



UNIVERSITÀ  
DEGLI STUDI  
DI UDINE

## Università degli studi di Udine

### A Low-Complexity Robust Beamforming Using Diagonal Unloading for Acoustic Source Localization

*Original*

*Availability:*

This version is available <http://hdl.handle.net/11390/1123492> since 2018-02-08T10:25:26Z

*Publisher:*

*Published*

DOI:10.1109/TASLP.2017.2789321

*Terms of use:*

The institutional repository of the University of Udine (<http://air.uniud.it>) is provided by ARIC services. The aim is to enable open access to all the world.

*Publisher copyright*

(Article begins on next page)

# A Low Complexity Robust Beamforming Using Diagonal Unloading for Acoustic Source Localization

Daniele Salvati, Carlo Drioli, *Member, IEEE*, and Gian Luca Foresti, *Senior Member, IEEE*

**Abstract**—In acoustic array processing, beamforming is a class of algorithms commonly used to estimate the position of a radiating sound source. This paper presents a diagonal unloading (DU) transformation method for the conventional response power beamforming to achieve robust localization with low computational complexity. The transformation is obtained by subtracting an opportune diagonal matrix from the covariance matrix of the array output vector. Specifically, the DU beamformer aims at subtracting the signal subspace from the noisy signal space. It is hence a data-dependent covariance matrix conditioning method. We show how to calculate precisely the unloading parameters, and we present a comparison of the proposed DU beamforming, the robust minimum variance distortionless response (MVDR) filter and the multiple signal classification (MUSIC) method, in terms of their respective eigenanalyses. Theoretical analysis and experiments conducted on both simulated and real acoustic data demonstrate that the DU beamformer localization performance is comparable to that of robust MVDR and MUSIC. Since its computational cost is equivalent to that of a conventional beamformer, the proposed DU beamformer method can thus be very attractive due to its effectiveness and computational efficiency.

**Index Terms**—Diagonal unloading beamforming, acoustic source localization, direction of arrival estimation, broadband robust beamforming, acoustic analysis, microphone array.

## I. INTRODUCTION

ACOUSTIC source localization is an important task in microphone array processing and it is of interest to numerous applications including teleconferencing systems, audio surveillance, autonomous robots, human-computer interaction and acoustic monitoring [1]–[10]. In general, the localization can be performed by indirect and direct methods. The indirect (two-step) approach computes a set of time difference of arrivals (TDOAs) using measurements across various combinations of microphones [11], [12], and then estimates the source position using geometric considerations [13]–[15]. Direct methods are based on the steered response

power (SRP) beamformers [16]–[23], on subspace algorithms [24]–[26], or on maximum-likelihood estimators [27]–[29].

Beamforming is very attractive in acoustic applications due to its robustness in noisy and reverberant conditions. A beamformer is a spatial filter whose goal is to achieve directional signal reception, to separate acoustic signals originating from different spatial positions. The conventional data-independent beamformer [30] is based on a delay-and-sum procedure, which has its roots in time-series analysis. Among conventional beamformers, the minimum variance distortionless response (MVDR) [31] filter is a well-known data-dependent beamformer that has a better resolution than the data-independent beamformer. However, its localization performance is not robust in most practical situations, since the spatial spectrum might be deteriorated by steering vector errors and discrete sampling effects. Therefore, a robust variant of the MVDR filter, obtained with regularization techniques [32], is often preferred. A popular approach is the class of diagonal loading (DL) robust MVDR techniques [33]–[38]. Another popular robust localization method is the multiple signal classification (MUSIC) [24], which exploits the subspace orthogonality property to build the spatial spectrum and to localize the sources. It can be interpreted as a beamformer that uses the noise subspaces of the covariance matrix in the computation of the SRP. Several robust MVDR-based and MUSIC-based algorithms can be found in [39]–[44] and in [45]–[49], respectively. The family of robust localization algorithms provides a better resolution and a better rejection of energy components from other directions if compared to the conventional beamformer, at the cost however of an increase of complexity. MVDR and MUSIC methods require a full-rank inversion matrix the former, and an eigendecomposition the latter. For both methods, the complexity is  $O(n^3)$ , which leads to a relevant computational cost in broadband applications.

In this paper, we focus on far-field source direction of arrival (DOA) estimation and we propose a low complexity and robust beamformer based on a diagonal unloading (DU) transformation of the covariance matrix involved in the conventional beamformer computations. The transformation is data-dependent and it is obtained by subtracting an opportune diagonal matrix from the covariance matrix of the array output vector. The DU beamforming aims at removing as much as possible the signal subspace from the covariance matrix to design a high resolution beampattern. This goal will be achieved through a DU transformation of the covariance matrix, whose objective is to set to zero the eigenvalue corresponding to the

Copyright (c) 2016 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

D. Salvati, C. Drioli, and G.L. Foresti are with the Department of Mathematics, Computer Science and Physics, University of Udine, Udine 33100, Italy, e-mail: daniele.salvati@uniud.it, carlo.drioli@uniud.it, gianluca.foresti@uniud.it. The work was partially supported by Regione Friuli Venezia-Giulia under the "Proactive Vision for advanced UAV systems for the protection of mobile units, control of territory and environmental prevention (SUPReME)" FVG L.R. 20/2015 project and by the Italian Ministry of Defence under the "PREscriptive Situational awareness for cooperative autoorganizing aerial sensor NETworks (PRESNET)" MoD 1889 12/2016 project.

signal subspace. This constraint determines exactly how to calculate the penalty weight for the DU operation. Furthermore, this DU procedure leads to a negative semidefinite transformed covariance matrix, which allows to put the proposed DU beamforming in a form that is comparable to that of the MVDR and MUSIC formulation. In this paper, we also provide a comparison analysis between DU, robust DL MVDR and MUSIC by taking into account their transformed covariance matrices and their eigenvalue decomposition, and it is shown that the proposed DU beamforming achieves state-of-the-art robust beamforming localization with significant reduction of the computational complexity. Hence, the advantage of the DU beamforming is that the computational cost is equivalent to that of the conventional beamforming, which has  $O(n)$  complexity, whereas the localization performance and the spatial resolution is comparable to that of the regularized MVDR and MUSIC methods.

In summary, the objective of this paper is twofold: 1) to develop a novel form of low complexity and robust beamformer, which has high resolution and data-dependent properties, by applying an appropriate diagonal unloading procedure to the conventional spatial filter; 2) to provide an eigenvalue analysis, which highlights the relationship between DU, robust MVDR and MUSIC by taking into account their transformed covariance matrices.

The paper is organized as follows. Section II provides the definition of the data model and of the broadband spatial filters. The DU beamforming is then described in Section III. We examine the importance of the attenuation of the signal subspace by discussing the eigendecomposition of the PSD matrix and the orthogonality property. We introduce the proposed beamformer by considering the single source case, and then an analysis in a multisource scenario is described. Next, we introduce a suboptimal implementation. Section IV provides an eigenvalue analysis of the covariance matrix and of the transformed covariance matrices in the DU beamformer, the MVDR filter, and the MUSIC method. We provide some properties for evaluating the performance of the proposed method in comparison with high resolution robust beamforming. Section V provides a computational complexity analysis. Experiments using artificially-generated and real-world signals are shown in Section VI. The conclusions are drawn in Section VII.

## II. BACKGROUND

### A. Notation

In what follows, we will make use of standard notational conventions.  $\mathbb{R}$  and  $\mathbb{C}$  denote the sets of all real and complex numbers respectively. Vectors and matrices are written in boldface with matrices in capitals. For a random matrix  $\mathbf{X}$ ,  $E\{\mathbf{X}\}$  denotes the expectation of  $\mathbf{X}$ . For a matrix  $\mathbf{A}$ ,  $\mathbf{A}^T$ ,  $\mathbf{A}^H$ , and  $\text{tr}(\mathbf{A})$  denote the transpose, the conjugate transpose, and the trace, i.e. the sum of diagonal elements of  $\mathbf{A}$ , respectively. The identity matrix of any size is denoted by  $\mathbf{I}$ . The symbol  $*$  stands for convolution.

### B. Data Model

Suppose that a single source impinges upon an array of  $N$  sensors and let  $s(t) \in \mathbb{R}$  denote the signal generated by a nonstationarity broadband source at time  $t$ . The output of the  $n$ th ( $n = 1, 2, \dots, N$ ) sensor is given by

$$x_n(t) = (h_n * s)(t) + v_n(t), \quad (1)$$

where  $h_n(t)$  is the impulse response from the source to the  $n$ th sensor, and  $v_n(t)$  is an additive noise that is assumed to be uncorrelated and spatially white Gaussian with zero mean and variance equal to  $\sigma^2$  for all sensors. This is a reasonable model for many real-world noise fields [50]. In the short-time Fourier transform domain, the data model of the array signals can be expressed as

$$\mathbf{x}(k, f) = \mathbf{h}(k, f)S(k, f) + \mathbf{v}(k, f), \quad (2)$$

where  $k$  is the block time index,  $f$  is the frequency bin, the vectors are defined as

$$\begin{aligned} \mathbf{x}(k, f) &= [X_1(k, f), X_2(k, f), \dots, X_N(k, f)]^T \in \mathbb{C}^N, \\ \mathbf{h}(k, f) &= [H_1(k, f), H_2(k, f), \dots, H_N(k, f)]^T \in \mathbb{C}^N, \\ \mathbf{v}(k, f) &= [V_1(k, f), V_2(k, f), \dots, V_N(k, f)]^T \in \mathbb{C}^N, \end{aligned}$$

and  $X_n(k, f)$ ,  $H_n(k, f)$ ,  $S(k, f)$ , and  $V_n(k, f)$  are the discrete-time Fourier transforms (DTFTs) of  $x_n(t)$ ,  $h_n(t)$ ,  $s(t)$ , and  $v_n(t)$  respectively. We now select the first sensor ( $n = 1$ ) as the reference sensor. Assuming anechoic conditions, and under the hypothesis that all the sensors are omnidirectional, identical and with time-invariant transfer functions, the expression (2) can be simplified in the far-field as [51]

$$\mathbf{x}(k, f) = \mathbf{a}(f, \theta)\bar{S}(k, f) + \mathbf{v}(k, f), \quad (3)$$

where  $\bar{S}(k, f)$  is the source signal at the reference sensor, and  $\mathbf{a}(f, \theta)$  is the array steering vector defined as

$$\mathbf{a}(f, \theta) = [1, e^{\frac{-j2\pi f\tau_{12}(\theta)}{L}}, \dots, e^{\frac{-j2\pi f\tau_{1N}(\theta)}{L}}]^T \in \mathbb{C}^N, \quad (4)$$

where  $L$  is the size of the DTFT,  $j$  is the imaginary unit,  $\theta$  is the DOA of the source and  $\tau_{1n}$  is TDOA between the reference sensor and sensor  $n$ . The relationship between the TDOA  $\tau_{1n}$  and the DOA  $\theta$  is given by  $\tau_{1n}(\theta) = d_{1n} \sin(\theta)c^{-1}$ , where  $c$  is the speed of wave propagation and  $d_{1n}$  is the distance between the reference and the  $n$ th sensor.

### C. Beamforming

An acoustic SRP beamformer is typically computed in the frequency-domain by calculating the response power of each frequency bin and by fusing the narrowband components.

The output of a narrowband beamformer  $Y(k, f, \theta) \in \mathbb{C}$  at block time  $k$ , for frequency  $f$  and look direction  $\theta$ , is obtained as

$$Y(k, f, \theta) = \mathbf{w}^H(k, f, \theta)\mathbf{x}(k, f), \quad (5)$$

where  $\mathbf{w}(k, f, \theta) \in \mathbb{C}^N$  is a column vector containing the beamformer coefficients for time-shifting, weighting, and summing the data, so to steer the array in the direction  $\theta$ . Then,

the power spectral density (PSD) of the spatially filtered signal is

$$\begin{aligned} P(k, f, \theta) &= E\{|Y(k, f, \theta)|^2\} \\ &= \mathbf{w}^H(k, f, \theta) \mathbf{\Phi}(k, f) \mathbf{w}(k, f, \theta), \end{aligned} \quad (6)$$

where  $\mathbf{\Phi}(k, f) = E\{\mathbf{x}(k, f)\mathbf{x}^H(k, f)\} \in \mathbb{C}^{N \times N}$  is the PSD matrix of the array signal, which is symmetric and positive definite. The PSD matrix  $\mathbf{\Phi}(k, f)$  is unknown and it has to be estimated through the averaging of the array signal blocks [52], [53]

$$\hat{\mathbf{\Phi}}(k, f) = \frac{1}{M} \sum_{k_p=0}^{M-1} \mathbf{x}(k - k_p, f) \mathbf{x}^H(k - k_p, f), \quad (7)$$

where  $M$  is the number of signal blocks for the averaging, which is typically called number of snapshots. In the conventional narrowband beamformer, whose implementation reflects the delay-and-sum scheme, all its weights are equal in magnitude, i.e.  $\mathbf{w}_{\text{SRP}}(k, f, \theta) = \mathbf{a}(f, \theta)$ .

The broadband SRP is obtained by integrating the narrowband SRP over all frequencies. To increase the spatial resolution, the narrowband components are in general normalized with respect to some spectral characteristic. Examples of normalization methods used in practice are the phase transform (PHAT) [17], a pre-filter that uses the magnitude information of the PSD matrix to normalize the narrowband components in the SRP conventional beamforming, or the incoherent frequency fusion [54], a post-filter normalization which is defined as

$$P(k, \theta) = \sum_{f=0}^{L-1} \frac{P(k, f, \theta)}{\max_{\theta} [P(k, f, \theta)]}, \quad (8)$$

and that will be adopted herein to perform the narrowband data fusion. The PSD  $P(k, \theta)$  of a beamformer conveys information on the acoustic energy coming from direction  $\theta$ , thus it will be characterized by a maximum peak corresponding to the source direction  $\hat{\theta}$ . Therefore, the DOA estimate of the source is obtained by

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} [P(k, \theta)]. \quad (9)$$

#### D. The MVDR Beamformer

The MVDR beamformer [31] is a data-dependent spatial filtering technique which is aimed at keeping constant the gain related to a desired direction, while minimizing the acoustic energy originating from noise and sources located elsewhere. To improve robustness, the MVDR filter is often computed through a diagonal loading (DL) regularization [33]–[38], which leads to the following minimization problem

$$\begin{aligned} &\text{minimize} \quad \mathbf{w}^H(k, f, \theta) (\mathbf{\Phi}(k, f) + \mu' \mathbf{I}) \mathbf{w}(k, f, \theta), \\ &\text{subject to} \quad \mathbf{w}^H(k, f, \theta) \mathbf{a}(f, \theta) = 1, \end{aligned} \quad (10)$$

where  $\mu'$  is a real-valued, positive scalar. Solving (10) using the method of Lagrange multipliers, leads to the solution

$$\mathbf{w}_{\text{MVDR}}(k, f, \theta) = \frac{(\mathbf{\Phi}(k, f) + \mu' \mathbf{I})^{-1} \mathbf{a}(f, \theta)}{\mathbf{a}^H(f, \theta) (\mathbf{\Phi}(k, f) + \mu' \mathbf{I})^{-1} \mathbf{a}(f, \theta)}, \quad (11)$$

and to the corresponding PSD

$$P_{\text{MVDR}}(k, f, \theta) = \frac{1}{\mathbf{a}^H(f, \theta) (\mathbf{\Phi}(k, f) + \mu' \mathbf{I})^{-1} \mathbf{a}(f, \theta)}. \quad (12)$$

The DL regularization mitigates steering vector errors and discrete sampling effects in practical applications. However, the penalty weight  $\mu'$  has to be determined empirically. In general, a limitation of the DL regularization is that it is not clear how to efficiently compute the penalty weight for optimal performance, although useful data-dependent methods have been proposed as *ad hoc* procedures for specific applications [55], [56]. Finally, note that when the source signal is not active, the MVDR solution reduces to that of the conventional beamformer since  $\mathbf{\Phi}(k, f) = \sigma^2 \mathbf{I}$ , thus resulting in  $\mathbf{w}_{\text{MVDR}}(k, f, \theta) = \mathbf{a}(f, \theta)/N$ . The same condition holds also when  $\mu'$  approaches infinity.

#### E. The MUSIC Method

The MUSIC method [24] is based on a matrix eigendecomposition which exploits the orthogonality between signal and noise subspaces. The estimated noise subspace is used for obtaining the steering vector that is as orthogonal to the noise subspace as possible. The subspace orthogonality property leads to define the MUSIC pseudo spatial spectrum as

$$P_{\text{MUSIC}}(k, f, \theta) = \frac{1}{\mathbf{a}^H(f, \theta) \mathbf{U}_v(k, f) \mathbf{U}_v^H(k, f) \mathbf{a}(f, \theta)}, \quad (13)$$

where  $\mathbf{U}_v(k, f) \in \mathbb{C}^{N \times (N-Z)}$  is a matrix containing the eigenvectors corresponding to the noise subspace of the PSD matrix  $\mathbf{\Phi}(k, f)$ , and  $Z$  is the number of sources. The implementation of MUSIC requires that the number of sources is known or has to be estimated.

### III. DIAGONAL UNLOADING BEAMFORMING

The proposed DU beamforming, which is discussed in depth in this section, has as its principal objective the subtraction of the signal subspace from the noisy signal space. The intuition that the removal of the signal subspace contribution from the PSD matrix is beneficial to the beamforming performance comes from the observation of the principles underlying other beamformers, such as MUSIC and MVDR (an eigenanalysis that highlights the relationships between these beamformers will be provided in Section IV). We first examine the importance of the attenuation of the signal subspace by discussing the eigendecomposition of the PSD matrix and the orthogonality property. Then, we illustrate the method for a single acoustic source and we derive an optimal implementation for the case in which the noise statistics is known or can be estimated. Moreover, we analyze the multisource case, and finally we derive a suboptimal implementation that is effective even if the information on the noise component is not available. For simplicity, we will omit the block time variable  $k$  in the rest of the paper.

### A. The Eigendecomposition of the PSD Matrix and the Orthogonality Property

The symmetric and positive definite PSD matrix  $\Phi(f)$  can be decomposed in its eigenvalues and their associated eigenvectors through a subspace decomposition. Organizing the eigenvalues of  $\Phi$  in descending order ( $\lambda_1 > \lambda_2 > \dots > \lambda_N, \lambda_n \in \mathbb{R}$ ) and denoting  $\mathbf{u}_n \in \mathbb{C}^N$  their corresponding eigenvectors, the PSD matrix takes the following form

$$\Phi(f) = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H, \quad (14)$$

where  $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_N) \in \mathbb{R}^{N \times N}$  and  $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N] \in \mathbb{C}^{N \times N}$ . Under the hypothesis of a single source, the eigenvector that corresponds to the largest eigenvalue spans the signal subspace, and the remaining eigenvectors, which correspond to the smaller eigenvalues, span the noise subspace. Let  $P_s(f) = E\{|\tilde{S}(f)|^2\}$  denote the power of the signal, then the PSD matrix can be written as  $\Phi(f) = P_s(f)\mathbf{a}(f, \theta)\mathbf{a}^H(f, \theta) + \sigma^2\mathbf{I}$ . Therefore, we have that  $\lambda_1 = NP_s(f) + \sigma^2, \lambda_2 = \lambda_3 = \dots = \lambda_N = \sigma^2$ , with  $\lambda_1$  being the signal eigenvalue corresponding to the signal eigenvector  $\mathbf{u}_s = \mathbf{u}_1$ .

A positive definite Hermitian matrix has the important property that its eigenvector matrix  $\mathbf{U}$  is unitary and that its eigenvalues are always real and positive. If we separate the signal and the noise components, we can thus write

$$\mathbf{U}\mathbf{U}^{-1} = \mathbf{U}\mathbf{U}^H = \mathbf{u}_s\mathbf{u}_s^H + \mathbf{U}_v\mathbf{U}_v^H = \mathbf{I}, \quad (15)$$

where  $\mathbf{U}_v = [\mathbf{u}_2, \mathbf{u}_3, \dots, \mathbf{u}_N]$  is the noise subspace matrix. Moreover, since the matrix  $\mathbf{U}$  is unitary, its columns form an orthonormal basis for the definite positive matrix, and in particular we have that the signal eigenvector is orthogonal to each noise eigenvector:

$$\mathbf{u}_s^H \mathbf{u}_i = 0, \quad i = 2, 3, \dots, N. \quad (16)$$

Finally, since it is  $\Phi(f) = P_s(f)\mathbf{a}(f, \theta)\mathbf{a}^H(f, \theta) + \sigma^2\mathbf{I} = NP_s(f)\mathbf{u}_s\mathbf{u}_s^H + \sigma^2\mathbf{I}$ , then the noise eigenvectors in  $\mathbf{U}_v$  are orthogonal to the steering vector  $\mathbf{a}(f, \theta)$ :

$$\mathbf{a}^H(f, \theta)\mathbf{U}_v = \mathbf{0}, \quad (17)$$

i.e., the array steering vector  $\mathbf{a}(f, \theta)$  corresponding to the true signal parameters spans the signal subspace,  $\mathbf{u}_s$ , and is orthogonal to the noise subspace,  $\mathbf{U}_v$ . The orthogonality property, which is fully exploited in the MUSIC method, is known to be strictly connected to its high resolution property [45]–[49]. In fact, we can note that  $\mathbf{a}^H(f, \theta)\mathbf{U}_v\mathbf{U}_v^H\mathbf{a}(f, \theta) = 0$  results into a theoretically infinite energy peak in the pseudo spatial spectrum defined as in (13). We will show in Section IV that also the MVDR beamformer exploits the subspace orthogonality.

In order to compute the pseudo spatial spectrum in (13), the MUSIC method must compute the noise subspace  $\mathbf{U}_v$  from the PSD matrix  $\Phi(f)$ . This involves some computationally intensive operation due to the eigendecomposition. It is important to stress, however, that the noise subspace computation can be also achieved through a transformation  $\mathcal{F}(\Phi(f))$  of the PSD matrix that removes the signal subspace, since we can write

$$\mathbf{U}_v\mathbf{U}_v^H = \mathbf{U}\text{diag}(0, 1, \dots, 1)\mathbf{U}^H = \left(\frac{1}{\lambda'_v}\mathbf{I}\right)\mathcal{F}(\Phi(f)), \quad (18)$$

where the eigenvalue  $\lambda_s = NP_s(f) + \sigma^2$  was imposed to be zero and  $\lambda'_v$  are the noise eigenvalues of  $\mathcal{F}(\Phi(f))$ . We can thus write  $\Phi(f) = \left(\frac{\sigma^2}{\lambda'_v}\mathbf{I}\right)\mathcal{F}(\Phi(f)) + NP_s(f)\mathbf{u}_s\mathbf{u}_s^H$ . We can hence say that the exploitation of the orthogonality property can be performed by the removal or the attenuation of the signal subspace from the PSD matrix. Specifically, we can define a gain  $G$  that represents the amount of signal subspace left over after the removal from the PSD matrix and that is normalized with respect to the noise eigenvalues, the noise eigenvectors will all have unitary gain,  $\mathcal{F}(\Phi(f)) = \lambda'_v\mathbf{U}\text{diag}(G, 1, \dots, 1)\mathbf{U}^H$ . When  $G = 0$ , the orthogonality property is fully exploited ( $\mathbf{a}^H(f, \theta)\mathcal{F}(\Phi(f))\mathbf{a}(f, \theta) = 0$ ). When  $0 < G < 1$  we have an attenuation of the signal subspace with respect to the noise subspace, and the orthogonality property is partially exploited, i.e.  $\mathbf{a}^H(f, \theta)\mathcal{F}(\Phi(f))\mathbf{a}(f, \theta)$  returns the closest value to zero in the source direction. This may be the cause of some degradation in the localization performance in comparison to the case  $G = 0$ . Hence, a lower value of the gain  $G$  indicates a theoretical better localization performance. When  $G > 1$ , the orthogonality property cannot be exploited, and we fall in the case of the conventional beamformer. In fact, the normalized eigenvalue matrix for the conventional SRP can be written as  $\bar{\mathbf{\Lambda}} = \text{diag}\left(\frac{NP_s(f) + \sigma^2}{\sigma^2}, 1, \dots, 1\right)$ . Let  $\text{SNR} = P_s(f)/\sigma^2$  denote the signal-to-noise ratio. The SRP gain is given by

$$G_{\text{SRP}} = \frac{NP_s(f) + \sigma^2}{\sigma^2} = N\text{SNR} + 1, \quad (19)$$

which is always greater than 1 (it will be 1 when  $P_s(f) = 0$ , i.e. in absence of source signal, or when  $\sigma^2 = \infty$ ).

### B. Diagonal Unloading: Single Source Case

We make the following assumptions: 1) At the time of the acoustic analysis a single source is active in the sensed region; 2) The signals sensed by the sensors are affected by uncorrelated spatially white Gaussian noise with zero mean and variance equal to  $\sigma^2$ ; 3) There exists a transformation  $\mathcal{F} : \mathbb{C}^{N \times N} \rightarrow \mathbb{C}^{N \times N}$  that turns a PSD matrix into a transformed PSD matrix  $\Phi_{\text{DU}}(f) = \mathcal{F}(\Phi(f)) \in \mathbb{C}^{N \times N}$ , which is negative semidefinite, and whose eigenvalue  $\lambda_s \in \mathbb{R}$  corresponding to the signal subspace is equal to zero, while the other eigenvalues, corresponding to the noise subspace, are different from zero.

We formulate the DU beamformer by using an optimization problem with an orthogonality constraint that aims to achieve the signal subspace removal discussed before. Following what is proposed in [57], the optimization reads as:

$$\begin{aligned} &\text{minimize} \quad \|\mathbf{w}(f, \theta) - \mathbf{a}(f, \theta)\|^2, \\ &\text{subject to} \quad \mathbf{u}_s^H \mathbf{w}(f, \theta) = 0. \end{aligned} \quad (20)$$

We thus design the DU beamformer by imposing that the beamformer output is zero in the look direction ( $\mathbf{a}^H(f, \theta)\mathbf{w}(f, \theta) = \mathbf{u}_s^H \mathbf{w}(f, \theta) = 0$ ), which correspond to pursue the removal of the signal subspace from PSD matrix, and thus the orthogonality in (17). Using the method of Lagrange multipliers, leads to the solution

$$\mathbf{w}_{\text{DU}}(f, \theta) = (\mathbf{I} - \mathbf{u}_s\mathbf{u}_s^H)\mathbf{a}(f, \theta), \quad (21)$$

and by using equations (15) and (18), we obtain

$$\begin{aligned}\mathbf{w}_{\text{DU}}(f, \theta) &= \mathbf{U}_v \mathbf{U}_v^H \mathbf{a}(f, \theta) \\ &= \mathbf{U} \text{diag}(0, 1, \dots, 1) \mathbf{U}^H \mathbf{a}(f, \theta) \\ &= \left( \frac{1}{\lambda'_v} \mathbf{I} \right) \mathcal{F}(\Phi(f)) \mathbf{a}(f, \theta) \\ &= \left( \frac{1}{\lambda'_v} \mathbf{I} \right) \Phi_{\text{DU}}(f) \mathbf{a}(f, \theta),\end{aligned}\quad (22)$$

where  $\lambda'_v \in \mathbb{R}$  is the noise eigenvalue of the matrix  $\Phi_{\text{DU}}(f)$ . Substituting (22) in (6), we have

$$\begin{aligned}P'_{\text{DU}}(f, \theta) &= \mathbf{a}^H(f, \theta) \Phi_{\text{DU}}(f) \Phi(f) \Phi_{\text{DU}}(f) \mathbf{a}(f, \theta) \\ &= \frac{\sigma^2}{\lambda'_v} \mathbf{a}^H(f, \theta) \mathbf{U} \text{diag}(0, \lambda'_v, \dots, \lambda'_v) \mathbf{U}^H \mathbf{a}(f, \theta) \\ &= \frac{\sigma^2}{\lambda'_v} \mathbf{a}^H(f, \theta) \Phi_{\text{DU}}(f) \mathbf{a}(f, \theta),\end{aligned}\quad (23)$$

where the quantity  $\frac{\sigma^2}{\lambda'_v}$  is a scalar factor that can be omitted since it has no influence on the DOA estimation. We can note that (23) has a form similar to that of the conventional SRP beamformer. However, since  $\Phi_{\text{DU}}$  is negative semidefinite, we have that  $P'_{\text{DU}}(f, \theta) \leq 0$ , and we can write the maximum search of the PSD in (23),  $\hat{\theta} = \underset{\theta}{\text{argmax}}[P'_{\text{DU}}(f, \theta)]$ , in the following equivalent form

$$\hat{\theta} = \underset{\theta}{\text{argmax}} \left[ -\frac{1}{P'_{\text{DU}}(f, \theta)} \right] = \underset{\theta}{\text{argmax}} [P_{\text{DU}}(f, \theta)], \quad (24)$$

where the pseudo spatial spectrum  $P_{\text{DU}}(f, \theta)$  is defined as

$$P_{\text{DU}}(f, \theta) = -\frac{1}{\mathbf{a}^H(f, \theta) \Phi_{\text{DU}}(f) \mathbf{a}(f, \theta)}. \quad (25)$$

To ensure that  $\Phi_{\text{DU}}$  is negative semidefinite, the Hermitian matrix must have all strictly nonpositive eigenvalues. This condition can be met by transforming the PSD matrix  $\Phi$  with a diagonal unloading. The DU transformed PSD matrix is given by

$$\Phi_{\text{DU}}(f) = \Phi(f) - \mu \mathbf{I}, \quad (26)$$

where  $\mu$  is a real-valued, positive scalar. Note that the negative semidefinite assumption is fundamental in the design of the DU beamformer, since it allows the exploitation of the high resolution orthogonality property. The matrix  $\Phi_{\text{DU}}(f)$  is negative semidefinite if  $\mu \geq NP_s(f) + \sigma^2$ . If  $\mu \leq \sigma^2$ , then  $\Phi_{\text{DU}}(f)$  is positive semidefinite, the gain of the signal subspace will be greater than 1, and the DU operation simply corresponds to a noise attenuation operated on a conventional beamformer. Substituting equation (26) in (25), the pseudo spatial spectrum becomes

$$P_{\text{DU}}(f, \theta) = -\frac{1}{\mathbf{a}^H(f, \theta) (\Phi(f) - \mu \mathbf{I}) \mathbf{a}(f, \theta)}. \quad (27)$$

By adding or subtracting the same real offset  $\mu$  to each element of the diagonal, all eigenvalues of the PSD matrix are increased or decreased by the value  $\mu$  while its eigenvectors remain the same. In other words, a transformation of the diagonal of  $\Phi$  modifies its eigenvalues while keeping the same signal and noise subspaces. This fact is very important since

it allows us to control the contribution of the subspaces in the computation of the beamformer through a transformation of the eigenvalues. The eigenvalue decomposition of the transformed PSD matrix (26) is therefore given by

$$\Phi_{\text{DU}}(f) = \mathbf{U} \text{diag}(NP_s(f) + \sigma^2 - \mu, \sigma^2 - \mu, \dots, \sigma^2 - \mu) \mathbf{U}^H. \quad (28)$$

Now, the constraint of having the eigenvalue corresponding to the signal subspace of the transformed PSD matrix  $\Phi_{\text{DU}}$  equal to zero ( $\lambda_s = 0$ ) becomes

$$NP_s(f) + \sigma^2 - \mu = 0. \quad (29)$$

The solution is easily found by considering that

$$\text{tr}(\Phi) = \text{tr}(\Lambda) = N(P_s(f) + \sigma^2), \quad (30)$$

and hence we have that the penalty weight of DU is data-dependent and is given by

$$\mu = \text{tr}(\Phi(f)) - (N - 1)\sigma^2. \quad (31)$$

The DU transformation with the penalty weight in (31) guarantees that the transformed PSD matrix (26) is negative semidefinite and that the eigenvalues corresponding to the noise subspace are non-zero, since they are set to  $\lambda'_v = \sigma^2 - \mu = -\text{tr}(\Phi(f)) + N\sigma^2 = -NP_s(f)$ . This fact guarantees the total removal of signal subspace in the transformed PSD matrix. Finally, substituting (31) in (27) and changing the sign, the pseudo spatial spectrum of the optimal DU beamforming becomes

$$P_{\text{DU}}(f, \theta) = \frac{1}{\mathbf{a}^H(f, \theta) \Phi_{\text{DU}}^{\text{opt}}(f) \mathbf{a}(f, \theta)}, \quad (32)$$

with

$$\Phi_{\text{DU}}^{\text{opt}}(f) = [\text{tr}(\Phi(f)) - (N - 1)\sigma^2] \mathbf{I} - \Phi(f). \quad (33)$$

The constraint on the signal eigenvalue is fundamental for the proposed DU beamforming and has the important effect of improving the narrowband spatial resolution with respect to the conventional beamforming, by exploiting the subspace orthogonality. In order to adopt in practice this optimal implementation of the DU beamforming, it is required that  $\sigma^2$  is known or can be estimated. Typically  $\sigma^2$  can be estimated from signal-free analysis blocks, if the noise can be considered stationary.

We can note that the PSD in (32) has a form similar to that of MVDR and MUSIC in (12) and (13), respectively. The PSD in (32) can be considered as a normalization constant that does not affect the output of the beamformer (similarly to the MVDR case [58]). We can thus write the weights of the DU beamformer by omitting the scalar factor  $\lambda'_v$  in (22) and by inserting the normalization constant as

$$\mathbf{w}_{\text{DU}}^{\text{opt}}(f, \theta) = \frac{\Phi_{\text{DU}}^{\text{opt}}(f) \mathbf{a}(f, \theta)}{\mathbf{a}^H(f, \theta) \Phi_{\text{DU}}^{\text{opt}}(f) \mathbf{a}(f, \theta)}. \quad (34)$$

The result has a form similar to that of an MDVR beamformer without the computational cost of the matrix inversion. When the source signal is not active, the DU beamformer reduces to the conventional beamformer due its data-dependent nature, resulting in  $\mathbf{w}_{\text{DU}}^{\text{opt}} = \mathbf{a}(f, \theta)/N$ .

### C. Multisource Case

Without loss of generality, we consider the case in which two sources impinge an array of sensors. Let  $s_1(t) \in \mathbb{R}$  and  $s_2(t) \in \mathbb{R}$  denote the signals generated by two sources at time  $t$ . The source signals are assumed to be uncorrelated. The multisource PSD matrix can be written as  $\Phi_{\text{MS}}(f) = \mathbf{A}\mathbf{S}\mathbf{A}^H + \sigma^2\mathbf{I}$  where  $\mathbf{A} = [\mathbf{a}_1(f, \theta), \mathbf{a}_2(f, \theta)]$  and  $\mathbf{S} = \text{diag}(P_{s_1}(f), P_{s_2}(f))$  with  $\mathbf{a}_1(f, \theta)$  and  $\mathbf{a}_2(f, \theta)$  being the array steering vectors for source  $s_1(t)$  and  $s_2(t)$ , and  $P_{s_1}$  and  $P_{s_2}$  being the power of the source signals. Let  $\lambda_1$  and  $\lambda_2$  denote the larger eigenvalues corresponding to the signal subspaces of  $s_1(t)$  and  $s_2(t)$ , then we have that the diagonal elements of eigenvalue matrix  $\Lambda_{\text{MS}}(f)$  are  $\lambda_1 = NP_{s_1}(f) + \sigma^2$ ,  $\lambda_2 = NP_{s_2}(f) + \sigma^2$ , and  $\lambda_v = \sigma^2$  with  $v = 3, \dots, N$ . The trace of  $\Phi_{\text{MS}}$  becomes

$$\text{tr}(\Phi_{\text{MS}}(f)) = \text{tr}(\Lambda_{\text{MS}}(f)) = N(P_{s_1}(f) + P_{s_2}(f) + \sigma^2). \quad (35)$$

In a multisource scenario, we can only minimize the signal eigenvalues with the DU procedure in (31). We can normalize to one the eigenvalues corresponding to the noise subspaces (the gain of each noise subspace will thus be one in the beamforming computation), and we can write the normalized eigenvalue matrix of  $\Phi_{\text{DU}}^{\text{opt}}$  as

$$\bar{\Lambda}_{\text{DU}}^{\text{opt}}(f) = \text{diag}\left(\frac{P_{s_2}(f)}{P_{s_1}(f) + P_{s_2}(f)}, \frac{P_{s_1}(f)}{P_{s_1}(f) + P_{s_2}(f)}, 1, \dots, 1\right). \quad (36)$$

Let  $G_1$  and  $G_2$  denote the theoretical gains of the signal subspace in the transformed PSD matrix, then we have

$$G_1 = \frac{1}{\gamma + 1}, \quad G_2 = \frac{1}{\gamma^{-1} + 1}, \quad (37)$$

where  $\gamma = P_{s_1}(f)/P_{s_2}(f)$ . Both gains are always minor than one. Hence, the DU operation guarantees a reduction of the signal subspaces in this scenario. The theoretical attenuation of the signal subspace is thus related to the signal power of both sources. In the general case of  $S$  sources, each signal subspace vector is attenuated by a factor proportional to the corresponding signal eigenvalue and the gain quantity is related to the sum of the other eigenvalues due to the subtraction operation.

### D. Suboptimal Implementation

If  $\sigma^2$  is unknown or cannot be estimated, the DU beamforming can be implemented through a suboptimal solution that still aims at the signal subspace subtraction, but does not guarantee its total removal, thus resulting in  $\lambda_s \neq 0$ . By ignoring the noise component in (33), the DU beamforming becomes

$$P_{\text{DU}}(f, \theta) = \frac{1}{\mathbf{a}^H(f, \theta) \Phi_{\text{DU}}^{\text{subopt}}(f) \mathbf{a}(f, \theta)}, \quad (38)$$

with

$$\Phi_{\text{DU}}^{\text{subopt}}(f) = \text{tr}(\Phi(f))\mathbf{I} - \Phi(f). \quad (39)$$

Equation (38) is an optimal solution in the ideal noise-free case  $\sigma^2 = 0$ . On the other hand, when a noise component is actually present, its adoption results in a suboptimal implementation of the DU algorithm due to the residual amount of the signal subspace. This residual will affect the beamforming computation also in the single source case and may be the cause of some degradation in the localization performance due to the reduced effectiveness in the exploitation of the orthogonality of subspaces (an example of the DU localization performance at variations of the signal subspace gain is shown in Figure 8). To analyze the amount of signal subspace involved in the computation of the spatial filter, note that the normalized eigenvalue matrix of  $\Phi_{\text{DU}}^{\text{subopt}}$  can be written as

$$\bar{\Lambda}_{\text{DU}}^{\text{subopt}}(f) = \text{diag}\left(\frac{(N-1)\sigma^2}{NP_s(f) + (N-1)\sigma^2}, 1, \dots, 1\right). \quad (40)$$

Let  $G_{\text{DU}}$  denote the signal subspace in the spatial filter. We can then write the DU theoretical gain as

$$G_{\text{DU}} = \frac{(N-1)\sigma^2}{NP_s(f) + (N-1)\sigma^2} = \frac{N-1}{N\text{SNR} + N-1}. \quad (41)$$

The gain is always less than 1 (it would be 1 when  $P_s(f) = 0$  and in case of infinite noise power), and its value decreases as the number of sensors in the array decreases or as the SNR increases. Note that the capability of signal subspace attenuation of the suboptimal implementation decreases when  $N$  increases, since the unknown noise component is proportional to  $(N-1)\sigma^2$ . The total removal of the signal subspace is not guaranteed when using the suboptimal DU implementation of equation (38) (unknown  $\sigma^2$ ), however a certain attenuation is always obtained. On the other hand, the total removal of signal subspace is achieved with equation (32) when  $\sigma^2$  is known and an accurate estimate of the PSD matrix is computed.

For the suboptimal DU implementation in a multisource scenario, the normalized eigenvalue matrix in case of two sources becomes

$$\bar{\Lambda}_{\text{DU}}^{\text{subopt}}(f) = \text{diag}\left(\frac{NP_{s_2}(f) + (N-1)\sigma^2}{N(P_{s_1}(f) + P_{s_2}(f)) + (N-1)\sigma^2}, \frac{NP_{s_1}(f) + (N-1)\sigma^2}{N(P_{s_1}(f) + P_{s_2}(f)) + (N-1)\sigma^2}, 1, \dots, 1\right). \quad (42)$$

We have that the signal subspace gains of equations (37) becomes

$$G_1 = \frac{N\text{SNR}_2 + N-1}{N(\text{SNR}_1 + \text{SNR}_2) + N-1}, \quad G_2 = \frac{N\text{SNR}_1 + N-1}{N(\text{SNR}_1 + \text{SNR}_2) + N-1}, \quad (43)$$

where  $\text{SNR}_1 = P_{s_1}(f)/\sigma^2$  and  $\text{SNR}_2 = P_{s_2}(f)/\sigma^2$ . Both gains are always less than one. The theoretical attenuation of the signal subspace is thus related to the SNR of both sources.

Note that when  $\Phi(f)$  is singular and has rank 1, i.e., in the case of single source with  $\sigma^2 = 0$ , the suboptimal implementation reduces to an optimal solution since the signal subspace is totally absent in the transformed matrix. The

weights of the suboptimal DU beamformer are determined substituting  $\Phi_{\text{DU}}^{\text{subopt}}(f)$  in (34):

$$\mathbf{w}_{\text{DU}}^{\text{subopt}}(f, \theta) = \frac{(\text{tr}(\Phi(f))\mathbf{I} - \Phi(f))\mathbf{a}(f, \theta)}{\mathbf{a}^H(f, \theta)(\text{tr}(\Phi(f))\mathbf{I} - \Phi(f))\mathbf{a}(f, \theta)}, \quad (44)$$

which is an exact solution when  $\text{rank}(\Phi(f))=1$ , otherwise it is only an approximation solution since we cannot derive the PSD in (38) substituting (44) in (6), unless the signal eigenvalue of  $\Phi_{\text{DU}}^{\text{subopt}}(f) = \text{tr}(\Phi(f))\mathbf{I} - \Phi(f)$  is zero.

#### IV. PSD MATRIX EIGENANALYSIS

In this section, we provide an analysis of the proposed DU beamforming properties, highlighting the role of the signal subspace attenuation. We analyze from a theoretical viewpoint the relationship between the DU method and some known beamformers by discussing the eigenanalysis of their respective transformed PSD matrices.

By expressing the DU, the MVDR and the MUSIC beamformers in terms of their transformed PSD matrix, we obtain the spatial spectrum of the three methods in a form which is easily comparable. The transformed PSD matrix  $\Phi_{\text{DU}}^{\text{opt}}(f)$  is given in (33). The corresponding eigenvalue matrix of the transformed matrix  $\Phi_{\text{DU}}^{\text{opt}}(f)$  can be derived from (28), and reads as

$$\Lambda_{\text{DU}}^{\text{opt}}(f) = \text{diag}(0, \text{tr}(\Phi(f)) - N\sigma^2, \dots, \text{tr}(\Phi(f)) - N\sigma^2). \quad (45)$$

Let us first compare  $\Phi_{\text{DU}}^{\text{opt}}(f)$  and  $\Lambda_{\text{DU}}^{\text{opt}}(f)$  to the PSD and eigenvalue matrices of the MUSIC method. From (13), we have that the transformed PSD matrix of MUSIC can be written as  $\Phi_{\text{MUSIC}}(f) = \mathbf{U}_v(f)\mathbf{U}_v^H(f) = \mathbf{U}(f)\Lambda_{\text{MUSIC}}(f)\mathbf{U}^H(f)$ . Assuming that a single source is present, the eigenvalue diagonal matrix is then

$$\Lambda_{\text{MUSIC}}(f) = \text{diag}(0, 1, \dots, 1). \quad (46)$$

That is, MUSIC assigns a zero value to the eigenvalue corresponding to the signal subspace, and sets to 1 all eigenvalues corresponding to the noise subspace, resulting a signal subspace gain  $G_{\text{MUSIC}} = 0$ . We can interpret MUSIC as a beamformer that removes the signal subspace and uses a transformed PSD matrix  $\Phi_{\text{MUSIC}}$ , in which the eigenvector matrix  $\mathbf{U}$  is that of  $\Phi$ , and the eigenvalues are set to zero or to one if corresponding to noise or signal eigenvectors respectively. If we now compare the DU and MUSIC eigenvalue matrices, we note that it is

$$\Lambda_{\text{DU}}^{\text{opt}}(f) = \alpha_1 \Lambda_{\text{MUSIC}}(f), \quad (47)$$

where  $\alpha_1 = \text{tr}(\Phi(f)) - N\sigma^2$ . In this single source case, since the signal eigenvalue of the DU transformed PSD matrix is zero, we have that DU differs from MUSIC only by a scaling factor of the spatial spectrum, i.e.  $P_{\text{DU}} = \frac{1}{\alpha_1} P_{\text{MUSIC}}$ .

If we now look at the MVDR beamformer, from (12) we have that the transformed PSD matrix is  $\Phi_{\text{MVDR}}(f) = (\Phi(f) + \mu'\mathbf{I})^{-1}$ , and the diagonal matrix of the  $\Phi_{\text{MVDR}}$  eigenvalues can be written as

$$\Lambda_{\text{MVDR}}(f) = \text{diag}\left(\frac{1}{NP_s(f) + \sigma^2 + \mu'}, \frac{1}{\sigma^2 + \mu'}, \dots, \frac{1}{\sigma^2 + \mu'}\right). \quad (48)$$

Since  $\Phi_{\text{MVDR}}$  is Hermitian and full-rank, the eigenvectors of the inverse matrix are equal to those of the matrix itself. We can normalize the eigenvalue matrix of  $\Phi_{\text{MVDR}}(f)$  by multiplying each element for  $\sigma^2 + \mu'$ , obtaining

$$\bar{\Lambda}_{\text{MVDR}}(f) = \text{diag}\left(\frac{\sigma^2 + \mu'}{NP_s(f) + \sigma^2 + \mu'}, 1, \dots, 1\right). \quad (49)$$

The theoretical total removal of the signal subspace is achieved when  $P_s(f) = \infty$  or  $N = \infty$ , and in such case we have that  $P_{\text{MVDR}} = \alpha_2 P_{\text{DU}}$  and  $P_{\text{MVDR}} = \alpha_3 P_{\text{MUSIC}}$ , where  $\alpha_2$  and  $\alpha_3$  are real positive values. Hence, MVDR is always affected by a certain amount of signal subspace in the SRP computation. Let  $G_{\text{MVDR}}$  denote the gain of the signal subspace in the MVDR. We can then write the theoretical gain as

$$G_{\text{MVDR}} = \frac{\sigma^2 + \mu'}{NP_s(f) + \sigma^2 + \mu'} = \frac{1 + \frac{\mu'}{\sigma^2}}{N\text{SNR} + 1 + \frac{\mu'}{\sigma^2}}. \quad (50)$$

The noise signal attenuation is always guaranteed since  $G_{\text{MVDR}}$  is always less than 1. Hence, the DU beamforming has the advantage, if compared to MVDR, that the signal subspace is not used in the orthogonality searching procedure. However, the DU method requires that  $\sigma^2$  is known to exploit its optimal solution. If  $\sigma^2$  is unknown, DU can be implemented through its suboptimal solution given by equation (38). By comparing (41) and (50) when  $\mu' = 0$ , we can write the following equation, valid when  $N > 2$

$$G_{\text{DU}} = \frac{N-1}{N\text{SNR} + N-1} > \frac{1}{N\text{SNR} + 1} = G_{\text{MVDR}}. \quad (51)$$

Hence, the theoretical MVDR signal subspace attenuation is greater than the DU subspace attenuation when the suboptimal solution is used.

In practical applications, there is always a certain mismatch between the estimated and the actual PSD matrix, due to the finite sample size (number of snapshots), to the signal model mismatches, and to the nonstationary nature of the source. It should be noted however that, in slowly varying acoustic conditions, the estimated PSD matrix approaches its true value as the number of snapshots increases [59]. In stationary conditions, it is well known that the estimated PSD matrix converges to the actual PSD matrix for infinity snapshots in a mean square sense:  $M \rightarrow \infty, \hat{\Phi}(f) \rightarrow \Phi(f)$ . On the other hand, when the acoustic conditions change too fast, the assumption that incrementing the number of snapshots improves the PSD estimation is not valid anymore. We can write the estimated eigenvalue matrix of  $\hat{\Phi}(f)$  as

$$\hat{\Lambda}(f) = \text{diag}(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_N). \quad (52)$$

For the suboptimal DU beamformer, we have that the estimated eigenvalue matrix of  $\hat{\Phi}_{\text{DU}}^{\text{subopt}}(f)$  is

$$\hat{\Lambda}_{\text{DU}}^{\text{subopt}}(f) = \text{diag}(\hat{\lambda}_2 + \hat{\lambda}_3 + \dots + \hat{\lambda}_N, \hat{\lambda}_1 + \hat{\lambda}_3 + \dots + \hat{\lambda}_N, \dots, \hat{\lambda}_1 + \hat{\lambda}_2 + \dots + \hat{\lambda}_{N-1}). \quad (53)$$

We can then note in (53) that the DU operation may result in different gains for the noise subspaces in the transformed PSD matrix if compared to the theoretical case (40). Specifically, the signal subspace has smallest gain, the noise eigenvector

TABLE I  
THE SPATIAL SPECTRUM, THE WEIGHTS AND SIGNAL SUBSPACE GAIN (SINGLE SOURCE SCENARIO) OF NARROWBAND BEAMFORMERS.

<b>Conventional</b>	$P_{\text{SRP}} = \mathbf{a}^H \Phi \mathbf{a}$	[30]	$\mathbf{w}_{\text{SRP}} = \mathbf{a}$	[30]	$G_{\text{SRP}} = \text{NSNR} + 1 > 1$
<b>MVDR</b>	$P_{\text{MVDR}} = \frac{1}{\mathbf{a}^H (\Phi + \mu' \mathbf{I})^{-1} \mathbf{a}}$	[31]	$\mathbf{w}_{\text{MVDR}} = \frac{(\Phi + \mu' \mathbf{I})^{-1} \mathbf{a}}{\mathbf{a}^H (\Phi + \mu' \mathbf{I})^{-1} \mathbf{a}}$	[31]	$G_{\text{MVDR}} = \frac{1 + \frac{\mu'}{\sigma^2}}{\text{NSNR} + 1 + \frac{\mu'}{\sigma^2}} < 1$
<b>MUSIC</b>	$P_{\text{MUSIC}} = \frac{1}{\mathbf{a}^H \mathbf{U}_v \mathbf{U}_v^H \mathbf{a}}$	[24]	$\mathbf{w}_{\text{MUSIC}} = \mathbf{U}_v \mathbf{U}_v^H \mathbf{a}$	[57]	$G_{\text{MUSIC}} = 0$
<b>DU</b>	$P_{\text{DU}} = \frac{1}{\mathbf{a}^H (\text{tr}(\Phi) \mathbf{I} - \Phi) \mathbf{a}}$		$\mathbf{w}_{\text{DU}} = \frac{(\text{tr}(\Phi) \mathbf{I} - \Phi) \mathbf{a}}{\mathbf{a}^H (\text{tr}(\Phi) \mathbf{I} - \Phi) \mathbf{a}}$		$G_{\text{DU}} = \frac{N-1}{\text{NSNR} + N-1} < 1$

associated with the smallest eigenvalue has gain equal to 1, and the other noise eigenvectors have gain minor than 1. Note that, in contrast to the MUSIC method, in which the gains of the noise subspace components are all set to 1 (see equation (46)), here the relative weighting of the different noise components are preserved. This fact results, as the simulations in Section VI will show, in different DU vs MUSIC localization performances in the single source scenario. For the robust MVDR beamformer, the estimated eigenvalue matrix of  $\hat{\Phi}_{\text{MVDR}}(f)$  is

$$\hat{\Lambda}_{\text{MVDR}}(f) = \text{diag}\left(\frac{1}{\hat{\lambda}_1 + \mu'}, \frac{1}{\hat{\lambda}_2 + \mu'}, \dots, \frac{1}{\hat{\lambda}_N + \mu'}\right). \quad (54)$$

If we normalize the eigenvalues with respect to the last eigenvalue, which is the largest one due to the diagonal removal and to the inversion operation, in the single source case we obtain the following estimated signal subspace gains

$$\begin{aligned} \hat{G}_{\text{DU}} &= \frac{\sum_{i=2}^N \hat{\lambda}_i}{\sum_{i=1}^{N-1} \hat{\lambda}_i}, \\ \hat{G}_{\text{MVDR}} &= \frac{\hat{\lambda}_N + \mu'}{\hat{\lambda}_1 + \mu'}, \end{aligned} \quad (55)$$

and we have that

$$\hat{G}_{\text{DU}} < \hat{G}_{\text{MVDR}}, \quad \text{if } \mu' > \sum_{i=2}^{N-1} \hat{\lambda}_i. \quad (56)$$

When the condition in equation (56) holds, DU will provide better localization performances. On the other hand, the MVDR will provide a better performance when  $\mu' < \sum_{i=2}^{N-1} \hat{\lambda}_i$ . When the PSD matrix is estimated with a limited number of snapshots ( $M < N$ ), the PSD matrix is singular and hence the noise eigenvalues are smaller than the true values. The DU signal subspace attenuation will be larger than the attenuation provided by MVDR, in which the DL regularization guarantees that  $\Phi_{\text{MVDR}}(f)$  is always full-rank. When increasing the number of snapshots, the noise eigenvalue estimation of the PSD matrix will be more accurate and the MVDR will provide a larger signal subspace attenuation, which confirms the theoretical comparison in (51).

It is interesting to observe that the MVDR beamforming can be considered as a special case of the DU beamforming under the hypothesis of the model in equation (3) (single plane wave plus spatially uncorrelated white Gaussian noise). From (45) and (48) it can be seen that if we add a constant equal to  $\sigma^2 + \mu'$  to the diagonal of the DU matrix, we have

$$\Lambda_{\text{DU}}^c(f) = \text{diag}(\sigma^2 + \mu', \text{tr}(\Phi(f)) - N\sigma^2 + \sigma^2 + \mu', \dots), \quad (57)$$

TABLE II  
A COMPARISON OF THE DU PROCEDURE AND OTHER BEAMFORMERS, FOR THE SINGLE SOURCE SCENARIO.

$\Phi_{\text{DU}}(f) = \Phi(f) - \mu \mathbf{I}$	Transformed PSD matrix	Equivalence
Positive Semidefinite	$\Phi_{\text{DU}}(f) = \Phi(f) - \mu \mathbf{I} \quad (\mu \leq \sigma^2)$	Conventional
Negative Semidefinite	$\Phi_{\text{DU}}^{\text{opt}}(f) = [\text{tr}(\Phi(f)) - (N-1)\sigma^2] \mathbf{I} - \Phi(f)$	MUSIC
Negative Semidefinite	$\Phi_{\text{DU}}^c(f) = [\text{tr}(\Phi(f)) - (N-2)\sigma^2 + \mu'] \mathbf{I} - \Phi(f)$	MVDR
Negative Semidefinite	$\Phi_{\text{DU}}^{\text{subopt}}(f) = \text{tr}(\Phi) \mathbf{I} - \Phi$	

and the two eigenvalue matrices will become equal apart from a scaling factor  $\alpha = \frac{1}{(\sigma^2 + \mu')(\text{tr}(\Phi(f)) - \sigma^2 + \mu')}$ , i.e.  $\Lambda_{\text{MVDR}}(f) = \alpha \Lambda_{\text{DU}}^c(f)$ . We thus can say that the robust MVDR can be considered as a special case of the proposed DU method with a suboptimal solution given by

$$\Phi_{\text{DU}}^c(f) = [\text{tr}(\Phi(f)) - (N-2)\sigma^2 + \mu'] \mathbf{I} - \Phi(f). \quad (58)$$

With this specific solution, given the equivalence with the MVDR beamformer in (11), we can write the weight of the DU beamformer as

$$\mathbf{w}_{\text{DU}}^c(f, \theta) = \frac{\Phi_{\text{DU}}^c(f) \mathbf{a}(f, \theta)}{\mathbf{a}^H(f, \theta) \Phi_{\text{DU}}^c(f) \mathbf{a}(f, \theta)}, \quad (59)$$

and, by using the weight  $\mathbf{w}_{\text{DU}}^c(f, \theta)$  in the output power beamforming equation, we obtain the following output PSD:

$$P_{\text{DU}}^c = \frac{\mathbf{a}^H(f, \theta) \Phi_{\text{DU}}^c(f) \Phi(f) \Phi_{\text{DU}}^c(f) \mathbf{a}(f, \theta)}{[\mathbf{a}^H(f, \theta) \Phi_{\text{DU}}^c(f) \mathbf{a}(f, \theta)] [\mathbf{a}^H(f, \theta) \Phi(f) \mathbf{a}(f, \theta)]}. \quad (60)$$

We now search for a simpler form of the matrix product  $\Phi_{\text{DU}}^c(f) \Phi(f)$  by referring to the eigenvalues of the two matrices and to their product. If for simplicity we set  $\mu' = 0$ , we have that  $\Lambda_{\text{DU}}^c(f) \Lambda(f) = (NP_s(f) + \sigma^2) \sigma^2 \mathbf{I} = \alpha' \mathbf{I}$ . This finally leads to

$$P_{\text{DU}}^c = \frac{\alpha'}{\mathbf{a}^H(f, \theta) \Phi_{\text{DU}}^c(f) \mathbf{a}(f, \theta)}. \quad (61)$$

The quantity  $\alpha'$  is a scalar factor that can be omitted since it has no influence on the DOA estimates.

In table I, we summarize the narrowband spatial spectrum, the weights and the signal subspace gain (single source scenario) equations of the conventional beamforming, the robust MVDR, the MUSIC and the suboptimal DU beamforming (i.e., when  $\sigma^2$  is unknown) omitting for simplicity the dependency from  $f$  and  $\theta$ . In table II, we summarize the proposed DU procedure and the equivalence with other beamformers for the single source scenario.

TABLE III  
THE COMPLEXITY OF BROADBAND BEAMFORMERS.

<b>Conventional - DU</b>	$O(L \log L + \frac{L}{2} N)$
<b>MVDR - MUSIC</b>	$O(L \log L + \frac{L}{2} N^3)$

## V. COMPUTATIONAL COMPLEXITY ANALYSIS

The computational complexity of the DU beamformer is that of a conventional spatial filter, which has  $O(n)$  complexity. In fact, the DU procedure adds few basic operations since it consists of  $N - 1$  additions for the calculation of the trace of the PSD matrix,  $N$  subtractions for the diagonal unloading procedure, and  $\Gamma L/2$  divisions, where  $\Gamma$  is the number of angles for searching the source and  $L$  is the size of the fast Fourier transform (FFT). On the other hand, the MVDR and MUSIC methods require a full-rank inversion matrix the former, and an eigendecomposition the latter. For both methods, the complexity is  $O(n^3)$ . In the frequency-domain broadband beamforming, the accuracy and the computational cost is related to the number of narrowband components. Therefore we have that the conventional and the DU beamforming has  $O(L \log L + \frac{L}{2} N)$  complexity, where  $L \log L$  is the complexity contribution of the FFT, while the MVDR and MUSIC methods require  $O(L \log L + \frac{L}{2} N^3)$  flops. Hence, the DU beamforming is attractive in array processing since it provides a robust localization beamforming at no significant additional cost.

## VI. EXPERIMENTAL RESULTS

In this section, we present some experiments, based on both simulated and real acoustic data, that validate the proposed DU beamformer. The acoustic data were obtained by simulating the wave propagation of a speech signal in different reverberating indoor environments, and by actually recording a speech signal reproduced in an reverberant environment. We have compared the DOA localization performance using the root mean square error (RMSE) of the proposed DU, the MVDR [31] with DL, the MUSIC [24], and the conventional beamforming [30], i.e., the steered response power (SRP), and the SRP-PHAT [17]. Note that in DU, MVDR, MUSIC, SRP the broadband fusion was computed with the post-filter normalization [54] of equation (8). A data-dependent DL for the MVDR is adopted to improve the robustness of the MVDR. The data-dependent DL factor used is given by  $\mu'(f) = \text{tr}(\Phi(f))\Delta$  [60], where  $\Delta$  is a parameter that controls the relative artificial noise level in the data. We have set  $\Delta$  equal to  $1/N$ , which has been demonstrated to be a robust solution in [60]. The inversion and eigendecomposition of the PSD matrix was performed by singular value decomposition since it provides some numerical advantages. Besides, an optimal frequency range between 80 Hz and 8000 Hz, since it is the typical spectrum range of speech signals, was used for all beamformers. The sampling frequency was 44.1 kHz and frames of the signal were obtained by a Hann window of size  $L = 2048$  samples with an overlap of 512 samples. Given the nonstationary nature of the speech signal, a small number of snapshots was used for the estimation of the PSD matrix. For

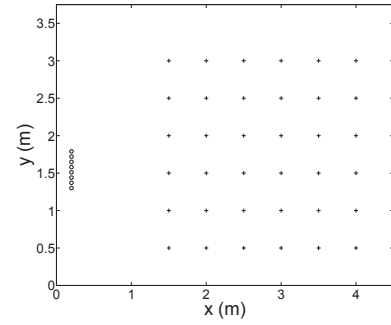


Fig. 1. The simulated room setup with the positions of sensors and sources.

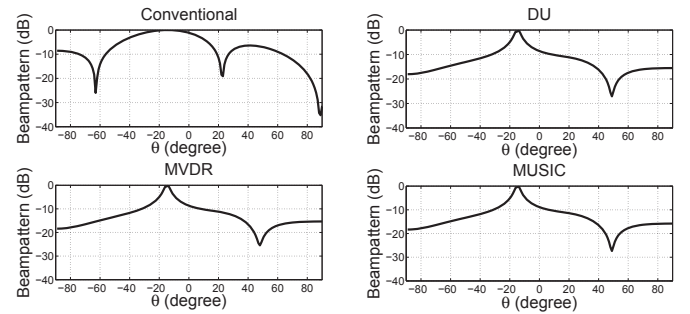


Fig. 2. The beampattern for the conventional, the DU, the MVDR, and the MUSIC beamformers of a 1000 Hz sinusoid impinging the array from a DOA of -18 degree.

the DU beamforming, we thus used the suboptimal version, as formulated in equation (38). We also provide some results that illustrate its optimal solution performances with known and estimated noise statistics.

### A. Synthetic Data

We consider an uniform linear array (ULA) of 8 sensors in noisy and reverberant conditions. The distance between microphones was set to 0.07 m. Acoustic simulations of reverberant environments were obtained with the image-source method [61], [62]. We configured a localization task in a room of 4.5 m  $\times$  3.75 m  $\times$  3.05 m. The simulations were conducted with different SNR levels, obtained by adding mutually independent white Gaussian noise to each channel. We have investigated both the single and the multiple source scenarios using speech signals from the TSP Speech Database<sup>1</sup>. The sources and microphones were considered omnidirectional. The room setup is shown in Figure 1, in which we can see the source positions used in the simulations. The spatial aliasing frequency for the array is  $f_a = \frac{c}{2d \sin(\theta_{\max})}$  [63], where  $\theta_{\max}$  is the maximum searching angle and  $d$  is the inter-microphone distance of the ULA. In our setup,  $\theta_{\max} = 48.22$  degree, resulting in  $f_a = 3285.5$  Hz. The DOA limit to avoid aliasing up to 8000 Hz is 17.83 degree. By referring to the source positions in Figure 1, we have that about half of them are located in the region where aliasing occurs above 8000 Hz (the positions in front of the array), whereas the other ones are located where aliasing occurs in the range [3285.5, 8000] Hz. In

<sup>1</sup><http://www-mmsep.ece.mcgill.ca/Documents/Data>

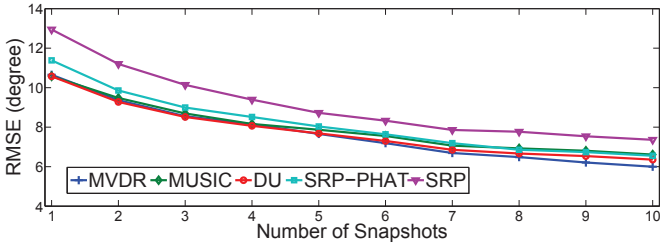


Fig. 3. Localization performance of a male speech signal at variation of number of snapshots with  $RT_{60}=0.4$  s and  $SNR=20$  dB.

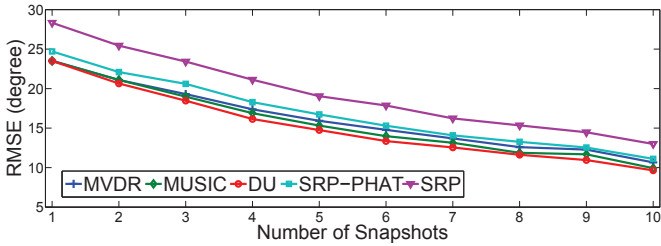


Fig. 4. Localization performance of a male speech signal at variation of number of snapshots with  $RT_{60}=0.4$  s and  $SNR=0$  dB.

our experiments, the percentage of narrowband components affected by aliasing is small in general, and do not exceed 40% in the worst case (i.e., when the source is at an angular position of  $\theta_{\max}$ ). We have observed that even in the worst case the aliasing related artifacts do not significantly affect the localization performances.

First, we analyzed the narrowband responses when a sinusoidal signal impinges the array. The beam patterns for a frequency of 1000 Hz with a DOA of -18 degree are shown in Figure 2. We can observe the high resolution response of the DU beamforming. We can also note that the beam patterns of DU, MVDR, and MUSIC are indistinguishable and, if compared to the conventional beamformer, they all provide better resolution and a better rejection of unsought energy components.

Next, the localization of a male speech signal of a duration of 25 s was investigated. The overall performance was computed with a simulation for each position depicted in Figure 1. The localization performance for different numbers of snapshots used to estimate the PSD matrix was conducted. Figures 3 and 4 show the results for a reverberation time ( $RT_{60}$ ) of 0.4 s with an SNR of 20 dB and 0 dB respectively. The DU performance is in general comparable to that of MVDR and MUSIC. In case of moderate noise (Figure 3), and when the number of snapshots does not exceed four, DU provides the best performance if compared to other methods. When the number of snapshots exceeds four, MVDR outperforms DU due to the improvement of accuracy in the estimated PSD matrix. When the noise is higher (Figure 4), DU provides the best performance for whatever number of snapshots, due to its robustness with respect to PSD estimation errors, for the reasons discussed in Section IV.

Next, we have conducted an analysis for different values of SNR in the range  $[-20, 20]$  dB with  $RT_{60}=0.4$  s and a number of snapshots equal to 10. Figure 5 shows the results

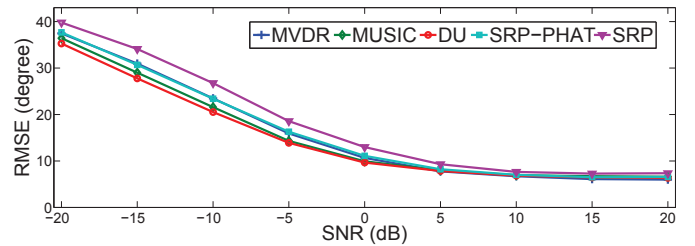


Fig. 5. Localization performance of a male speech signal at variation of SNR with  $RT_{60}=0.4$  s and a number of snapshots equal to 10.

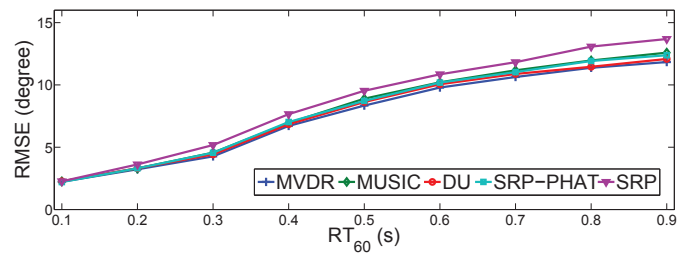


Fig. 6. Localization performance of a male speech signal at variation of  $RT_{60}$  with  $SNR=10$  dB and a number of snapshots equal to 10.

TABLE IV  
THE RMSE (DEGREE) OF LOCALIZATION PERFORMANCE WITH AND WITHOUT NORMALIZATION OF A SINGLE SOURCE WITH  $RT_{60}=0.4$  s,  $SNR=10$  dB, AND A NUMBER OF SNAPSHOTS EQUAL TO 10.

	MVDR	MUSIC	DU	SRP-PHAT	SRP
with norm.	6.708	6.946	6.847	7.025	7.655
without norm.	13.852	10.643	8.311	13.175	13.175

that demonstrate the robustness against noise of the DU. We can notice that the DU provides a comparable performance if compared to MUSIC method when SNR is 0 and -5 dB. Afterwards, simulations were conducted to analyze the effects of reverberation on the DU algorithm with  $SNR=10$  dB and a number of snapshots equal to 10. As it can be seen in Figure 6, DU is characterized by a performance very close to that of the MVDR, which has the lowest RMSE in all reverberant conditions.

To evaluate the impact of normalization of equation (8) in the DU, we have performed a simulation without it. Table IV shows a localization comparison with  $RT_{60}=0.4$  s,  $SNR=10$  dB, and a number of snapshots equal to 10, with and without the normalization of narrowband components for all methods. As we can observe, the normalization provides an error reduction of the DOA estimation for all methods.

Moreover, we have evaluated the localization performance by using the optimal DU implementation, assuming two cases: 1) an estimation of the noise, which is computed when the source signal is not active; 2) a perfect knowledge of the noise in each frame, i.e. guaranteeing that the signal subspace is always completely removed. Simulation was computed with  $RT_{60}=0.4$  s,  $SNR=10$  dB, and a number of snapshots equal to 10. We can note in Table V that the performance of the DU is the same when the noise is estimated, since the signal subspace is not completely canceled due to the errors in the PSD matrix estimation and the averaging in the noise power estimation. On

TABLE V

THE RMSE (DEGREE) OF LOCALIZATION PERFORMANCE WITH OPTIMAL DU IMPLEMENTATION OF A SINGLE SOURCE WITH  $RT_{60}=0.4$  s,  $SNR=10$  dB, AND A NUMBER OF SNAPSHOTS EQUAL TO 10.

MVDR	MUSIC	DU	DU (noise est.)	DU (noise know.)	SRP-PHAT	SRP
5,9758	6,5893	6,3285	6,3285	6,288	6,5493	7,3803

the other hand, when the noise was assumed to be known in each frame, only a very slight improvement can be noticed, which confirms the good performance of the suboptimal DU implementation in practical applications. Note that the MUSIC performance has a higher RMSE if compared to the DU with the noise known in each frame. Both methods totally remove the signal subspace in the localization computation. However, as we discussed in Section IV, the gains of noise subspaces in the MUSIC are all set to unity, whereas they can be minor or equal to 1 in the DU method, since the diagonal procedure assigns a lower gain to larger noise eigenvalues and a gain equal to 1 to the smallest eigenvalue.

The effect of the signal subspace contribution is then reported in Figure 8, that shows the DU localization performance at variations of the signal subspace gain given from (55). The gain was set to a specific value in the range  $[0.1, 0.9]$  in the simulation with a male speech signal,  $RT_{60}=0.4$  s,  $SNR=20$  dB, and a single snapshot. After the diagonal removal using the suboptimal implementation, the PSD matrix is analyzed using the single value decomposition. The eigenvalues are normalized, and the signal eigenvalue, i.e the signal gain  $\hat{G}_{DU}$  from (55), is forced to a specific value. Then, the transformed PSD matrix is reconstructed with the new eigenvalues. We can observe a degradation performance when the gain increases. Figure 8 reports also the localization performance of others methods. When the gain increases, the exploitation of the orthogonality property is attenuated, and the performance approaches that of the conventional beamformer. Note that when the gain is equal to 1, it corresponds to the case of absence of the signal, resulting in a flat response, and when the gain is greater than 1, the DU reduces to the conventional SRP.

Next, the single source localization performance with a larger number of snapshots, ranging from 20 to 80, is evaluated in very low noise condition. Figure 7 shows a localization comparison with  $RT_{60}=0.4$  s and  $SNR=-10$  dB. The results shows that when the number of snapshots increases, the estimation of the PSD matrix will be more accurate and the MVDR provides a larger signal subspace attenuation, and hence a better localization performance, which confirms the theoretical comparison in (51).

Then, we present numerical examples to verify the DOA estimation in case of two sources. A male speech signal and a female speech signal, both of duration equal to 25 s and without silent parts, were used. In each trial, two positions were randomly selected from the ones depicted in Figure 1, and were used as source positions to simulate the acoustic propagation. A mean power ratio of 1 was imposed between them. We have assumed that the noise eigenvectors have dimension  $N-2$  for the MUSIC method. The results of 30 trials

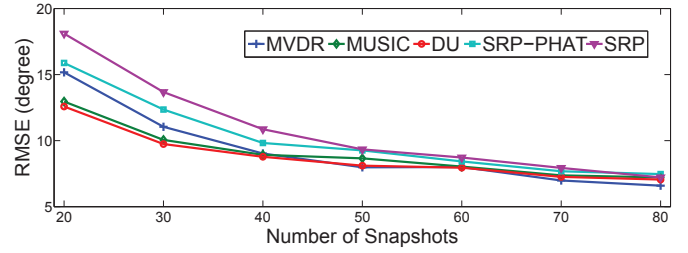


Fig. 7. Localization performance of the DU beamformer with a larger number of snapshots with  $RT_{60}=0.4$  s and  $SNR=-10$  dB.

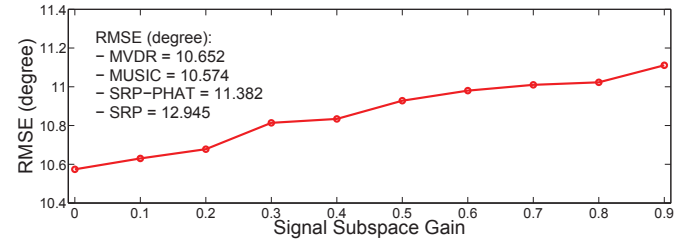


Fig. 8. Localization performance of the DU beamformer at variation of signal subspace gain  $\hat{G}_{DU}$  with  $RT_{60}=0.4$  s and  $SNR=20$  dB.

TABLE VI

THE COMPLEXITY EXPRESSED AS TOTAL NUMBER OF FLOPS OF BROADBAND BEAMFORMERS WITH AN FFT SIZE  $L$  OF 2048 SAMPLES.

	N=8	N=16
Conventional - DU	$24 \cdot 10^3$	$32 \cdot 10^3$
MVDR - MUSIC	$540 \cdot 10^3$	$4210 \cdot 10^3$

at variation of the number of snapshots is depicted in Figure 9 with  $RT_{60}=0.4$  s and  $SNR=10$  dB. MUSIC outperforms DU when the number of snapshots increases since it guarantees the total removal of both signal subspaces. However, DU provides a good performance in the multisource scenario, comparable to that of MUSIC for the most part of the parameters investigated. This is also confirmed in the results depicted in Figure 10 and 11 that show the localization performance of two speech signals at variation of SNR and of  $RT_{60}$ , respectively. We can observe in Figure 10 that the performance of the DU, the MVDR, the MUSIC and the SRP-PHAT is very similar in the SNR range  $[10, 20]$  dB, whereas in the range  $[-5, 5]$  dB a better performance of the DU and the MUSIC is measured. In Figure 11, MUSIC outperforms other methods at moderate reverberation, and all methods provide degraded localization performances when the reverberation time increases.

Finally, a complexity comparison is depicted in Table VI using an FFT size  $L$  of 2048 samples. We can observe the significant reduction of total number of flops of the DU beamformer in comparison of state-of-the-art robust beamforming.

### B. Real Data

The experiments were performed in a room of  $4.5 \text{ m} \times 3.75 \text{ m} \times 3.05 \text{ m}$  with  $RT_{60}=0.4$  s. The same array setup (i.e., an ULA of 8 sensors) and system parameters used for the simulated experiments were adopted. The distance between microphones was set to 0.07 m. A speech signal having

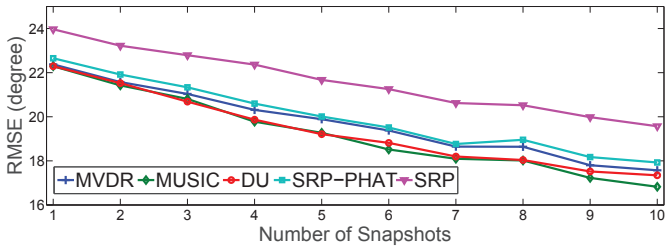


Fig. 9. Localization performance of two speech signals at variation of number of snapshots with  $RT_{60}=0.4$  s and  $SNR=10$  dB.

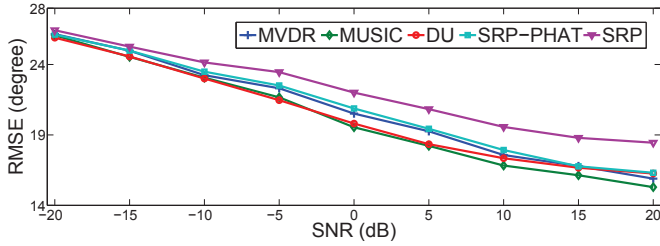


Fig. 10. Localization performance of two speech signals at variation of SNR with  $RT_{60}=0.4$  s and a number of snapshots equal to 10.

TABLE VII

THE RMSE (DEGREE) OF LOCALIZATION PERFORMANCE OF A SINGLE SOURCE WITH REAL DATA AND  $RT_{60}=0.4$  s AND A NUMBER OF SNAPSHOTS EQUAL TO 10.

MVDR	MUSIC	DU	SRP-PHAT	SRP
4.888	4.962	4.980	5.338	5.255

duration of 25 s, from a male speaker, was reproduced with a loudspeaker at a distance from the array of about 2 m with  $DOA=[26, 13, 6, -13, -19, -26]$  degrees. The results reported in Table VII confirm the effectiveness of the proposed DU beamformer. MVDR provides the best DOA estimation, which is however very close to that of the proposed DU.

## VII. CONCLUSIONS

We have proposed a data-dependent DU beamformer for acoustic source localization in microphone array signal processing. It consists of a transformation of the conventional beamformer into a high resolution and robust method with low complexity by an opportune covariance matrix conditioning operation. The DU procedure is designed to attenuate the signal subspace in the transformed covariance matrix in order to exploit the subspace orthogonality property. We have highlighted the role of the eigenvalues as weights for the attenuation of the signal eigenvectors in the calculation of the response power on one hand and, on the other, for the amplification of noise subspaces. We have introduced an eigenvalue analysis for a clear understanding of DU, robust MVDR and MUSIC, and of their relationships from the point of view of the transformed covariance matrices. The theoretical DU derivation for a single acoustic source in noisy conditions is based on an optimization problem with an orthogonality constraint and on a PSD matrix transformation imposing that the eigenvalue corresponding to the signal subspace is zero.

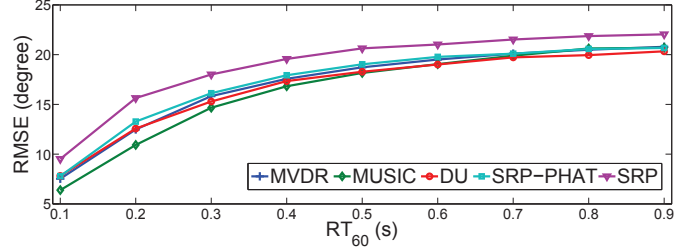


Fig. 11. Localization performance of two speech signals at variation of  $RT_{60}$  with  $SNR=10$  dB and a number of snapshots equal to 10.

This constraint determines how to exactly calculate the penalty weight in the DU operation. We have then analyzed the DU beamforming in a multisource scenario and we have introduced a suboptimal implementation for practical applications. Experiments conducted on both simulated and recorded acoustic data have been presented, which confirm the effectiveness of the method. Thus, the DU offers an attractive alternative to the current robust state-of-the-art localization beamformers at a computational cost comparable to that of the conventional beamformer. DU is robust with respect to errors in the covariance matrix estimation, and its performance does not depend on the heuristic determination of critical parameters, as it is the case of the penalty weight for the regularization of MVDR or the case of the estimated covariance matrix analysis in the MUSIC method.

## VIII. ACKNOWLEDGEMENTS

The authors are grateful to the anonymous reviewers for their constructive comments that greatly contributed to improve the manuscript.

## REFERENCES

- [1] F. Ribeiro, C. Zhang, D. A. Florencio, and D. E. Ba, "Using reverberation to improve range and elevation discrimination for small array sound source localization," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 7, pp. 1781–1792, 2010.
- [2] H. Sun, H. Teutsch, E. Mabande, and W. Kellermann, "Robust localization of multiple sources in reverberant environments using EB-ESPRIT with spherical microphone arrays," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, 2011, pp. 117–120.
- [3] A. Lombard, Y. Zheng, H. Buchner, and W. Kellermann, "TDOA estimation for multiple sound sources in noisy and reverberant environments using broadband independent component analysis," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1490–1503, 2011.
- [4] D. Salvati, S. Canazza, and A. Rodà, "A sound localization based interface for real-time control of audio processing," in *Proceedings of the International Conference on Digital Audio Effects*, 2011, pp. 177–184.
- [5] A. Canclini, F. Antonacci, A. Sarti, and S. Tubaro, "Acoustic source localization with distributed asynchronous microphone networks," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 21, no. 2, pp. 439–443, 2013.
- [6] D. Salvati and S. Canazza, "Incident signal power comparison for localization of concurrent multiple acoustic sources," *The Scientific World Journal*, vol. 2014, pp. 1–13, 2014.
- [7] M. J. Taghizadeh, A. Asaei, S. Haghighatshoar, P. N. Garner, and H. Bourlard, "Spatial sound localization via multipath euclidean distance matrix recovery," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 802–814, 2015.

- [8] A. Canclini, P. Bestagini, F. Antonacci, M. Compagnoni, A. Sarti, and S. Tubaro, "A robust and low-complexity source localization algorithm for asynchronous distributed microphone networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 10, pp. 1563–1575, 2015.
- [9] D. Salvati, C. Drioli, and G. L. Foresti, "Sound source and microphone localization from acoustic impulse responses," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1459–1463, 2016.
- [10] X. Li, L. Girin, R. Horaud, and S. Gannot, "Estimation of the direct-path relative transfer function for supervised sound-source localization," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 11, pp. 2171–2186, 2016.
- [11] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.
- [12] J. Benesty, "Adaptive eigenvalue decomposition algorithm for passive acoustic source localization," *Journal of the Acoustical Society of America*, vol. 107, no. 1, pp. 384–391, 2000.
- [13] J. O. Smith and J. S. Abel, "Closed-form least-squares source location estimation from range-difference measurements," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 12, pp. 1661–1669, 1987.
- [14] Y. Huang, J. Benesty, G. W. Elko, and R. M. Mersereau, "Real-time passive source localization: a practical linear-correction least-squares approach," *IEEE Transactions on Speech and Audio Processing*, vol. 9, no. 8, pp. 943–956, 2001.
- [15] P. Stoica and J. Li, "Source localization from range-difference measurements," *IEEE Signal Processing Magazine*, vol. 23, no. 3, pp. 63–66, 2006.
- [16] M. Omologo, P. Svaizer, and R. De Mori, *Spoken Dialogue with Computers*. Academic Press, 1998, ch. Acoustic Transduction.
- [17] J. H. DiBiase, H. F. Silverman, and M. S. Brandstein, *Microphone Arrays: Signal Processing Techniques and Applications*. Springer, 2001, ch. Robust localization in reverberant rooms.
- [18] P. Aarabi, "The fusion of distributed microphone arrays for sound localization," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 4, pp. 338–347, 2003.
- [19] D. B. Ward, E. A. Lehmann, and R. C. Williamson, "Particle filtering algorithms for tracking an acoustic source in a reverberant environment," *IEEE Transactions on Speech and Audio Processing*, vol. 11, no. 6, pp. 826–836, 2003.
- [20] P. Pertilä, T. Korhonen, and A. Visa, "Measurement combination for acoustic source localization in a room environment," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2008, pp. 1–14, 2008.
- [21] J. Velasco, D. Pizarro, and J. Macias-Guarasa, "Source localization with acoustic sensor arrays using generative model based fitting with sparse constraints," *Sensors*, vol. 12, no. 10, pp. 13 781–13 812, 2012.
- [22] L. O. Nunes, W. A. Martins, M. V. S. Lima, L. W. P. Biscainho, M. V. M. Costa, F. M. Gonalves, A. Said, and B. Lee, "A steered-response power algorithm employing hierarchical search for acoustic source localization using microphone arrays," *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 5171–5183, 2014.
- [23] D. Salvati, C. Drioli, and G. L. Foresti, "A weighted MVDR beamformer based on SVM learning for sound source localization," *Pattern Recognition Letters*, vol. 84, pp. 15–21, 2016.
- [24] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.
- [25] R. Roy and T. Kailath, "ESPRIT - estimation of signal parameters via rotational invariance techniques," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.
- [26] B. D. Rao and K. V. S. Hari, "Performance analysis of root-music," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 12, pp. 1939–1949, 1989.
- [27] K. Harmanci, J. Tabrikian, and J. L. Krolik, "Relationships between adaptive minimum variance beamforming and optimal source localization," *IEEE Transactions on Signal Processing*, vol. 48, no. 1, pp. 1–12, 2000.
- [28] C. Zhang, D. Florencio, D. E. Ba, and Z. Zhang, "Maximum likelihood sound source localization and beamforming for directional microphone arrays in distributed meetings," *IEEE Transactions on Multimedia*, vol. 10, no. 3, pp. 538–548, 2008.
- [29] J. Traa, D. Wingate, N. D. Stein, and P. Smaragdakis, "Robust source localization and enhancement with a probabilistic steered response power model," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 3, pp. 493–503, 2016.
- [30] M. S. Bartlett, "Smoothing periodograms from time-series with continuous spectra," *Nature*, vol. 161, pp. 686–687, 1948.
- [31] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969.
- [32] J. Li and P. Stoica, Eds., *Robust Adaptive Beamforming*. Wiley, 2005.
- [33] H. Cox, R. Zeskind, and M. Owen, "Robust adaptive beamforming," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 10, pp. 1365–1376, 1987.
- [34] J. Li, P. Stoica, and Z. Wang, "On robust Capon beamforming and diagonal loading," *IEEE Transactions on Signal Processing*, vol. 51, no. 7, pp. 1702–1715, 2003.
- [35] L. Hongqing, L. Guisheng, and Z. Jie, "A robust adaptive Capon beamforming," *Signal Processing*, vol. 86, no. 10, pp. 2820–2826, 2006.
- [36] X. Mestre and M. A. Lagunas, "Finite sample size effect on minimum variance beamformers: optimum diagonal loading factor for large arrays," *IEEE Transactions on Signal Processing*, vol. 54, no. 1, pp. 69–82, 2006.
- [37] C. Liu and G. Liao, "Robust Capon beamformer under norm constraint," *Signal Processing*, vol. 90, no. 5, pp. 1573–1581, 2010.
- [38] Y. Xua, J. Maa, Z. Liua, and W. Liub, "A class of diagonally loaded robust Capon beamformers for noncircular signals of interest," *Signal Processing*, vol. 94, pp. 670–680, 2014.
- [39] J. Dmochowski, J. Benesty, and S. Affes, "Linearly constrained minimum variance source localization and spectral estimation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 16, no. 8, pp. 1490–1502, 2008.
- [40] T. Yardibi, J. Li, P. Stoica, M. Xue, and A. B. Baggeroer, "Source localization and sensing: A nonparametric iterative adaptive approach based on weighted least squares," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 1, pp. 425–443, 2010.
- [41] Y. Gu and A. Leshem, "Robust adaptive beamforming based on interference covariance matrix reconstruction and steering vector estimation," *IEEE Transactions on Signal Processing*, vol. 60, no. 7, pp. 3881–3885, 2012.
- [42] S. D. Somasundaram, A. Jakobsson, and N. H. Parsons, "Robust and automatic data-adaptive beamforming for multidimensional arrays," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 11, pp. 4642–4656, 2012.
- [43] H. Abeida, Q. Zhang, J. Li, and N. Merabtine, "Iterative sparse asymptotic minimum variance based approaches for array processing," *IEEE Transactions on Signal Processing*, vol. 61, no. 4, pp. 933–944, 2013.
- [44] L. Tzafri and A. J. Weiss, "High-resolution direct position determination using MVDR," *IEEE Transactions on Wireless Communications*, vol. 15, no. 9, pp. 6449–6461, 2016.
- [45] D. W. Rieken and D. R. Fuhrmann, "Generalizing MUSIC and MVDR for multiple noncoherent arrays," *IEEE Transactions on Signal Processing*, vol. 52, no. 9, pp. 2396–2406, 2004.
- [46] J. P. Dmochowski, J. Benesty, and S. Affes, "Broadband MUSIC: opportunities and challenges for multiple source localization," in *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2007, pp. 18–21.
- [47] B. Wang, Y. Zhao, and J. Liu, "Mixed-order MUSIC algorithm for localization of far-field and near-field sources," *IEEE Signal Processing Letters*, vol. 20, no. 4, pp. 311–314, 2013.
- [48] L. Kumar, A. Tripathy, and R. M. Hegde, "Robust multi-source localization over planar arrays using MUSIC-group delay spectrum," *IEEE Transactions on Signal Processing*, vol. 62, no. 17, pp. 4627–4636, 2014.
- [49] G. Liu, H. Chen, X. Sun, and R. C. Qiu, "Modified MUSIC algorithm for DOA estimation with nystrm approximation," *IEEE Sensors Journal*, vol. 16, no. 12, pp. 4673–4674, 2016.
- [50] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 1053–5888, 1996.
- [51] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Prentice-Hall, 2005.
- [52] B. D. Carlson, "Covariance matrix estimation errors and diagonal loading in adaptive arrays," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 24, no. 4, pp. 397–401, 1988.
- [53] L. Zhang, W. Liu, and L. Yu, "Performance analysis for finite sample mvdr beamformer with forward backward processing," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2427–2431, 2011.
- [54] D. Salvati, C. Drioli, and G. L. Foresti, "Incoherent frequency fusion for broadband steered response power algorithms in noisy environments," *IEEE Signal Processing Letters*, vol. 21, no. 5, pp. 581–585, 2014.
- [55] W. Liu and S. Ding, "An efficient method to determine the diagonal loading factor using the constant modulus feature," *IEEE Transactions on Signal Processing*, vol. 56, no. 12, pp. 6102–6106, 2008.

- [56] L. Du, J. Li, and P. Stoica, "Fully automatic computation of diagonal loading levels for robust adaptive beamforming," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, no. 1, pp. 449–458, 2010.
- [57] B. Ng, M. Er, and C. Kot, "A MUSIC approach for estimation of directions of arrival of multiple narrowband and broadband sources," *Signal Processing*, vol. 40, no. 2, pp. 319–323, 1994.
- [58] S. A. Vorobyov, "Principles of minimum variance robust adaptive beamforming design," *Signal Processing*, vol. 93, no. 12, pp. 3264–3277, 2013.
- [59] P. Gerstoft, R. Menon, W. S. Hodgkiss, and C. F. Mecklenbräuker, "Eigenvalues of the sample covariance matrix for a towed array," *Journal of the Acoustical Society of America*, vol. 132, no. 4, pp. 2388–2396, 2012.
- [60] J. F. Synnevag, A. Austeng, and S. Holm, "Adaptive beamforming applied to medical ultrasound imaging," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 54, no. 8, pp. 1606–1613, 2007.
- [61] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.
- [62] E. Lehmann and A. Johansson, "Prediction of energy decay in room impulse responses simulated with an image-source model," *Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 269–277, 2008.
- [63] J.-W. Shin, Y.-J. Lee, and H.-N. Kim, "Reduced-complexity maximum likelihood direction-of-arrival estimation based on spatial aliasing," *IEEE Transactions on Signal Processing*, vol. 62, no. 24, pp. 6568–6581, 2014.



**Daniele Salvati** received the Laurea degree in Environmental Engineering from the Sapienza University of Rome, Italy, in 2003, the postgraduate Master degree in Sound Engineering from Tor Vergata University of Rome, Italy, in 2006, and the Ph.D. degree in Multimedia Communication from the University of Udine, Italy, in 2012, with a research on acoustic source localization using microphone arrays. Since 2013 he is a postdoctoral researcher at the Department of Mathematics, Computer Science and Physics, University of Udine, Italy. He was a system

and audio consultant to many information technology companies from 2001 to 2008. His research interests are in audio and acoustic signal processing and multimedia communications.



**Carlo Drioli** is Assistant Professor at the Department of Mathematics, Computer Science and Physics of the University of Udine. He received the Laurea degree in Electronic Engineering and the Ph.D. degree in Electronic and Telecommunications Engineering from the University of Padova, in 1996 and 2003, respectively. He has been a researcher with the Centro di Sonologia Computazionale (CSC) of the University of Padova, in the field of sound and voice analysis and processing; visiting researcher at the Royal Institute of Technology (KTH), Stock-

holm, Sweden, with the support of the European Community through a Marie Curie Fellowship; researcher at the Department of Phonetics and Dialectology of the Institute of Cognitive Sciences and Technology of the Italian National Research Council (ISTC-CNR), where he pursued research on voice processing and emotional speech synthesis; research assistant and adjunct professor at the Department of Computer Science of the University of Verona. His current research interests are in the fields of multimedia signal processing, sound and voice coding by means of physical modeling, speech analysis and synthesis, array signal processing. He is a member of the Institute of Electrical and Electronics Engineers (IEEE), of the Acoustical Society of America (ASA), and of the International Speech Communication Association (ISCA).



**Gian Luca Foresti** is Full Professor of Computer Science at the University of Udine and Director of the Dept. of Mathematics, Computer Science and Physics. He is the appointed Italian member of the NATO RTO Information System Technology Panel. He was Finance Chair of the 11th IEEE Conference on Image Processing (ICIP05), General Chair of the 16th Int. Conf. on Image Analysis and Processing (ICIAP11) and of the 8th IEEE Conf. on Advanced Video and Signal Based Surveillance (AVSS11). His main interests involve Computer Vision and Image Processing, Multisensor Data and Information Fusion, Pattern Recognition and Deep Neural Networks. Prof. Foresti is author of more than 300 papers published in International Journals and International Conferences and he has been co-editor of several books in the field of Multimedia and Computer Vision. He has been Guest Editor of a Special Issue of the Proceedings of the IEEE on Video Communications, Processing and Understanding for Third Generation Surveillance Systems. In 2002, he has been awarded of best IEEE Vehicular Electronics paper, in 2010 of the Best paper Award at the International Conference on Distributed Smart Cameras (ICDSC 2010) and in 2016 of the Best Industry Related Paper Award (BIRPA) at the International Conference on Pattern Recognition (ICPR16). Prof. Foresti is Fellow member of IAPR and Senior member of IEEE.