fullband algorithm is tested with real speech signal. For demonstrating the dereverberation ability we used a high SNR level (45dB) and $n_a = 32$. The frequency response of the estimated filter is depicted in Fig. 17, for either using the entire null subspace vectors or only part of it. Performance degradation in using only part of the subspace is evident. The dereverberated speech signal is shown in Fig. 18. It is clearly shown from the figure, that while the microphone sig-

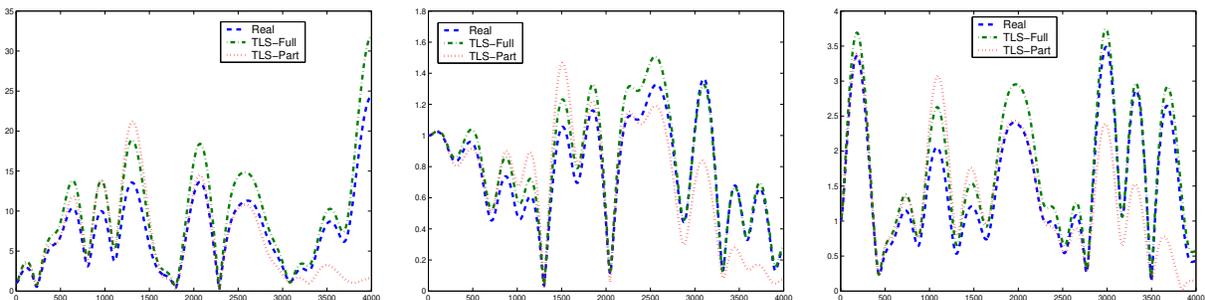


Fig. 17. Real and estimated frequency response of an arbitrary ATF of order 32 (frequency axis in Hz). Speech signal input. Fullband and Part methods. SNR=45dB.

nal is different from the original signal (due to the filtering by the ATF), the dereverberated signal resembles it. This is also supported by a more than 9dB decrease in the power of the difference.

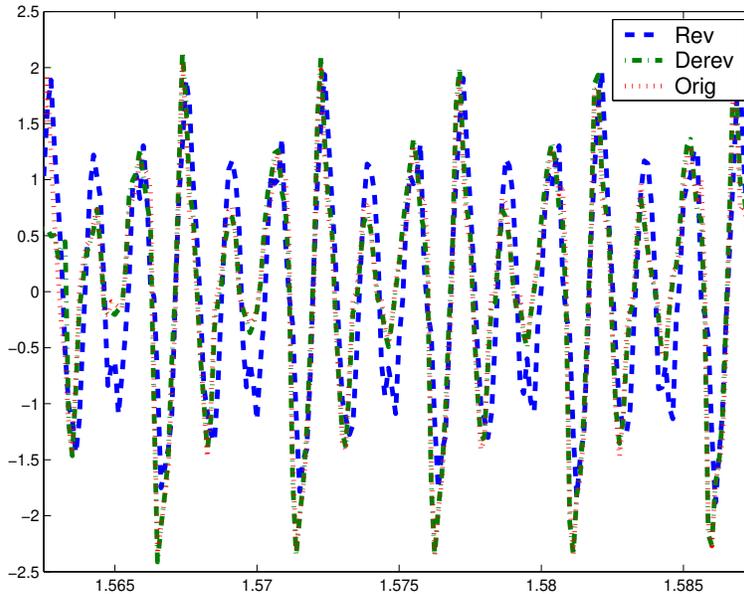


Fig. 18. Dereverberated speech. ATF of order 32. Speech signal input. Subband method. SNR=45dB.

VIII. CONCLUSIONS

A novel method for speech dereverberation, based on null subspace extraction (applying either GSVD to a noisy data matrix or GEVD to the corresponding correlation matrix) is suggested. An ATF estimation procedure is obtained by exploiting the special Sylvester structure of the corresponding filtering matrix by using TLS fitting. An alternative, more efficient method, is proposed, based on the same null subspace structure, and on the QR decomposition. The TLS approach, although imposing a high computational burden, is found to be superior to the cheaper QRD method. The desired signal is obtained by incorporating the estimated ATFs into an extended GSC structure.

Of special interest is the subband framework, in which the ATFs estimation is done in each decimated band separately and then combined to form the fullband ATFs. This technique allows an increase of the filter order which can be treated by the proposed system, while maintaining good performance even with real speech signals and higher noise levels. This method still suffers from the gain ambiguity problem, and thus, should be further explored. We note, that such subband structure might be incorporated into other methods as well (e.g. the EVAM algorithm). An experimental study supports the above conclusions. It is worth mentioning that the method (as

well as other methods in the literature) should be further explored and tested in actual scenarios, where the ATFs are much longer.

REFERENCES

- [1] Q.-G. Liu, B. Champagne, and P. Kabal, "A microphone Array Processing Technique for Speech Enhancement in a Reverberant Space," *Speech Communication*, vol. 18, pp. 317–334, 1996.
- [2] J. González-Rodríguez, J.L. Sánchez-Bote, and J. Ortega-García, "Speech dereverberation and noise reduction with a combined microphone array approach," in *International Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Istanbul, Turkey, 2000.
- [3] B. Yegnanarayana, P. Satyanarayana Murthy, C. Avendano, and H. Hermansky, "Enhancement of reverberant speech using lp residual," in *International Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Seattle, USA, 1998.
- [4] S. M. Griebel and M. S. Brandstein, "Microphone array speech dereverberation using coarse channel modeling," in *International Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Salt Lake City, Utah, USA, 2001.
- [5] S. Affes and Y. Grenier, "A Signal Subspace Tracking Algorithm for Microphone Array Processing of Speech," *IEEE trans. on Speech and Audio Proc.*, vol. 5, no. 5, pp. 425–437, Sep. 1997.
- [6] S. Doclo and M. Moonen, "Combined Frequency-Domain Dereverberation and Noise Reduction Technique for Multi-Microphone Speech Enhancement," in *The 2001 International Workshop on Acoustic Echo and Noise Control (IWAENC01)*, Darmstadt, Germany, Sep. 2001.
- [7] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace Methods for the Blind Identification of Multichannel FIR Filters," *IEEE trans. on Sig. Proc.*, vol. 43, no. 2, pp. 516–525, Feb. 1995.
- [8] G. Xu, H. Liu, L. Tong, and T. Kailath, "A Least-Squares Approach to Blind Channel Identification," *IEEE trans. on Sig. Proc.*, vol. 43, no. 12, pp. 2982–2993, Dec. 1995.
- [9] M. İ. Güreli and L. Nikias, "EVAM: An Eigenvector-Based Algorithm for Multichannel Blind Deconvolution of Input Colored Signals," *IEEE trans. on Sig. Proc.*, vol. 43, no. 1, pp. 134–149, Jan. 1995.
- [10] S. Van Huffel, H. Park, and J. Ben Rosen, "Formulation and Solution of Structured Total Least Norm Problems for Parameter Estimation," *IEEE Trans. on Signal Proc.*, vol. 44, no. 10, pp. 2464–2474, Oct. 1996.
- [11] G.H Golub and C.F. Van Loan, *Matrix Computations*, chapter 7.2, pp. 341–350, The Johns Hopkins University Press, Baltimore and London, 2nd edition, 1989.
- [12] S. Weiß, R.W. Stewart, A. Stenger, and R. Rabenstein, "Performance Limitations of Subband Adaptive Filters," in *9th European Signal Processing Conference (EUSIPCO)*, Rhodes, Greece, Sep. 1998, pp. 1245–1248.
- [13] S. Weiß, G.W. Rice, and R.W. Stewart, "Multichannel Equalization in Subbands," in *IEEE Workshop on Application of Signal Proc. to Audio and Acoustics*, New Paltz, New York, USA, Oct. 1999.
- [14] A. Spriet, M. Moonen, and J. Wouters, "A Multichannel Subband GSVD Based Approach for Speech Enhancement in Hearing Aids," in *The 2001 International Workshop on Acoustic Echo and Noise Control (IWAENC01)*, Darmstadt, Germany, Sep. 2001.

- [15] K. Eneman and M. Moonen, “DFT modulated Filter Bank Design for Oversampled Subband Systems,” *Signal Processing*, vol. 81, pp. 1947–1973, 2001.
- [16] K. Rahbar, J. P. Reilly and J.H. Manton, “A Frequency Domain Approach to Blind Identification of MIMO FIR Systems Driven by Quasi-Stationary Signals,” in *Int. Conf. on Acoustics, Speech and Signal Proc.*, 2002.
- [17] M. Miyoshi and Y. Kaneda, “Inverse Filtering of Room Acoustics,” *IEEE trans. on Acoustics, Speech and Signal Proc.*, vol. 36, no. 2, pp. 145–152, Feb. 1988.
- [18] S. Gannot, D. Burshtein, and E. Weinstein, “Signal Enhancement Using Beamforming and Non-Stationarity with application to Speech,” *IEEE Trans. on Sig. Proc.*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.
- [19] R.E Crochiere, “A weighted overlap-add method of short-time fourier analysis/synthesis,” *IEEE Trans. on ASSP*, vol. Vol 28, no. No 1, pp. 99–102, Feb. 1980.
- [20] A. Varga and H.J.M. Steeneken, “Assessment for Automatic Speech Recognition: II. NOISEX-92: A Database and an Experiment to Study the Effect of Additive Noise on Speech Recognition Systems,” *Speech Comm.*, vol. 12, pp. 247–251, 1993.
- [21] National Institute of standards and Technology, “The DARPA TIMIT acoustic-phonetic continuous speech corpus,” CD-ROM NIST Speech Disc 1-1.1, Oct. 1991.