SPEECH DEREVERBERATION VIA SUB-BAND IMPLEMENTATION OF SUBSPACE METHODS

Sharon Gannot
Faculty of Electrical Engineering, Technion, Technion City, 32000 Haifa, Israel
e-mail: gannot@siglab.technion.ac.il
Marc Moonen
Dept. of Elect. Eng. (ESAT-SISTA), K.U.Leuven, B-3001 Leuven, Belgium
e-mail: Marc.Moonen@esat.kuleuven.ac.be

ABSTRACT

A novel approach for speech dereverberation via sub-band implementation of subspace methods is presented[1]. In recent work we presented a method utilizing the null subspace of the spatial-temporal correlation matrix of the received signals (obtained by the *generalized eigenvalue decomposition* (GEVD) procedure). The desired *acoustic transfer functions* (ATF-s) are shown to be embedded in these generalized eigenvectors. The special Silvester structure of the filtering matrix, related to this subspace, was exploited for deriving a *total least squares* (TLS) estimate for the ATF-s. The high sensitivity of the GEVD procedure to noise, especially when the involved ATF-s are very long, together with the wide dynamic range of the speech signal, make the proposed method problematic in realistic scenarios. In this contribution we suggest to incorporate the TLS subspace method into a sub-band structure. The novel method proves to be efficient, although some new problems arise and other remain open. A preliminary experimental study supports the potential of the proposed method.

1 INTRODUCTION AND PROBLEM FORMULATION

The dereverberation problem, although explored for a long period, still remains an unsolved issue. The null subspace of the correlation matrix of the received signal was shown by Gürelli and Nikias [1] to maintain information on the transfer function relating the source and the receivers. This observation constitute the basis for their EVAM algorithm. This method, although originally aimed at solving communications problems, has also a potential in the speech processing framework. The same observation was recently exploited by the authors [2],[3] as the basis of a TLS based approach. We proceed now by formally introducing the problem.

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

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

where $m = 1, \ldots, M$ and $t = 0, 1, \ldots, T$. $z_m(t)$ is the $m-$th received signal, $y_m(t)$ is the corresponding desired signal part, $v_m(t)$ is the noise signal received at the $m-$th microphone, $s(t)$ is the desired speech signal and $T+1$ is the number of samples observed. Define the $Z-$transform of each of the M filters as,

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

The goal of the dereverberation problem is to reconstruct the speech signal $s(t)$ from the noisy observations $z_m(t)$, $m = 1, 2, \ldots, M$. In both full-band and sub-band approaches we try to achieve this goal by first estimating the ATF-s, and then, based on these estimates, to reconstruct the desired signal. Schematically, an *ATF Estimation* procedure, depicted in Fig. 1 is searched for.

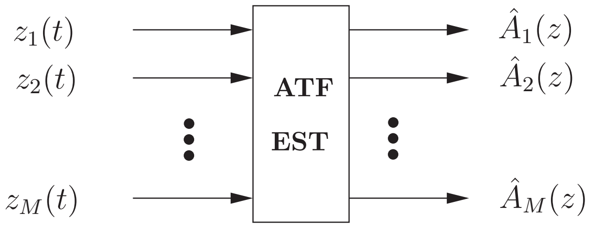

Figure 1: ATF-s estimation procedure.

The structure of the rest of this paper is as follows. In Section 2.1 we start by exploring the full-band algorithm. The drawbacks of this algorithm are stated in Section 2.2. The new sub-band method is presented in Section 3. A preliminary experimental study is given in Section 4. The open issues related with the proposed method and some future research directions are discussed in Section 5.

2 FULL-BAND ALGORITHM

In this section we briefly overview the full-band approach [2] and state its drawbacks.

[1] This research work was 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-poject *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.

2.1 Review

The essence of the use of the null subspace lies in Eq. (2).

$$[a_m(t) * y_n(t) - a_n(t) * y_m(t)] * e_l(t) = 0;\ m, n = 1, \ldots, M \tag{2}$$

($*$ denotes the convolution operation), where it can be seen that the desired ATF-s are embedded in the null subspace of the reverberated (but not noisy) signals. To exploit this observation, the *data matrices* of $y_m(t);\ m = 1, \ldots, M$ are constructed. The *data matrix* of the m-th signal is given by Eq. (3), on the top of the next page. $\hat{n}_a$ is the estimated ATF-s order, assumed to be larger than the real order n_a, i.e., the ATF-s order is always overestimated. The data matrix of all the received signals may then be constructed. In the two channel case the entire data matrix is given by,

$$\mathcal{Y}^T = \begin{bmatrix} \mathcal{Y}_2^T & -\mathcal{Y}_1^T \end{bmatrix}$$

otherwise a proper pairing of the channels may be applied [2]. The $2(\hat{n}_a+1) \times 2(\hat{n}_a+1)$ spatial-temporal correlation matrix of the data is given by $\hat{R}_y = \frac{\mathcal{Y}\mathcal{Y}^T}{T+1}$. The null subspace of the matrix $\hat{R}_y$ is the basis of the proposed algorithm, as we show in the sequel. As, usually, only noisy observations are available, it can be shown that the GEVD of the corresponding correlation matrices, $\hat{R}_z$ and $\hat{R}_v$, can be applied instead. The generalized eigenvectors related to the generalized eigenvalues of value 1 are then used. Denote these generalized eigenvectors by $\boldsymbol{g}_l$, $l = 0, 1, 2, \ldots, \hat{n}_a - n_a$. Then, splitting each null subspace vector into M parts of equal length $\hat{n}_a + 1$ we obtain,

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

From the above discussion, each of the vectors $\tilde{\boldsymbol{a}}_{m,l}$ of order $\hat{n}_a$ have the following transfer function,

$$\tilde{A}_{ml}(z) = \sum_{k=0}^{\hat{n}_a} \tilde{a}_{ml}(k) z^{-k} = A_m(z) E_l(z)$$
$$l = 0, 1, \ldots, \hat{n}_a - n_a,\ m = 1, \ldots, M. \tag{4}$$

Concatenation of these filters nullifies the noiseless data matrix. Thus, the zeros of the filters $\tilde{A}_{ml}(z)$ comprise the roots of the desired filters as well as some extraneous zeros. The common zeros of $\tilde{A}_{ml}(z);\ m = 1, \ldots, M$ constitutes the filters $E_l(z)$. Gürelli and Nikias proposed [1] a method for eliminating these common zeros.

We proceed from Eq. (4) in a different manner. In matrix form, Eq. (4) may be written in the following manner. Define the $(\hat{n}_a + 1) \times (\hat{n}_a - n_a + 1)$ Silvester filtering matrix (recall $\hat{n}_a \geq n_a$ is assumed),

$$\mathcal{A}_m = \underbrace{\begin{bmatrix} a_m(0) & 0 & 0 & \cdots & 0 \\ a_m(1) & a_m(0) & 0 & \cdots & 0 \\ \vdots & a_m(1) & \ddots & & \vdots \\ a_m(n_a) & \vdots & \ddots & \ddots & 0 \\ 0 & a_m(n_a) & & \ddots & a_m(0) \\ \vdots & 0 & & & a_m(1) \\ & \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & a_m(n_a) \end{bmatrix}}_{\hat{n}_a - n_a + 1}. \tag{5}$$

Then,

$$\tilde{\boldsymbol{a}}_{ml} = \mathcal{A}_m \mathbf{e}_l, \tag{6}$$

where, $\mathbf{e}_l^T = \begin{bmatrix} e_l(0)\, e_l(1) \ldots e_l(\hat{n}_a - n_a) \end{bmatrix}$ are vectors of the coefficients of the arbitrary unknown filters $E_l(z)$. Thus, the number of different filters (as shown in Eq. (4)) is $\hat{n}_a - n_a + 1$ and their order is $\hat{n}_a - n_a$. Let $\mathcal{E} = \begin{bmatrix} \mathbf{e}_0\, \mathbf{e}_1 \cdots \mathbf{e}_{\hat{n}_a - n_a} \end{bmatrix}$ be an $(\hat{n}_a - n_a + 1) \times (\hat{n}_a - n_a + 1)$ unknown matrix, then

$$\mathcal{G} = \begin{bmatrix} \mathcal{A}_1 \\ \vdots \\ \mathcal{A}_M \end{bmatrix} \mathcal{E} \overset{\triangle}{=} \mathcal{A}\mathcal{E}. \tag{7}$$

Note, that in the special case where the order of the ATF-s is known, i.e. $\hat{n}_a = n_a$, there is only one vector in the null subspace and its partitions $\tilde{\boldsymbol{a}}_{m0}\ ;\ m = 1, \ldots, M$ are equal to the desired filters $\boldsymbol{a}_m$ up to a (common) scaling factor ambiguity. In the case where $\hat{n}_a > n_a$, the actual ATF-s $A_m(z)$ are embedded in $\tilde{A}_{ml}(z)\ ;\ l = 0, 1, \ldots, \hat{n}_a - n_a$. The case $\hat{n}_a < n_a$ could not be treated properly by the proposed method. Based on the special structure of Eq. (7) and in particular on the Silvester structure of $\mathcal{A}_m$, we derive now an algorithm for extracting the ATF-s $A_m(z)$. $\mathcal{E}$ in Eq. (7) is a square and arbitrary matrix, implying that its inverse usually exists. Denote this inverse by $\mathcal{E}^i = \mathrm{inv}(\mathcal{E})$. Then.

$$\mathcal{G}\mathcal{E}^i = \mathcal{A} \tag{8}$$

Denote the columns of $\mathcal{E}^i$ by $\mathcal{E}^i = \begin{bmatrix} \mathbf{e}_0^i\, \mathbf{e}_1^i \cdots \mathbf{e}_{\hat{n}_a - n_a}^i \end{bmatrix}$. Then, Eq. (8) can be rewritten as,

$$\tilde{\mathcal{G}} \boldsymbol{x} = \mathbf{0}. \tag{9}$$

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

$$\tilde{\mathcal{G}} = \begin{bmatrix} \mathcal{G} & \mathcal{O} & \cdots & \cdots & \cdots & \mathcal{O} & -\mathcal{I}^{(0)} \\ \mathcal{O} & \mathcal{G} & \mathcal{O} & \cdots & \cdots & \mathcal{O} & -\mathcal{I}^{(1)} \\ \vdots & \mathcal{O} & \ddots & & & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & \mathcal{O} & \vdots \\ \mathcal{O} & \mathcal{O} & \cdots & \cdots & \mathcal{O} & \mathcal{G} & -\mathcal{I}^{\hat{n}_a - n_a} \end{bmatrix} \tag{10}$$

The vector of unknowns is defined by,

$$\boldsymbol{x}^T = \left[{\mathbf{e}_0^i}^T\, {\mathbf{e}_1^i}^T \cdots {\mathbf{e}_{\hat{n}_a - n_a}^i}^T\, \boldsymbol{a}_1{}^T\, \boldsymbol{a}_2{}^T \ldots \boldsymbol{a}_M{}^T. \right]$$

$\mathbf{0}$ and $\mathcal{O}$ are vector and matrix, respectively, of zeros of proper dimensions. $\mathcal{I}^{(l)}\ ;\ l = 0, 1, \ldots, \hat{n}_a - n_a$ is a fixed shift-by-l matrix.

Note however, that in most cases equality in Eq. (9) only approximately holds. Therefore, we suggest to use the *total least squares* (TLS) algorithm by picking the eigenvector $\boldsymbol{x}$ which corresponds to the smallest eigenvalue of the matrix $\tilde{\mathcal{G}}$.

2.2 Drawbacks

The proposed full-band method although theoretically supported have several severe 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 case, the GEVD procedure is not robust enough and quite sensitive to small estimation errors in the correlation matrix. Furthermore, the matrices involved become extremely large causing huge memory and computational requirements. Another problem arise from the wide dynamic

$$\mathcal{Y}_m = \begin{bmatrix} y_m(0) & y_m(1) & \cdots & & y_m(\hat{n}_a) & y_m(\hat{n}_a+1) & \cdots & y_m(T) & 0 & \cdots & 0 \\ 0 & y_m(0) & y_m(1) & \cdots & \vdots & \vdots & & \cdots & y_m(T) & 0 & 0 \\ \vdots & 0 & \ddots & \ddots & & & & & & \ddots & \vdots \\ 0 & & & \ddots & \vdots & & \ddots & & 0 & & \ddots \\ 0 & & \cdots & 0 & y_m(0) & y_m(1) & \cdots & y_m(\hat{n}_a) & \cdots & & y_m(T) \end{bmatrix} \quad (3)$$

range of the speech signal. This phenomenon may result in an erroneous estimates of the frequency response of the ATF-s in the low energy bands of the input signal.

Altogether these drawbacks render the proposed method useless in most practical speech dereverberation applications.

3 SUB-BAND APPROACH

To tackle the problems which arise in the full-band approach, sub-band implementation of the TLS subspace method is proposed. The use of sub-bands for splitting adaptive filters, especially in the context of echo cancellation, has gained recent interest in the literature. However, the use of sub-bands in subspace methods is not as common.

The M microphone signals are filtered by a sub-band structure, yielding a total of LM signals, $z_m^l(t)$; $l = 0, \ldots, L-1$; $m = 1, \ldots, M$. The signals are depicted in Fig. 2. The full-band subspace methods presented above is now applied to each sub-band signal separately. Although the resulting sub-band signals effectively correspond to a longer filter (which is the convolution of the corresponding ATF and the sub-band filter), the algorithm is aimed at reconstructing the ATF alone, ignoring the filter-bank roots. This is due to the fact that the zeros of the sub-band filter are common to all channels $z_m^l(t)$; $m = 1, \ldots, M$, with l fixed, and that subspace method is blind to common zeros (see (4)). For properly exploiting the benefits of the sub-band structure, each sub-band signal should be decimated. We choose critically decimated filter-bank, i.e. the decimation factor equals the number of bands.

This procedure has a twofold advantage. First, the ATF order in each band is approximately reduced by the decimation factor, making the estimation task easier. Second, after filtering and decimating the signals at each sub-band become flatter, making the signals effectively whiter, resulting again an improved performance. After estimating the decimated ATF-s, they are combined together using a proper synthesis filter-bank, comprised of interpolation followed by a filter-bank similar to the analysis filter-bank.

The design of the filter-bank is of crucial importance. Special emphasis should be given to adjusting the sub-band 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 sub-band structure, depicted in Fig. 3, is used. 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. These filters constitute the analysis and synthesis filter-banks H_l, G_l; $l = 0, 1, \ldots, L-1$.

Gain ambiguity may be a major drawback of the sub-band method. Recall that the TLS-subspace method is estimating the ATF-s up to a common gain factor. In the full-band scheme this does not impose any problem, since it results in an overall scaling of the output. However, in the sub-band scheme, the gain factor is common for all sub-band signals but is generally different from band to band. Thus, the estimated ATF-s (and the reconstructed signal) is effectively filtered by an arbitrary filter, which can be regarded as a new reverberation term. Although several methods can be applied to overcome this gain ambiguity problem, in this contribution we assume that the gain in each sub-band is known. Thus only the ability of the method to estimate the frequency shaping in each band is demonstrated. The gain ambiguity problem is left for further research.

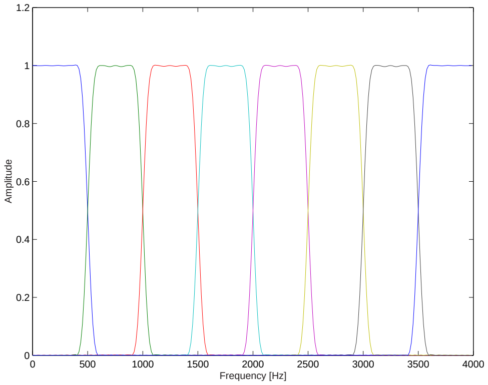

Figure 3: Sub-band structure. 8 equi–spaced equi–bandwidth filters.

4 EXPERIMENTAL STUDY

A preliminary experimental study is conducted to test the potential of the proposed method. Filters with exponentially decaying envelope and of order $n_a = 32$ are used to simulate the ATF-s. Speech-like noise presented input signal with wide dynamic range. The 8 channel sub-band structure depicted in Fig. 3 is used. Decimation in each channel by a factor of 8 (critically decimated) allow for a significant order reduction. In particular, the approximate order of the filter in each band is $\frac{32}{8} = 4$. While applying the TLS estimation algorithm, this order is overestimated only by 2. In Fig. 4 (Left) the estimated response in each sub-band is depicted, together with the sub-band structure used. The response is given for each band separately. In Fig. 4 (Right) all the bands are combined to form the entire frequency response of the ATF-s. The results demonstrate the ability of the algorithm to work well at lower SNR levels (25dB) while the filter order is still relatively high, even for the speech-like signal. This is in contrast to the full-band method which collapses even in a lower order. It is worth noting that errors in the frequency response are mainly encountered in the transition regions between the frequency bands. This phenomenon should be explored in depth, to enable a filter-bank design, which is more suited to the problem at hand.

5 DISCUSSION

The incorporation of the sub-band structure partially solves the problems encountered in the full-band algorithm. Longer ATF-s may now be dealt with, since in each sub-band only

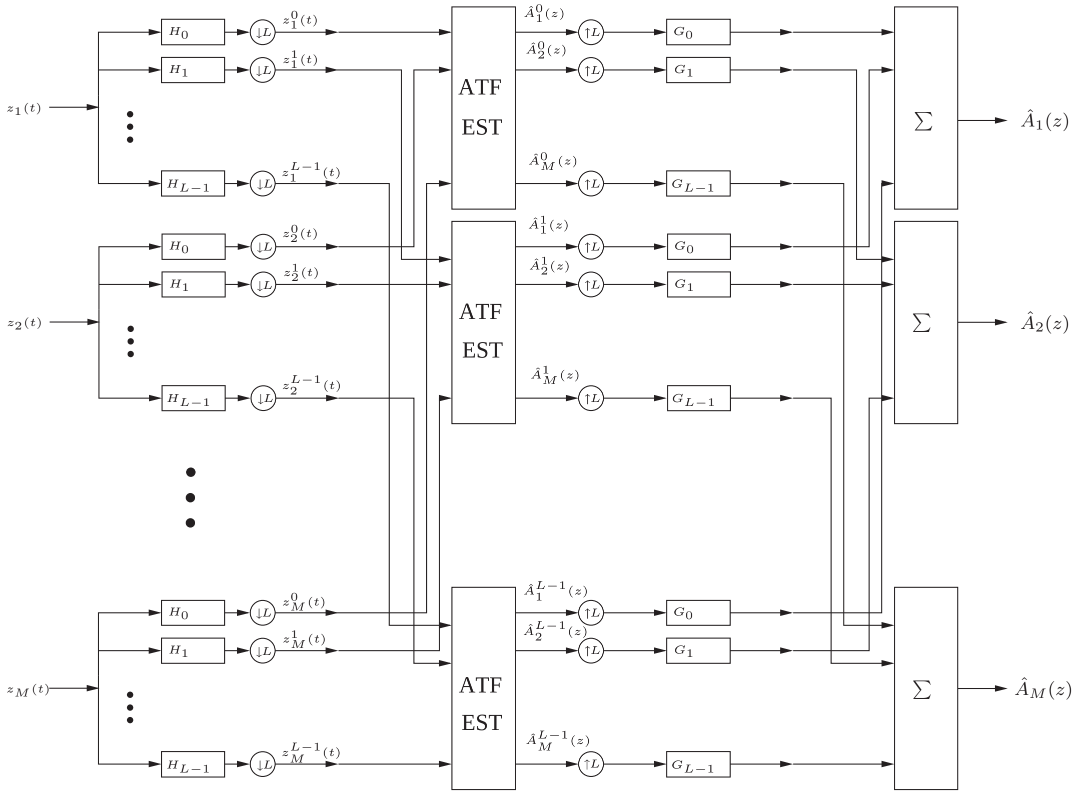

Figure 2: Null subspace in the two microphone noiseless case.

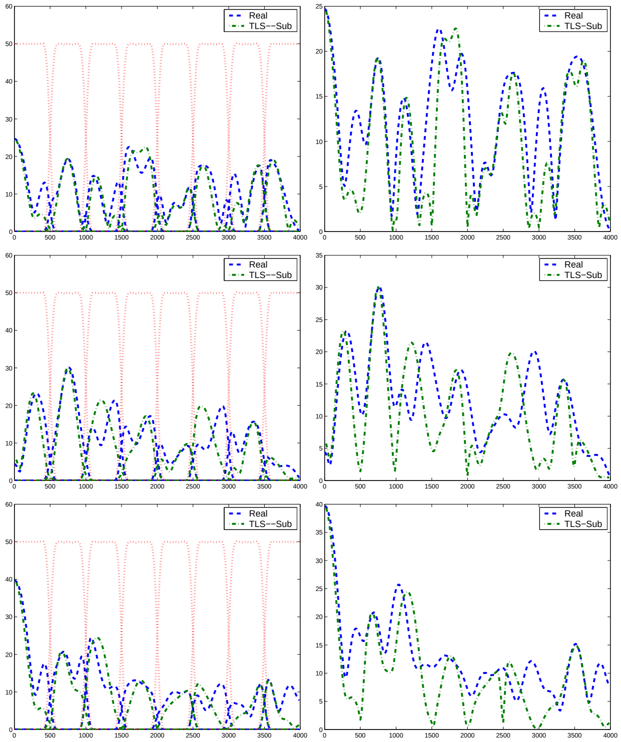

Figure 4: Sub-band method: estimated frequency response (frequency axis in Hz) of an ATF. Order 32, speech-like input, SNR=25dB. Separate bands (Left). Combined bands (Right).

shorter ATF-s are estimated. Besides, as the sub-bands become narrower, the input signal turns flatter, enabling the algorithm to deal with signals with wide dynamic range, like the speech signal.

Nevertheless, Several issues remain open. First, the gain ambiguity problem is not solved. Overlapping between bands or non-equal bands, might be ways to mitigate this problem. Another way might be to use the original input signals gain. Second, the estimation in the transition between bands is poor. Oversampled bands should be tested as a way to overcome this problem. Third, the SNR tested is still too high and the ATF-s are still very short to represent realistic scenarios. Finally, the proposed structure is not computationally efficient enough. The use of the *short time Fourier transform* (STFT) as a filter-bank is under current investigation. However, the potential of the sub-band method encourages further research on the structure.

6 *

References

[1] M. İ. Gürelli 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.

[2] S. Gannot and M. Moonen, "Subspace Methods for Multi-Microphone Speech Dereverberation," in *The 2001 International Workshop on Acoustic Echo and Noise Control (IWAENC01)*, Darmstadt, Germany, Sep. 2001.

[3] S. Gannot and M. Moonen, "Subspace Methods for Multi-Microphone Speech Dereverberation," CCIT report 398, Technion - Israel Institute of Technology, Haifa, Israel, Oct 2002.