

Subspace Methods for Multimicrophone Speech Dereverberation

Sharon Gannot

School of Engineering, Bar-Ilan University, Ramat-Gan 52900, Israel
Email: gannot@siglab.technion.ac.il

Marc Moonen

Department of Electrical Engineering, Katholieke Universiteit Leuven, ESAT-SISTA, Kasteelpark Arenberg 10, B-3001 Heverlee, Belgium
Email: Marc.Moonen@esat.kuleuven.ac.be

Received 2 September 2002 and in revised form 14 March 2003

A novel approach for multimicrophone 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 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 and phrases: speech dereverberation, subspace methods, subband processing.

1. 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 multimicrophone measurements.

Spatiotemporal methods, which are directly applied to the received signals, have been presented by Liu et al. [1] and by Gonzalez-Rodriguez 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 acoustical transfer functions (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 prove to be efficient and quite robust to small speaker movements. However, if this information is not available, these methods cannot 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 (see, e.g., [7, 8]). The EVAM algorithm presented by Güreli and Nikias [9] is of special interest. 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 several algorithms. Both fullband and subband versions are derived. A shorter preliminary conference version of the fullband methods has been published in [10].

The general dereverberation problem is presented in Section 2. The proposed method is outlined in Section 3. 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 total least squares (TLS) approach for ATF estimation. Suboptimal procedures, based on the QR decomposition (QRD), are derived in Section 4. The use of decimated subbands for reducing the complexity of the algorithm and increasing its robustness is explored in Section 5. A reconstruction procedure, based on the ATFs' matched filter and incorporated into an extension of the *generalized sidelobe canceller* (GSC) is proposed in Section 6. The derivation of the algorithms is followed by an experimental study presented in Section 7.

2. 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

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

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

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

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

$$A_m(z) = \sum_{k=0}^{n_a} a_m(k)z^{-k}, \quad m = 1, 2, \dots, M. \quad (3)$$

All the involved signals and ATFs are depicted in Figure 1. The goal of the 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 estimates. Schematically, an *ATF estimation* procedure, depicted in Figure 2, is searched for.

3. 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

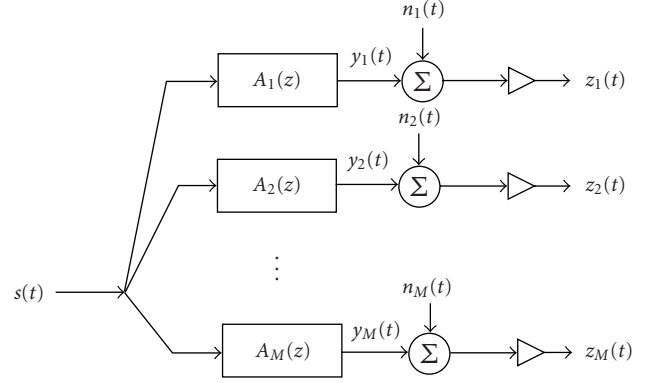


FIGURE 1: The general dereverberation problem.

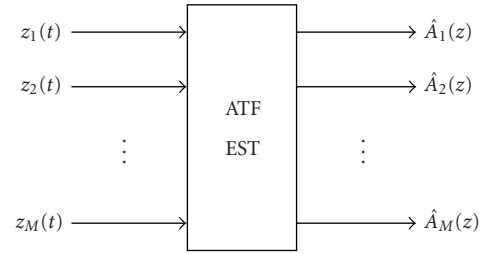


FIGURE 2: ATF estimation.

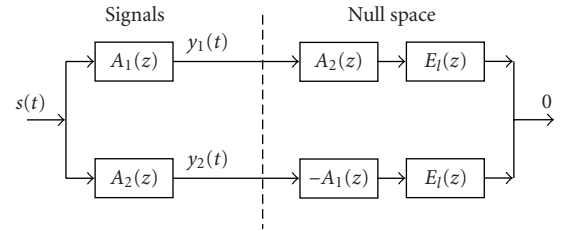


FIGURE 3: Null subspace in the two-microphone noiseless case.

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-microphone noiseless* case. We proceed through the *two microphones contaminated by colored noise* case. Then the general *multimicrophone colored noise* case is addressed. Special treatment for the case when only part of the null subspace vectors are determined concludes this section.

3.1. Two-microphone noiseless case

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 (see, e.g., [9]), but in a more intuitive way of presentation.

3.1.1. Preliminaries

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

$$\begin{aligned} y_1(t) &= a_1(t) * s(t), \\ y_2(t) &= a_2(t) * s(t). \end{aligned} \quad (4)$$

Clearly, as depicted in the right-hand side of Figure 3, the following identity holds:

$$[y_2(t) * a_1(t) - y_1(t) * a_2(t)] * e_l(t) = 0, \quad (5)$$

where $e_l(t)$, $l = 0, 1, 2, \dots$, are arbitrary and unknown filters, the number and the order of which 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, 8, 9].

Define the $(\hat{n}_a + 1) \times (T + \hat{n}_a + 1)$ single-channel data matrix

$$\mathcal{Y}_m = \begin{bmatrix} y_m(0) & y_m(1) & \cdots & y_m(\hat{n}_a) & y_m(\hat{n}_a + 1) & \cdots & y_m(T) & 0 & \cdots & 0 \\ 0 & y_m(0) & y_m(1) & \cdots & \vdots & \vdots & \cdots & y_m(T) & 0 & 0 \\ \vdots & 0 & \ddots & \ddots & & & & & \ddots & \vdots \\ 0 & & & \ddots & \vdots & & \ddots & 0 & & \ddots \\ 0 & & \cdots & 0 & y_m(0) & y_m(1) & \cdots & y_m(\hat{n}_a) & \cdots & y_m(T) \end{bmatrix}. \quad (6)$$

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

$$\mathcal{Y} = \begin{bmatrix} \mathcal{Y}_2 \\ -\mathcal{Y}_1 \end{bmatrix}. \quad (7)$$

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

Now, following [7, 9], the null subspace of the correlation matrix can be calculated by virtue of the eigenvalue decomposition (EVD). 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

$$\begin{aligned} \lambda_l &= 0, \quad l = 0, 1, \dots, \hat{n}_a - n_a, \\ \lambda_l &> 0, \quad \text{otherwise.} \end{aligned} \quad (8)$$

Thus, as proven by Gurelli 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 EVD 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 $l = 0, 1, 2, \dots, \hat{n}_a - n_a$. Then, splitting each null subspace

vector into two parts of equal length $\hat{n}_a + 1$, we obtain

$$\mathcal{G} = [\mathbf{g}_0 \quad \mathbf{g}_1 \quad \cdots \quad \mathbf{g}_{\hat{n}_a - n_a}] = \begin{bmatrix} \tilde{\mathbf{a}}_{1,0} & \tilde{\mathbf{a}}_{1,1} & \cdots & \tilde{\mathbf{a}}_{1,\hat{n}_a - n_a} \\ \tilde{\mathbf{a}}_{2,0} & \tilde{\mathbf{a}}_{2,1} & \cdots & \tilde{\mathbf{a}}_{2,\hat{n}_a - n_a} \end{bmatrix}. \quad (9)$$

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

$$\tilde{A}_{ml}(z) = \sum_{k=0}^{\hat{n}_a} \tilde{a}_{ml}(k) z^{-k}, \quad l = 0, 1, \dots, \hat{n}_a - n_a, \quad m = 1, 2. \quad (10)$$

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

$$\tilde{A}_{ml}(z) = A_m(z) E_l(z), \quad l = 0, 1, \dots, \hat{n}_a - n_a, \quad m = 1, 2. \quad (11)$$

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 Gurelli 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 ATFs $A_m(z)$, 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, that is, $\tilde{A}_{ml}(z)$ for $m = 1, \dots, M$.*

Under several regularity conditions (stated, e.g., by Moulines et al. [7]), the filters $A_m(z)$ can be found. An observation of special interest is that common roots of the filters $A_m(z)$ cannot be extracted by the algorithm as they are treated as the extraneous roots which constitute $E_l(z)$. Although this is a drawback of the method, we will take benefit of it while constructing a subband structure in Section 5.

In matrix form, equation (11) may be written in the following manner. Define the $(\hat{n}_a + 1) \times (\hat{n}_a - n_a + 1)$ Sylvester filtering matrix (recall that we assume $\hat{n}_a \geq n_a$)

$$\mathcal{A}_m = \begin{bmatrix} a_m(0) & 0 & 0 & \cdots & 0 \\ a_m(1) & a_m(0) & 0 & \cdots & 0 \\ \vdots & a_m(1) & \ddots & & \vdots \\ a_m(n_a) & \vdots & \ddots & \ddots & 0 \\ 0 & a_m(n_a) & & \ddots & a_m(0) \\ \vdots & 0 & & & a_m(1) \\ & \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & a_m(n_a) \end{bmatrix}. \quad (12)$$

$\underbrace{\hspace{10em}}_{\hat{n}_a - n_a + 1}$

Then,

$$\tilde{\mathbf{a}}_{ml} = \mathcal{A}_m \mathbf{e}_l, \quad (13)$$

where $\mathbf{e}_l^T = [e_l(0) \ e_l(1) \ \cdots \ e_l(\hat{n}_a - n_a)]$ are vectors of the coefficients of the arbitrary unknown filters $E_l(z)$. Thus, the number of different filters (as shown in (11)) is $\hat{n}_a - n_a + 1$ and their order is $\hat{n}_a - n_a$. Using Figure 3 and identity (5), and denoting

$$\mathcal{E} = [\mathbf{e}_0 \ \mathbf{e}_1 \ \cdots \ \mathbf{e}_{\hat{n}_a - n_a}], \quad (14)$$

we conclude that

$$\mathcal{G} = \begin{bmatrix} \mathcal{A}_1 \\ \mathcal{A}_2 \end{bmatrix} \mathcal{E} \triangleq \mathcal{A} \mathcal{E}, \quad (15)$$

where \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, that is, $\hat{n}_a = n_a$, 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 $\tilde{A}_m(z)$, $l = 0, 1, \dots, \hat{n}_a - n_a$. The case $\hat{n}_a < n_a$ could not be treated properly by the proposed method.

The special structure depicted in (15) and (12) forms the basis of our suggested algorithm.

3.1.2. Algorithm

Based on the special structure of (15) and, in particular, on the Sylvester structure of \mathcal{A}_1 and \mathcal{A}_2 , found in (12), we derive now an algorithm for finding the ATFs $A_m(z)$.

Note that \mathcal{E} in (15) is a square and arbitrary matrix, implying that its inverse usually exists. Denote this inverse by

$\mathcal{E}^i = \text{inv}(\mathcal{E})$. Then

$$\mathcal{G} \mathcal{E}^i = \mathcal{A}. \quad (16)$$

Denote the columns of \mathcal{E}^i by $\mathcal{E}^i = [\mathbf{e}_0^i \ \mathbf{e}_1^i \ \cdots \ \mathbf{e}_{\hat{n}_a - n_a}^i]$. Equation (16) can be then rewritten as

$$\tilde{\mathcal{G}} \mathbf{x} = \mathbf{0}, \quad (17)$$

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

$$\tilde{\mathcal{G}} = \begin{bmatrix} \mathcal{G} & \mathcal{O} & \cdots & \cdots & \cdots & \mathcal{O} & -\mathcal{J}^{(0)} \\ \mathcal{O} & \mathcal{G} & \mathcal{O} & \cdots & \cdots & \mathcal{O} & -\mathcal{J}^{(1)} \\ \vdots & \mathcal{O} & \ddots & & & \vdots & \vdots \\ \vdots & \vdots & & \ddots & \ddots & \vdots & \vdots \\ \vdots & \vdots & & & \ddots & \mathcal{O} & \vdots \\ \mathcal{O} & \mathcal{O} & \cdots & \cdots & \mathcal{O} & \mathcal{G} & -\mathcal{J}^{\hat{n}_a - n_a} \end{bmatrix} \quad (18)$$

and the vector of unknowns is defined as

$$\mathbf{x}^T = [\mathbf{e}_0^{iT} \ \mathbf{e}_1^{iT} \ \cdots \ \mathbf{e}_{\hat{n}_a - n_a}^{iT} \ \mathbf{a}_1^T \ \mathbf{a}_2^T], \quad (19)$$

where $\mathbf{0}$ is a vector of zeros: $\mathbf{0}^T = [0 \ 0 \ \cdots \ 0]$. 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 $\mathcal{J}^{(l)}$, $l = 0, 1, \dots, \hat{n}_a - n_a$, is a fixed shifting matrix given by

$$\mathcal{J}^{(l)} = \begin{bmatrix} \mathcal{O}_{l \times (n_a + 1)} & \mathcal{O}_{(\hat{n}_a + 1) \times (n_a + 1)} \\ I_{(n_a + 1) \times (n_a + 1)} & \mathcal{O}_{(\hat{n}_a + 1) \times (n_a + 1)} \\ \mathcal{O}_{(\hat{n}_a - n_a - l) \times (n_a + 1)} & \mathcal{O}_{l \times (n_a + 1)} \\ \mathcal{O}_{(\hat{n}_a + 1) \times (n_a + 1)} & I_{(n_a + 1) \times (n_a + 1)} \\ \mathcal{O}_{(\hat{n}_a - n_a - l) \times (n_a + 1)} & \mathcal{O}_{(\hat{n}_a - n_a - l) \times (n_a + 1)} \end{bmatrix}, \quad (20)$$

where $I_{(n_a + 1) \times (n_a + 1)}$ is the $(n_a + 1) \times (n_a + 1)$ identity matrix. A nontrivial (and exact) solution for the homogenous set of (17) 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 first part of the eigenvector is comprised of the nuisance parameters \mathbf{e}_l^i , $l = 0, 1, \dots, \hat{n}_a - n_a$. In the presence of noise, the somewhat nonstraightforward procedure will prove to be useful.

3.2. 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 (1), can be stacked into a data matrix fulfilling

$$\mathcal{Z} = \mathcal{Y} + \mathcal{N}, \quad (21)$$

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

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

$$\mathcal{N}_m = \begin{bmatrix} n_m(0) & n_m(1) & \cdots & n_m(\hat{n}_a) & n_m(\hat{n}_a + 1) & \cdots & n_m(T) & 0 & \cdots & 0 \\ 0 & n_m(0) & n_m(1) & \cdots & \vdots & \vdots & \cdots & n_m(T) & 0 & 0 \\ \vdots & 0 & \ddots & \ddots & \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & & \ddots & \vdots & \vdots & \ddots & 0 & \ddots & \ddots & \vdots \\ 0 & & \cdots & 0 & n_m(0) & n_m(1) & \cdots & n_m(\hat{n}_a) & \cdots & n_m(T) \end{bmatrix}. \quad (22)$$

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

$$\hat{R}_z \approx \hat{R}_y + \hat{R}_n, \quad (23)$$

where $\hat{R}_z = \mathcal{X}\mathcal{X}^T/(T+1)$ and $\hat{R}_n = \mathcal{N}\mathcal{N}^T/(T+1)$ are the noisy signal and noise-only signal correlation matrices, respectively. Now, (17) will not be accurate anymore. First, the null subspace matrix \mathcal{G} should be determined in a slightly different manner than suggested in (8). The white noise and colored noises cases are treated separately in the sequel. Second, the matrix $\tilde{\mathcal{G}}$ will not in general have an eigenvalue of value 0. A reasonable approximation for the solution, although not exact, would be to transform (17) into the following problem:

$$\tilde{\mathcal{G}}\mathbf{x} = \boldsymbol{\mu}, \quad (24)$$

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 [11] are called for. This issue will not be treated in this work anymore.

3.2.1. White noise case

In the case of spatiotemporally white noise—that is, $\hat{R}_n \approx \sigma^2 I$, where I is the identity matrix—the first $\hat{n}_a - n_a + 1$ eigenvalues in (8) will be σ^2 instead of zero. The corresponding eigenvectors will remain intact. Thus, the algorithm remains unchanged.

3.2.2. Colored noise case

The case of nonwhite noise signal was addressed in [7, 9]. In contrast to the noise balancing method presented in [9] and the prewhitening of the noise correlation matrix presented

in [7], the problem is treated here more rigorously, 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 GSVD of the corresponding data matrices \mathcal{X} and \mathcal{N} can be used. After determining the null subspace matrix, subsequent steps of the algorithm remain intact.

3.3. Multimicrophone case ($M > 2$)

In the multimicrophone case, a reasonable extension would be based on channel pairing (see [9]). Each of the pairs $M \times (M - 1)/2$ fulfills

$$[y_i(t) * a_j(t) - y_j(t) * a_i(t)] * e_l(t) = 0 \quad (25)$$

$$i, j = 1, 2, \dots, M; l = 0, 1, \dots, \hat{n}_a - n_a.$$

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

$$\mathcal{X} = \begin{bmatrix} \mathcal{X}_2 & \mathcal{X}_3 & \cdots & \mathcal{X}_M & \mathbb{O} & \cdots & \mathbb{O} & \cdots & \mathbb{O} \\ -\mathcal{X}_1 & \mathbb{O} & \cdots & \mathbb{O} & \mathcal{X}_3 & \cdots & \mathcal{X}_M & \cdots & \mathbb{O} \\ \mathbb{O} & -\mathcal{X}_1 & & -\mathcal{X}_2 & \mathbb{O} & & & & \vdots \\ \vdots & \mathbb{O} & \ddots & & \vdots & & & & \mathbb{O} \\ & \vdots & & & & \mathbb{O} & & & \mathcal{X}_M \\ \mathbb{O} & \mathbb{O} & \cdots & -\mathcal{X}_1 & \cdots & -\mathcal{X}_2 & \cdots & -\mathcal{X}_{M-1} \end{bmatrix}, \quad (26)$$

where \mathbb{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 \mathcal{G} , we can construct a new TLS equation:

$$\tilde{\mathcal{G}}\mathbf{x} = \boldsymbol{\mu}, \quad (27)$$

where $\tilde{\mathcal{G}}$ is constructed in a similar way as \mathcal{G} was constructed in (18). The vector of unknowns \mathbf{x} is given by

$$\mathbf{x}^T = [\mathbf{e}_0^T \ \mathbf{e}_1^T \ \cdots \ \mathbf{e}_{\hat{n}_a - n_a}^T \ \mathbf{a}_1^T \ \mathbf{a}_2^T \ \cdots \ \mathbf{a}_M^T]. \quad (28)$$

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

3.4. 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 (15) is then of dimensions $(\hat{n}_a - n_a + 1) \times \bar{L}$, and is thus noninvertible. To overcome this problem, we suggest concatenating several shifted versions of (15) in the following manner:

$$\tilde{\mathcal{G}} = \begin{bmatrix} \mathcal{G} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{G} & \mathbf{0} & \mathbf{0} \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & & & \mathcal{G} \end{bmatrix} = \tilde{\mathcal{A}} \underbrace{\begin{bmatrix} \mathcal{E} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathcal{E} & \mathbf{0} & \mathbf{0} \\ \vdots & & \ddots & \mathbf{0} \\ \mathbf{0} & & & \mathcal{E} \end{bmatrix}}_{L > \hat{n}_a - n_a + \hat{L}} = \tilde{\mathcal{A}}\tilde{\mathcal{E}}. \quad (29)$$

The new dimensions of $\tilde{\mathcal{E}}$ 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 column dimension.

The matrix $\tilde{\mathcal{A}}$ has a similar structure as \mathcal{A} in (12) and (15) but with more columns. The resulting matrix $\tilde{\mathcal{E}}$ has now more columns than rows and thus can generally be pseudo-inverted:

$$\mathcal{E}^{Pi} = \text{Pinv}(\tilde{\mathcal{E}}) = \tilde{\mathcal{E}}^T (\tilde{\mathcal{E}}\tilde{\mathcal{E}}^T)^{-1}, \quad (30)$$

resulting in

$$\tilde{\mathcal{G}}\mathcal{E}^{Pi} = \tilde{\mathcal{A}}. \quad (31)$$

Now the extended matrix $\tilde{\mathcal{G}}$ can be used in (24) instead of \mathcal{G}

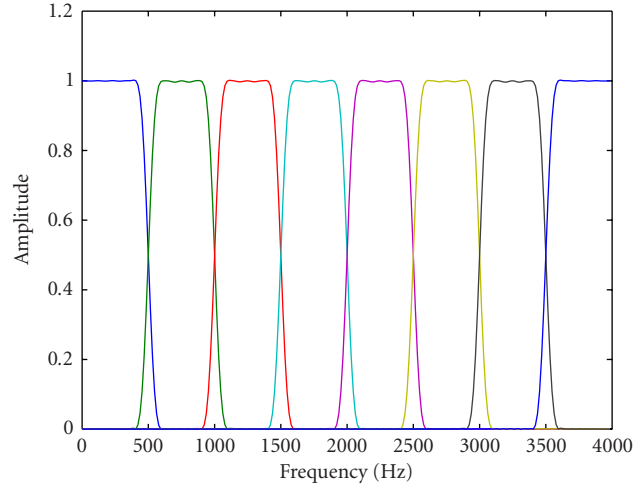


FIGURE 4: Subband structure. Eight equispaced equi-bandwidth filters.

to construct $\tilde{\mathcal{G}}$ in a similar manner to (18). Subsequent stages of the algorithm remain intact.

4. 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.

4.1. Complete knowledge of the null subspace

Apply the transpose operation to (15),

$$\mathcal{G}^T = \mathcal{E}^T \mathcal{A}^T. \quad (32)$$

As \mathcal{E}^T is an arbitrary matrix, it will usually have a QRD. Denote $\mathcal{E}^T = Q_{\mathcal{E}} R_{\mathcal{E}}$. Then,

$$\mathcal{G}^T = Q_{\mathcal{E}} R_{\mathcal{E}} \mathcal{A}^T = Q_{\mathcal{E}} R_{\mathcal{G}}, \quad (33)$$

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, equation (33) constitutes the QRD of \mathcal{G}^T . As $R_{\mathcal{E}}$ is a square and upper triangular matrix, it has only one nonzero 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.

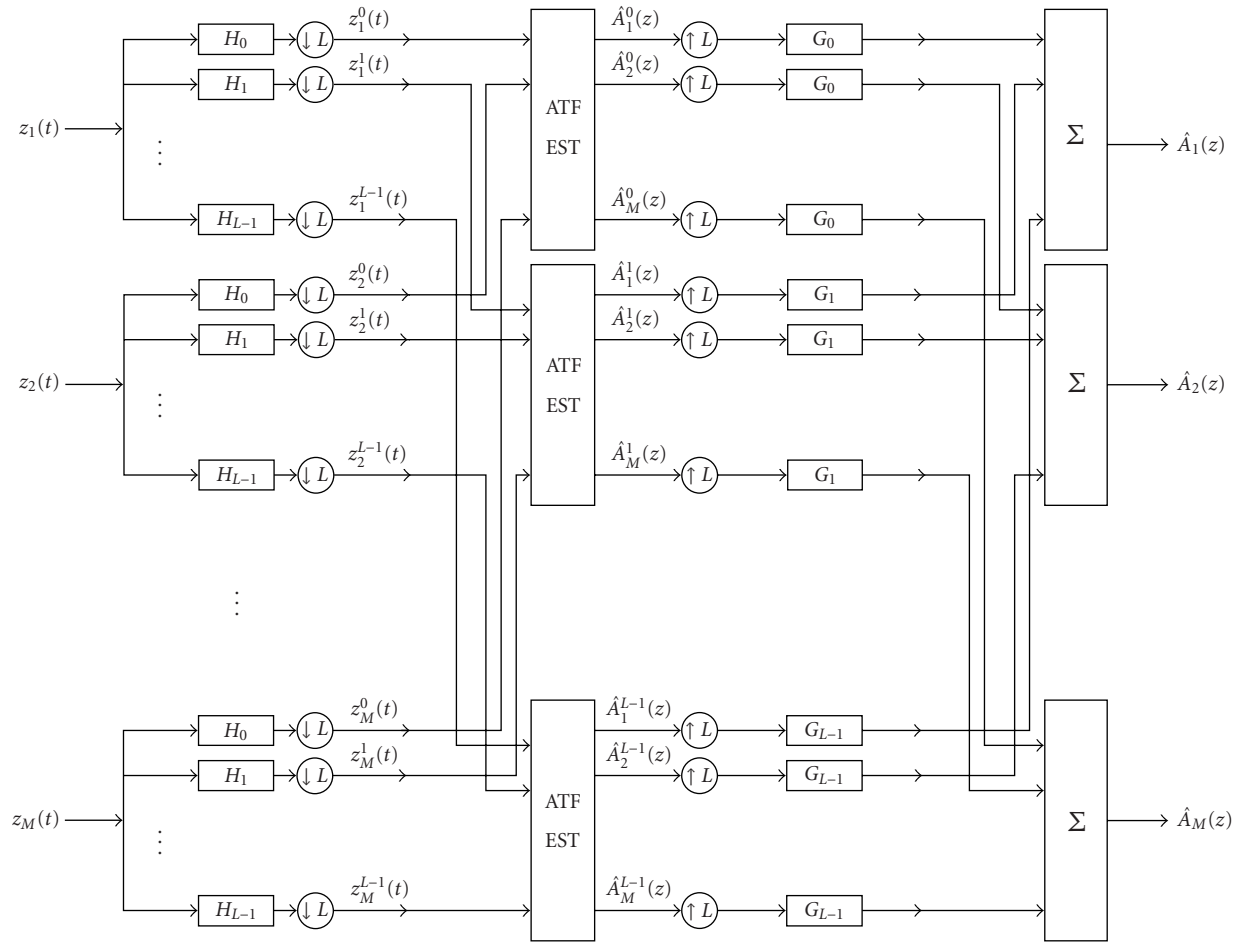


FIGURE 5: Null subspace in the two-microphone noiseless case.

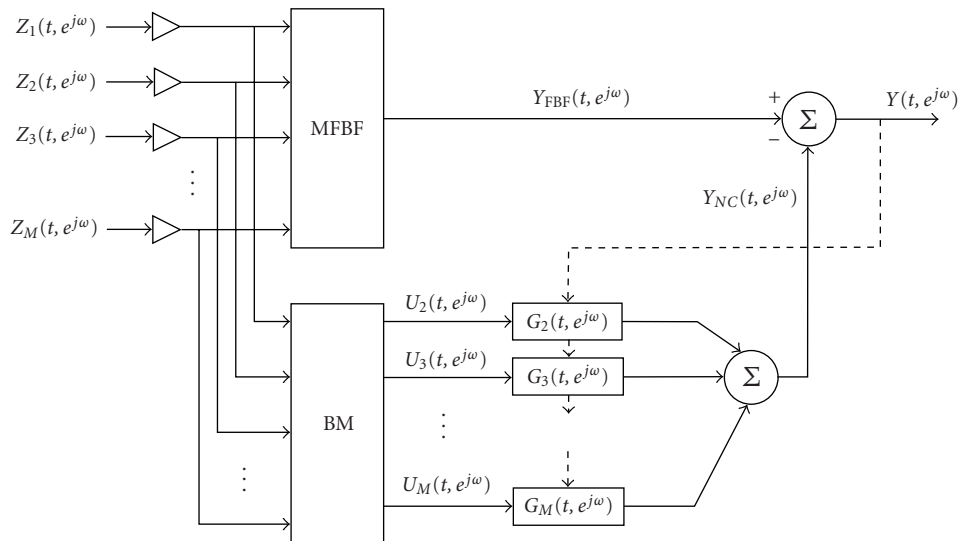


FIGURE 6: Extended GSC structure for joint noise reduction and dereverberation.

- (1) *Estimate ATFs:* $A_m(e^{j\omega})$, $m = 1, 2, \dots, M$.
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) $\mathbf{W}_0(t, e^{j\omega}) = \frac{\mathbf{A}(t, e^{j\omega})}{\|\mathbf{A}(t, e^{j\omega})\|^2}$.
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 $m = 1, \dots, M - 1$:
 $\tilde{G}_m(t+1, e^{j\omega}) = G_m(t, e^{j\omega}) + \mu \frac{U_m(t, e^{j\omega})Y^*(t, e^{j\omega})}{P_{\text{est}}(t, e^{j\omega})}$
 $G_m(t+1, e^{j\omega}) \xrightarrow{\text{FIR}} \tilde{G}_m(t+1, e^{j\omega})$
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 nonaliased samples, according to the *overlap and save* method [12].

ALGORITHM 1: Summary of the TF-GSC algorithm.

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

$$J_K = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & \dots & & 0 \\ 0 & \ddots & 0 & \dots & 0 \\ \vdots & 0 & \ddots & & \vdots \\ 0 & \dots & & 1 & 0 \end{bmatrix}. \quad (34)$$

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 an 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 QRD 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 k th column (counting from the rightmost 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,

$$\mathcal{G}^R = J_{M(\hat{n}_a+1)}^k \mathcal{G} = J_{M(\hat{n}_a+1)}^k \mathcal{A} \mathcal{E}. \quad (35)$$

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

$$\mathcal{G}^R = J_{M(\hat{n}_a+1)}^k \mathcal{A} (J_{\hat{n}_a-n_a+1}^k)^T J_{\hat{n}_a-n_a+1}^k \mathcal{E}. \quad (36)$$

Then, applying the transpose operation,

$$(\mathcal{G}^R)^T = (J_{\hat{n}_a-n_a+1}^k \mathcal{E})^T (J_{M(\hat{n}_a+1)}^k \mathcal{A} (J_{\hat{n}_a-n_a+1}^k)^T)^T. \quad (37)$$

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

$$(J_{\hat{n}_a-n_a+1}^k \mathcal{E})^T = \text{QR}. \quad (38)$$

Then,

$$(\mathcal{G}^R)^T = \text{QR} \left(J_{M(\hat{n}_a+1)}^k \mathcal{A} (J_{\hat{n}_a-n_a+1}^k)^T \right)^T = Q\tilde{R}. \quad (39)$$

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 k th column (counting from the rightmost column) \mathcal{A}^T , provided that $k \leq \hat{n}_a - n_a + 1$ and it is still an upper triangular matrix. Thus, the same statements regarding the nonrotated matrices apply for the rotated matrices. \square

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 estimates can be used to obtain a more robust estimate.

4.2. 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,

$$\check{\mathcal{G}} = \mathcal{A} \check{\mathcal{E}}, \quad (40)$$

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 4.1 is not applicable anymore.

Let

$$\check{\mathcal{G}}^T = \begin{bmatrix} (\tilde{\mathbf{a}}_{1,0})^T & (\tilde{\mathbf{a}}_{2,0})^T & \dots & (\tilde{\mathbf{a}}_{M,0})^T \\ (\tilde{\mathbf{a}}_{1,1})^T & (\tilde{\mathbf{a}}_{2,1})^T & \dots & (\tilde{\mathbf{a}}_{M,1})^T \end{bmatrix}. \quad (41)$$

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

$\tilde{\mathcal{G}}^T$ will yield the following $R_{\tilde{\mathcal{G}}}$ matrix:

$$R_{\tilde{\mathcal{G}}} = \begin{bmatrix} \cdots & \cdots \\ [0(\tilde{\mathbf{a}}'_{1,1})^T] & [0(\tilde{\mathbf{a}}'_{2,1})^T] & \cdots & [0(\tilde{\mathbf{a}}'_{M,1})^T] \end{bmatrix}. \quad (42)$$

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 $R_{\tilde{\mathcal{G}}}$ 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. This set may be obtained in several ways. One possibility (although others are also applicable) is to rotate each part of $\tilde{\mathcal{G}}$, that is, $\tilde{\mathbf{a}}_{m,l}$, downwards and apply the QRD again. After this two-steps stage, we obtain a shorter null subspace

$$\tilde{\mathcal{G}}'^T = \begin{bmatrix} (\tilde{\mathbf{a}}'_{1,0})^T & (\tilde{\mathbf{a}}'_{2,0})^T & \cdots & (\tilde{\mathbf{a}}'_{M,0})^T \\ (\tilde{\mathbf{a}}'_{1,1})^T & (\tilde{\mathbf{a}}'_{2,1})^T & \cdots & (\tilde{\mathbf{a}}'_{M,1})^T \end{bmatrix}. \quad (43)$$

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].

5. 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 [13]. 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 incorporating 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 [14, 15, 16, 17]. However, the use of subbands in subspace methods is not that 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 eight-channel subband structure was used as depicted in Figure 4. Each of the channel filters is an FIR filter of order 150. The filters are equispaced 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 signal corresponds to a longer filter (which

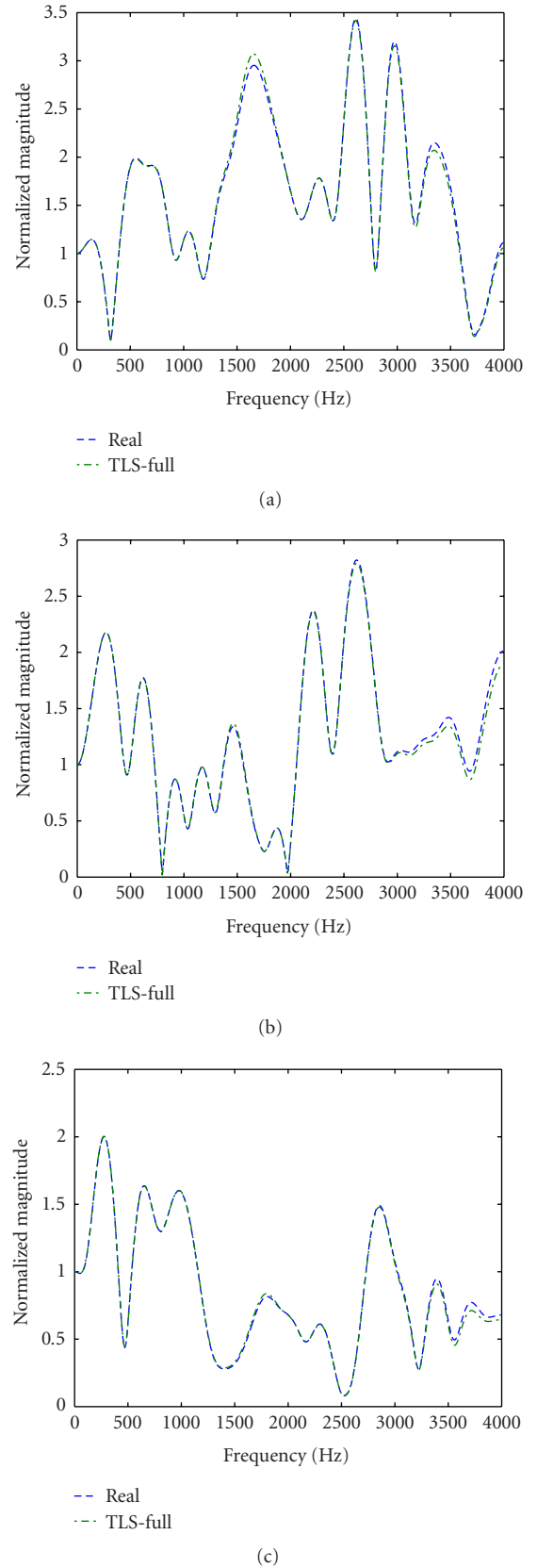
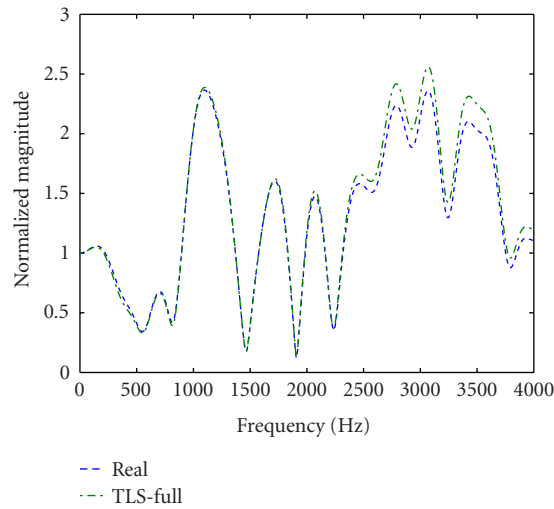
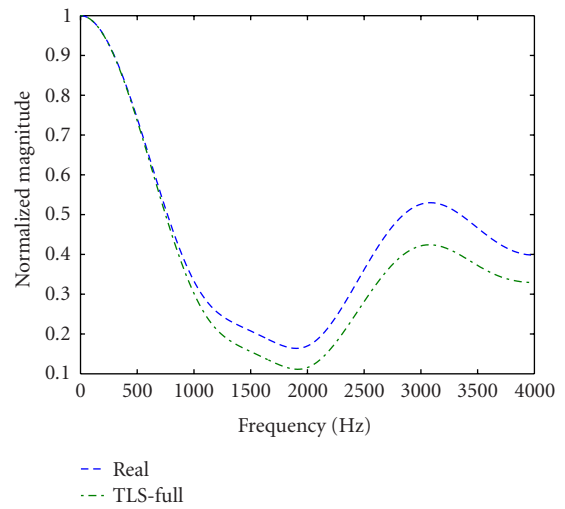


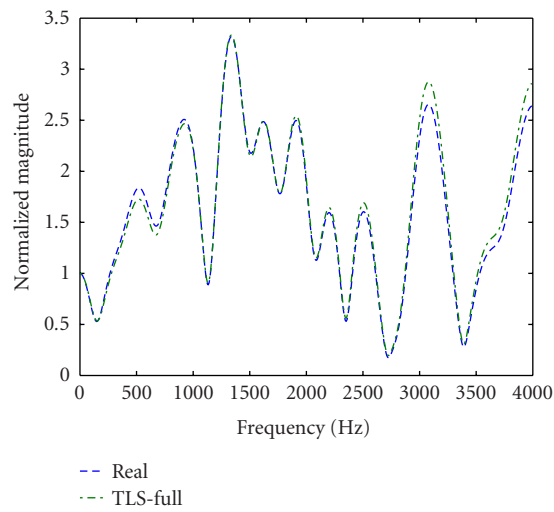
FIGURE 7: Real and estimated frequency response of an ATF with exponential decaying envelope of order 32. Results with speech-like noise input using fullband method at SNR = 45 dB.



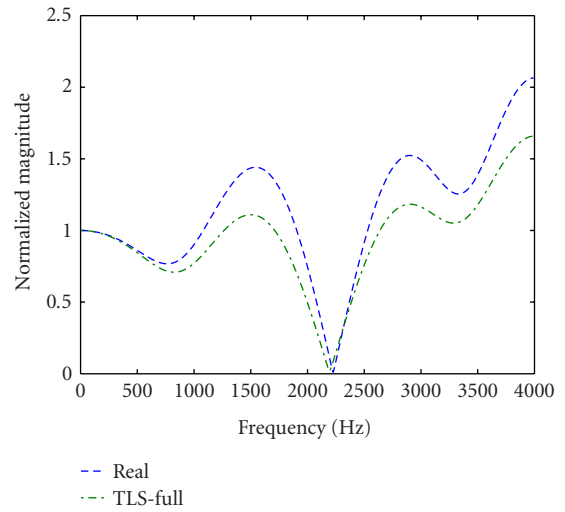
(a)



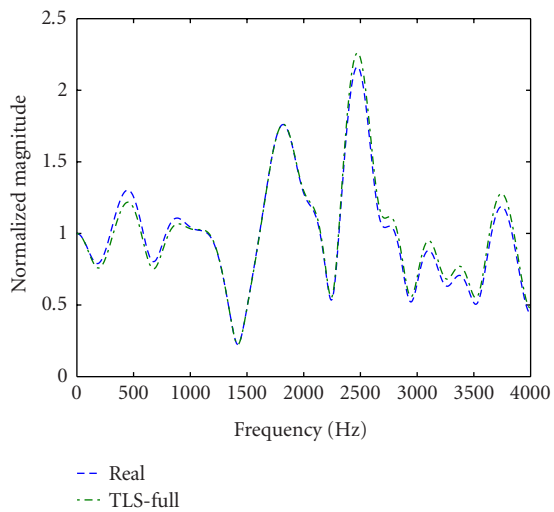
(a)



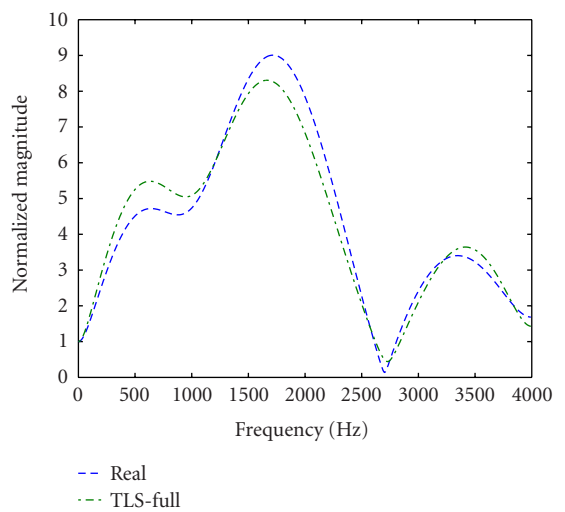
(b)



(b)



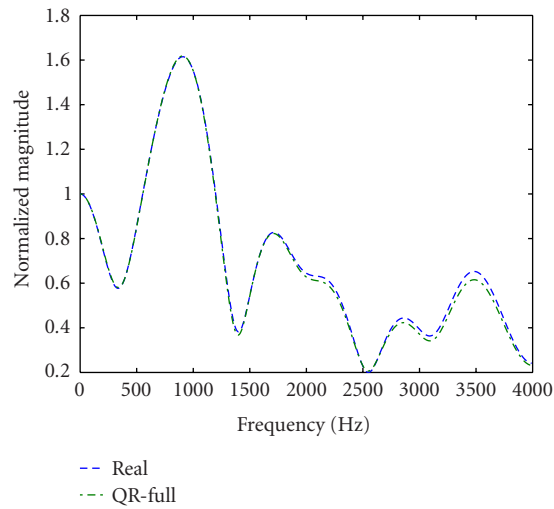
(c)



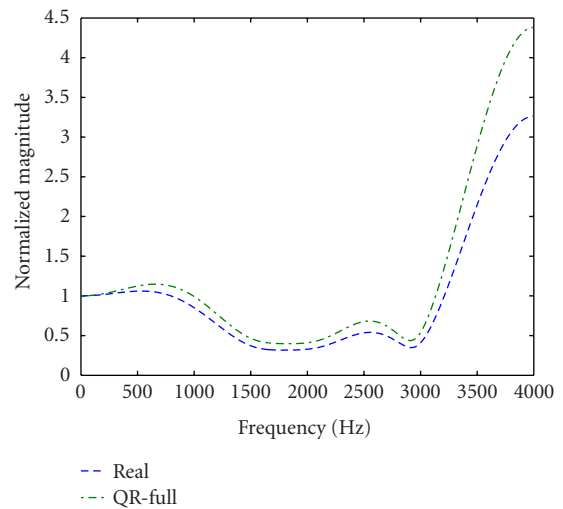
(c)

FIGURE 8: Real and estimated frequency response of an ATF with exponential decaying envelope of order 32. Results with white noise input using fullband method at SNR = 35 dB.

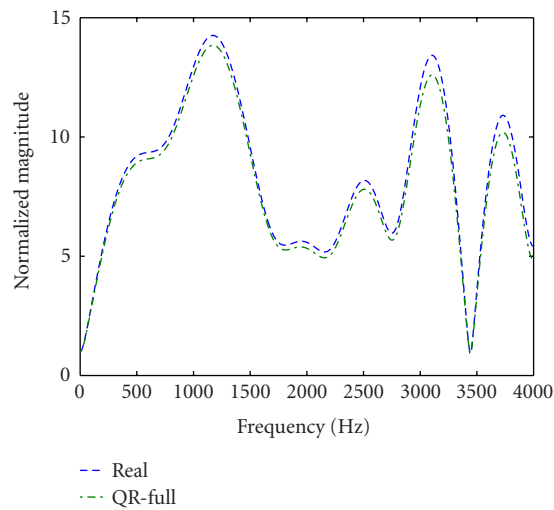
FIGURE 9: Real and estimated frequency response of an ATF with exponential decaying envelope of order 8. Results with white noise input using fullband method at SNR = 25 dB.



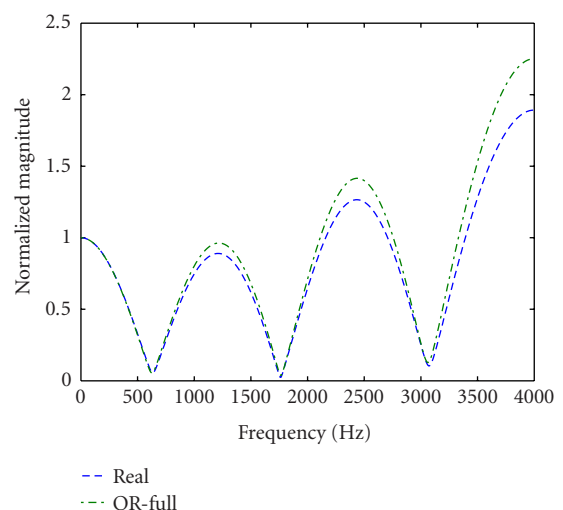
(a)



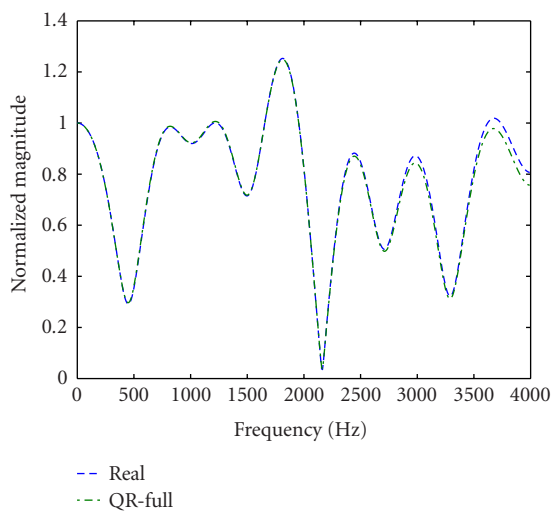
(a)



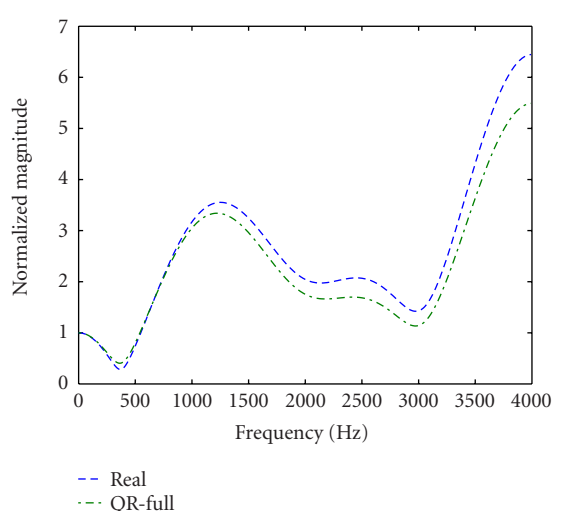
(b)



(b)



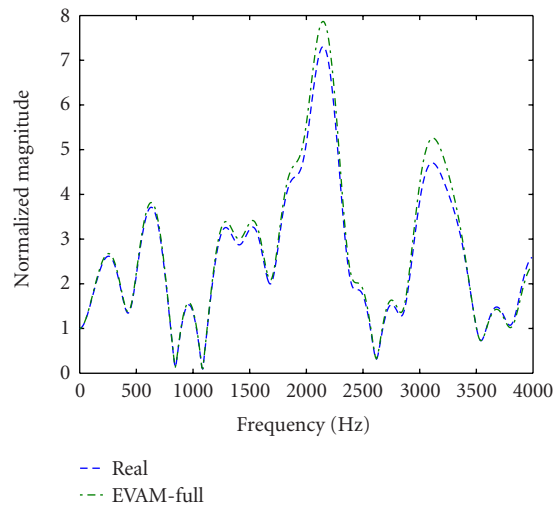
(c)



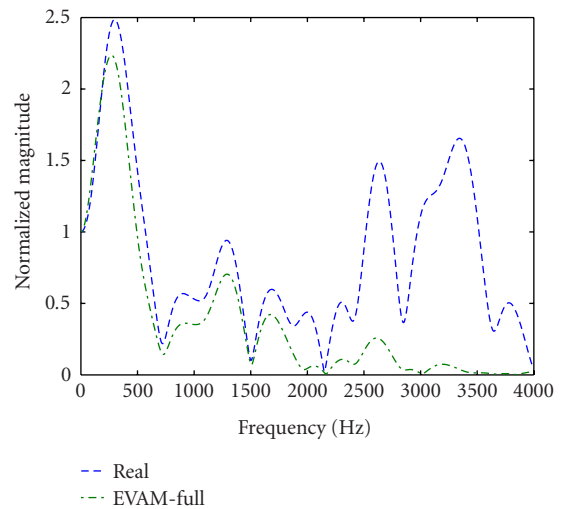
(c)

FIGURE 10: Real and estimated frequency response of an ATF with exponential decaying envelope of order 16. Results with speech-like noise input using QRD method at SNR = 45 dB.

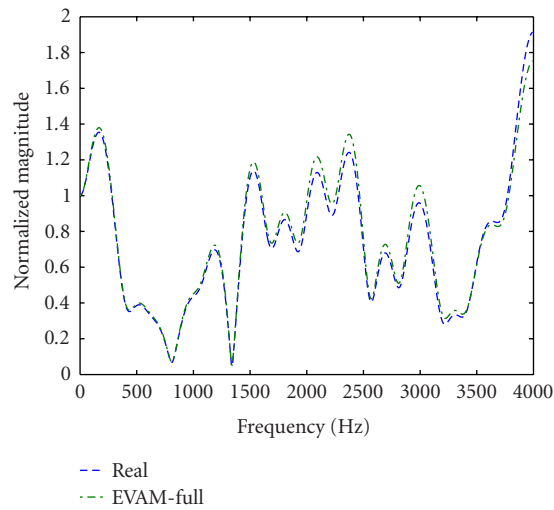
FIGURE 11: Real and estimated frequency response of an ATF with exponential decaying envelope of order 8. Results with speech-like noise input using QRD method at SNR = 35 dB.



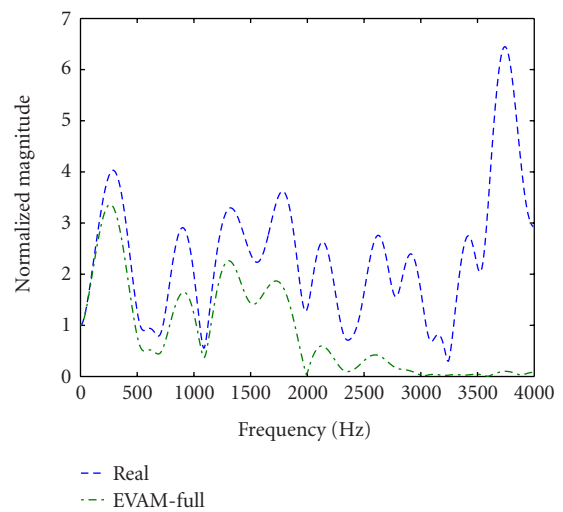
(a)



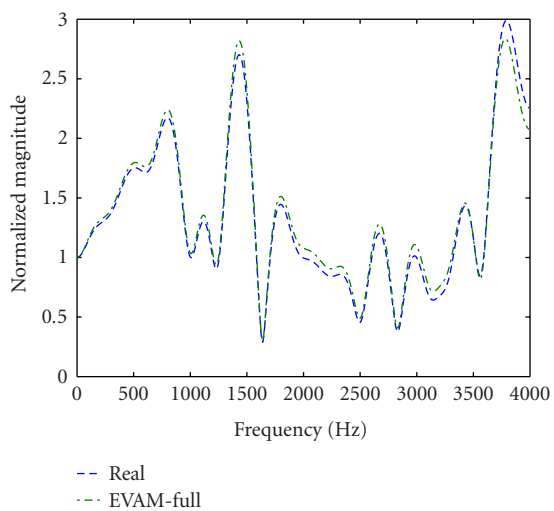
(a)



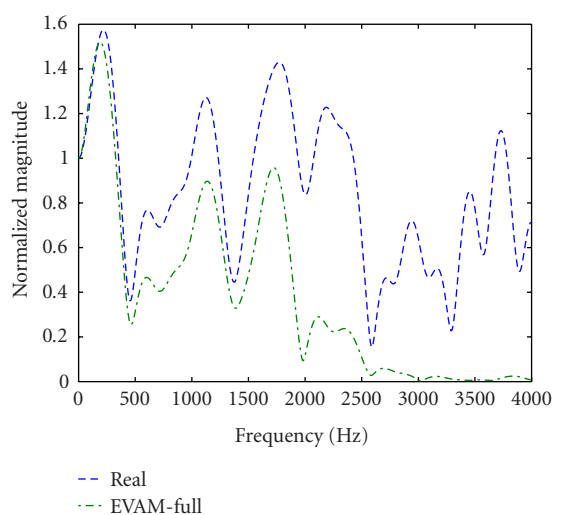
(b)



(b)



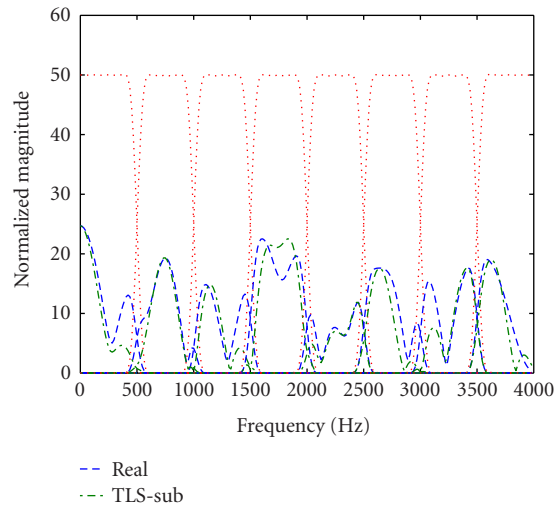
(c)



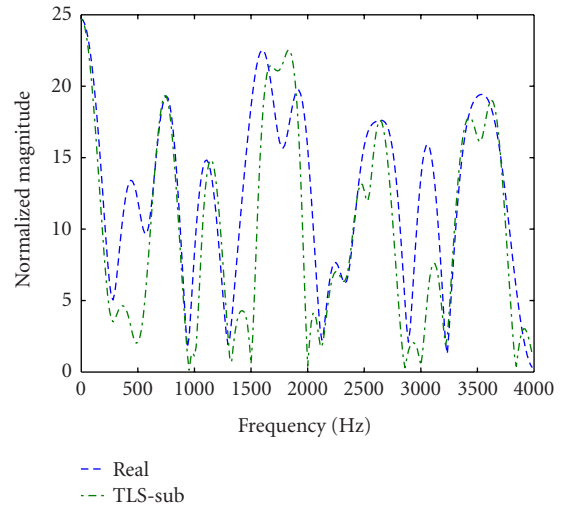
(c)

FIGURE 12: Real and estimated frequency response of an ATF with exponential decaying envelope of order 32. Results with speech-like input using EVAM method at SNR = 45 dB.

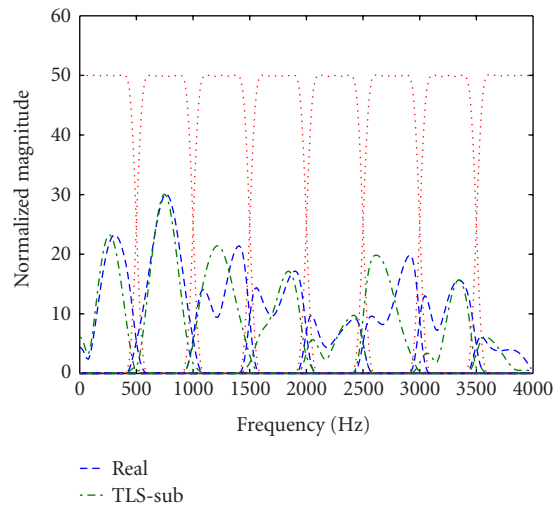
FIGURE 13: Real and estimated frequency response of an ATF with exponential decaying envelope of order 32. Results with speech-like input using EVAM method at SNR = 35 dB.



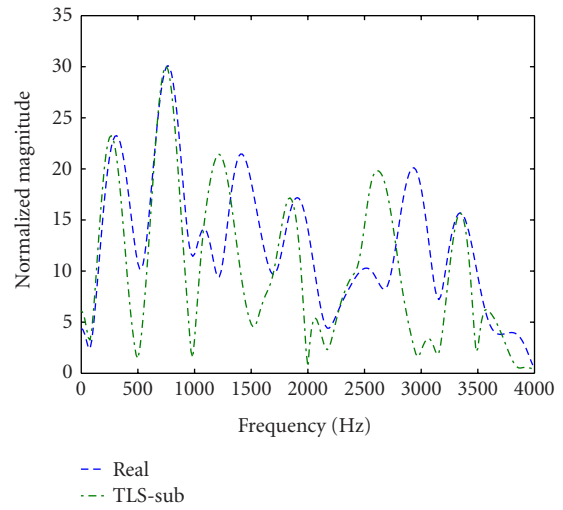
(a)



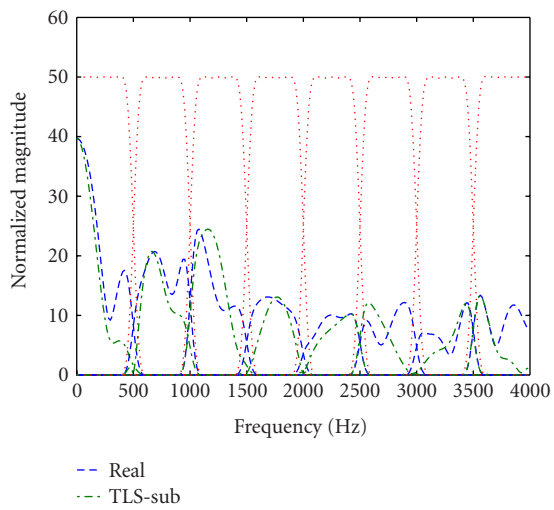
(a)



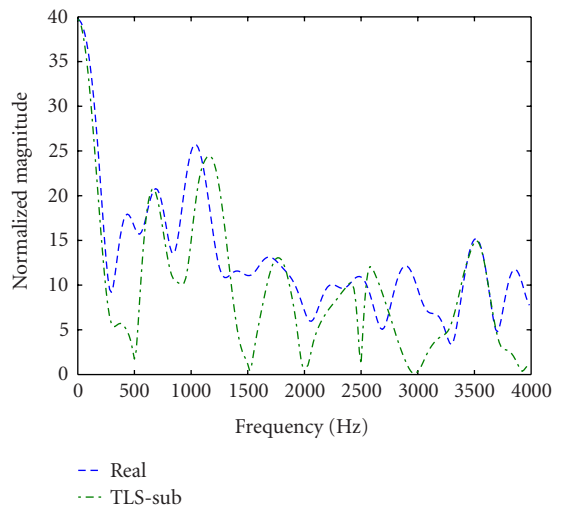
(b)



(b)



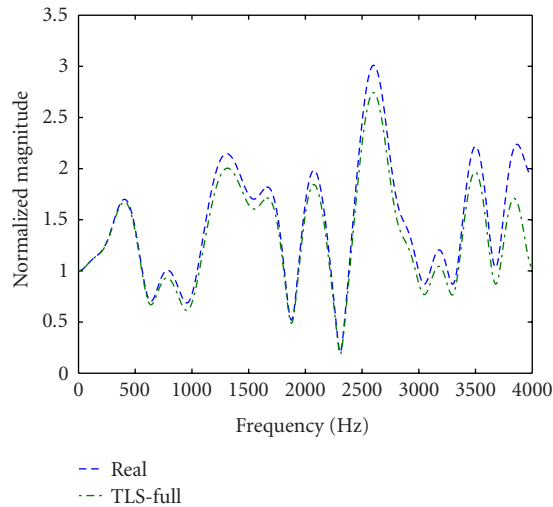
(c)



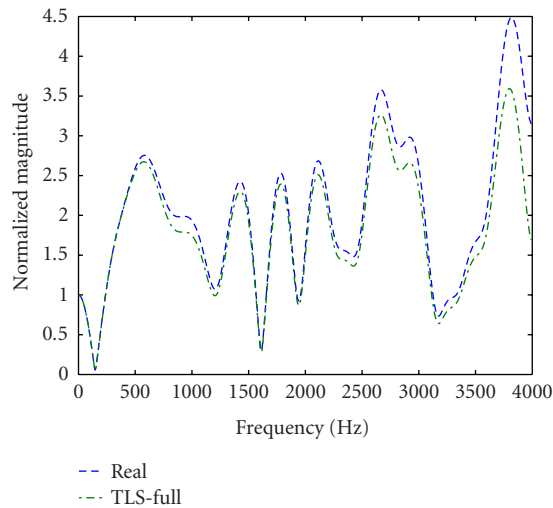
(c)

FIGURE 14: Real and estimated frequency response of an ATF with exponential decaying envelope of order 32. Results with speech-like input using subband method (separate bands) at SNR = 25 dB.

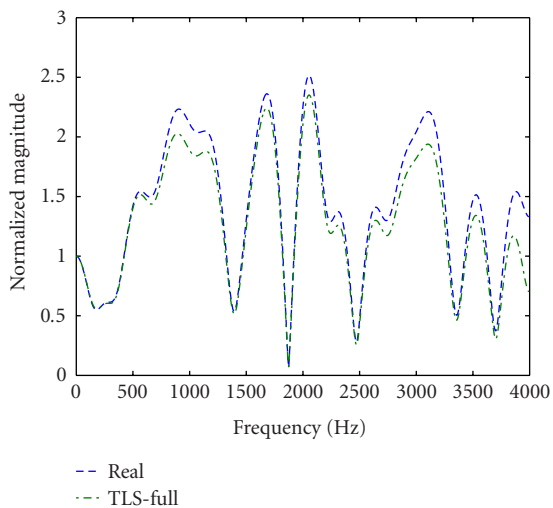
FIGURE 15: Real and estimated frequency response of an arbitrary ATF with exponential decaying envelope of order 32. Results with speech-like input using subband method (combined bands) at SNR = 25 dB.



(a)



(b)



(c)

FIGURE 16: Real and estimated frequency response of an arbitrary ATF of order 32. Results with speech signal input using fullband method at SNR = 50 dB.

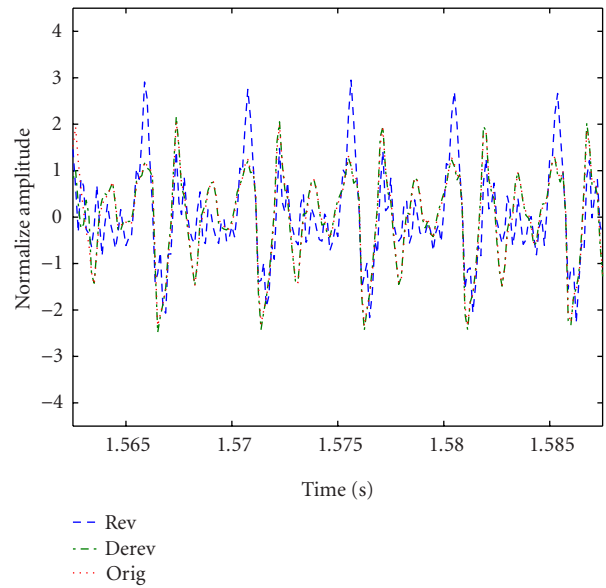


FIGURE 17: Dereverberated speech signal due to the filtering by the ATF of order 32 with speech signal input using subband method at SNR = 50 dB.

is the convolution of the corresponding ATF and the subband filter), the algorithm aims at reconstructing 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 3. For properly exploiting the benefits of the subband structure, each subband signal should be decimated. We took a critically decimated filterbank, that is, decimation factor equals the number of bands. By doing so, the ATF order in each band is reduced by approximately the decimation factor, making the estimation task easier. Note that now we need only to overestimate 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 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 Figure 5, where the *ATF estimation* block is shown schematically in Figure 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 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) are 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 as an approximate gain adjustment.

Another method was suggested by Rahbar et al. [18]. 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 using the ATFs order estimation obtained by the subspace method. The use of this method is a topic of further research. As far as this paper goes, the gain ambiguity problem is ignored, and we will assume that the gain in each subband is known. Thus we would only demonstrate the ability of the method to estimate the frequency shaping of the method in each band.

6. 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 multichannel FIR filters by a multichannel set of FIR filters is presented in [19].

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

$$Z_m(t, e^{j\omega}) = A_m(t, e^{j\omega})S(t, e^{j\omega}) + N_m(t, e^{j\omega}). \quad (44)$$

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

$$\begin{aligned} \hat{S}(t, e^{j\omega}) &= \frac{1}{\sum_{m=1}^M |\hat{A}_m(t, e^{j\omega})|^2} \sum_{m=1}^M Z_m(t, e^{j\omega}) \hat{A}_m^*(t, e^{j\omega}) \\ &= S(t, e^{j\omega}) \frac{1}{\sum_{m=1}^M |\hat{A}_m(t, e^{j\omega})|^2} \\ &\quad \times \sum_{m=1}^M A_m(t, e^{j\omega}) \hat{A}_m^*(t, e^{j\omega}) \\ &\quad + \frac{1}{\sum_{m=1}^M |\hat{A}_m(t, e^{j\omega})|^2} \sum_{m=1}^M N_m(t, e^{j\omega}) \hat{A}_m^*(t, e^{j\omega}). \end{aligned} \quad (45)$$

It is easily verified that if the estimation of the ATFs is sufficiently accurate, that is, $\hat{A}_m(t, e^{j\omega}) \simeq A_m(t, e^{j\omega})$, then the first term in (45) 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 incorporating the procedure into the recently proposed extended GSC, derived by Gannot et al. [20], shown schematically in Figure 6 and summarized in Algorithm 1. 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 MFBF described in (45), a noise reference construction block—which uses the ATFs ratios (note that $U_m(t, e^{j\omega})$ contain only noise terms), and a multichannel 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 MFBF branch produces sufficiently accurate estimate.

7. EXPERIMENTAL STUDY

The validity of the proposed methods was tested using various input signals and ATFs modelled as an FIR filters with exponential decaying envelope, and compared with the EVAM algorithm [9]. This input signal consisted of either white noise, speech-like noise (white signal colored to have a speech-like spectrum, drawn from the NOISEX-92 database [21]), or a real speech signal comprised of a concatenation of several speech signals drawn from the TIMIT database [22]. The input signal was 32000 samples long (corresponding to 4 seconds for the 8-kHz sampled speech signal, including silence periods). Three-microphone signals were simulated by filtering the input signal by various ATFs. In most applications, the order of realistic ATFs is at least several hundred taps. However, our method (as well as other methods in the literature) is incapable of dealing with this order. Thus, we settled for short ATFs of order of either 8, 16, or 32. To approximate more realistic scenario, we used ATFs with exponential decaying envelope. Various SNR levels were taken to test the robustness of the algorithms to additive noise. Temporally nonwhite 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 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 is depicted in Figures 7, 8, and 9 for SNR levels of 45 dB, 35 dB, and 25 dB, respectively. 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 32 (Figure 7), when the SNR decreases to 35 dB, the performance is maintained only with a white noise input signal (Figure 8). For SNR level of 25 dB, the algorithm fails to work with the longer filter and its order had to be reduced to only 8 (Figure 9). The sensitivity to the noise level is thus clearly indicated.

Results for the suboptimal QRD-based algorithm are depicted in Figure 10 for the speech-like input and an SNR level of 45 dB, and in Figure 11 for a white noise input and an SNR level of 35 dB. The QRD-based method is more sensitive to the noise level. At 45 dB, good results could be obtained only with filter of order 16, and at 35 dB, only with 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 12 and 13 for SNR levels of 45 dB and 35 dB, respectively. It is clearly shown 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.

The incorporation of the subspace method into a subband structure is depicted in Figures 14 and 15. We used the eight-subband structure shown in Figure 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 Figure 14, the subband structure is depicted, and the estimated response is given in each band separately. In Figure 15, all the bands are combined to form the entire frequency response of the ATFs. The results demonstrate the ability of the algorithm to work well at lower SNR levels (25 dB) 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 should be explored indepth 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 (50 dB) and $n_a = 32$. The frequency response of the estimated filter is depicted in Figure 16 for the fullband method. The dereverberated speech signal is shown in Figure 17. It is clearly shown from the figure, that while the microphone signal 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 12 dB decrease in the power of the difference.

8. 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 QRD. 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 ATF 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. Unfortunately,

this issue is left for further research. However, 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 more realistic and much longer.

ACKNOWLEDGMENTS

This 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 *Multimicrophone Signal Enhancement Techniques for hands-free telephony and voice-controlled systems* (MUSETTE-2) of the IWT, and was partially sponsored by Philips ITCL.

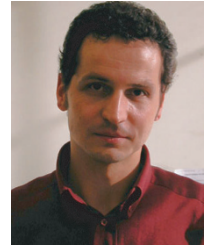
REFERENCES

- [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, no. 4, pp. 317–334, 1996.
- [2] J. Gonzalez-Rodriguez, J. L. Sanchez-Bote, and J. Ortega-Garcia, "Speech dereverberation and noise reduction with a combined microphone array approach," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '00)*, vol. 2, Istanbul, Turkey, June 2000.
- [3] B. Yegnanarayana, P. Satyanarayana Murthy, C. Avendano, and H. Hermansky, "Enhancement of reverberant speech using LP residual," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '98)*, vol. 1, Seattle, Wash, USA, May 1998.
- [4] S. M. Griebel and M. S. Brandstein, "Microphone array speech dereverberation using coarse channel modeling," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '01)*, vol. 1, Salt Lake City, Utah, USA, May 2001.
- [5] S. Affes and Y. Grenier, "A signal subspace tracking algorithm for microphone array processing of speech," *IEEE Trans. Speech, and Audio Processing*, vol. 5, no. 5, pp. 425–437, 1997.
- [6] S. Doclo and M. Moonen, "Combined frequency-domain dereverberation and noise reduction technique for multimicrophone speech enhancement," in *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC '01)*, Darmstadt, Germany, September 2001.
- [7] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, no. 2, pp. 516–525, 1995.
- [8] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Processing*, vol. 43, no. 12, pp. 2982–2993, 1995.
- [9] M. İ. Gürelli and C. L. Nikias, "EVAM: an eigenvector-based algorithm for multichannel blind deconvolution of input colored signals," *IEEE Trans. Signal Processing*, vol. 43, no. 1, pp. 134–149, 1995.
- [10] S. Gannot and M. Moonen, "Subspace methods for multimicrophone speech dereverberation," in *Proc. International*

Workshop on Acoustic Echo and Noise Control (IWAENC '01), Darmstadt, Germany, September 2001.

- [11] S. Van Huffel, H. Park, and J. B. Rosen, "Formulation and solution of structured total least norm problems for parameter estimation," *IEEE Trans. Signal Processing*, vol. 44, no. 10, pp. 2464–2474, 1996.
- [12] R. E. Crochiere, "A weighted overlap-add method of short-time fourier analysis/synthesis," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 28, no. 1, pp. 99–102, 1980.
- [13] G. H. Golub and C. F. Van Loan, *Matrix Computations*, The John Hopkins University Press, Baltimore, Md, USA, 2nd edition, 1989.
- [14] S. Weiss, R. W. Stewart, A. Stenger, and R. Rabenstein, "Performance limitations of subband adaptive filters," in *Proc. 9th European Signal Processing Conference (EUSIPCO '98)*, pp. 1245–1248, Island of Rhodes, Greece, September 1998.
- [15] S. Weiss, G. W. Rice, and R. W. Stewart, "Multichannel equalization in subbands," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, USA, October 1999.
- [16] A. Spriet, M. Moonen, and J. Wouters, "A multichannel subband GSVD based approach for speech enhancement in hearing aids," in *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC '01)*, Darmstadt, Germany, September 2001.
- [17] K. Eneman and M. Moonen, "DFT modulated filter bank design for oversampled subband systems," *Signal Processing*, vol. 81, no. 9, pp. 1947–1973, 2001.
- [18] 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 *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '02)*, Orlando, Fla, USA, May 2002.
- [19] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 36, no. 2, pp. 145–152, 1988.
- [20] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and non-stationarity with applications to speech," *IEEE Trans. Signal Processing*, vol. 49, no. 8, pp. 1614–1626, 2001.
- [21] 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 Communication*, vol. 12, no. 3, pp. 247–251, 1993.
- [22] National Institute of Standards and Technology, *The DARPA TIMIT acoustic-phonetic continuous speech corpus*, CD-ROM NIST Speech Disc 1-1.1, October 1990.

Marc Moonen received the Electrical Engineering degree and the Ph.D. degree in applied sciences from the Katholieke Universiteit Leuven, Leuven, Belgium, in 1986 and 1990, respectively. Since 2000, he has been an Associate Professor at the Electrical Engineering Department of Katholieke Universiteit Leuven, where he is currently heading a research team of sixteen Ph.D. candidates and postdocs, working in the area of signal processing for digital communications, wireless communications, DSL and audio signal processing. He received the 1994 KULeuven Research Council Award, the 1997 Alcatel Bell (Belgium) Award (with Piet Vandaele), and was a 1997 "Laureate of the Belgium Royal Academy of Science." He was the Chairman of the IEEE Benelux Signal Processing Chapter (1998–2002), and is currently a EURASIP AdCom Member (European Association for Signal, Speech, and Image Processing, 2000). He is Editor-in-Chief for the EURASIP Journal on Applied Signal Processing (2003), and a member of the editorial board of *Integration*, the *VLSI Journal*, *IEEE Transactions on Circuits and Systems II*, and *IEEE Signal Processing Magazine*.



Sharon Gannot received his B.S. degree (summa cum laude) from the Technion – Israel Institute of Technology, Israel in 1986 and the M.S. (cum laude) and Ph.D. degrees from Tel Aviv University, Tel Aviv, Israel in 1995 and 2000, respectively, all in electrical engineering. Between 1986 and 1993, he was the Head of a research and development section in R&D center of the Israel Defense Forces. In 2001, he held a postdoctoral position at the Department of Electrical Engineering (SISTA) at Katholieke Universiteit Leuven, Belgium. From 2002 to 2003, he held a research and teaching position at the Signal and Image Processing Lab (SIPL), Faculty of Electrical Engineering, The Technion – Israel Institute of Technology, Israel. Currently, he is affiliated with the School of Engineering, Bar-Ilan University, Israel.

