This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2017.2700945, IEEE/ACM Transactions on Audio, Speech, and Language Processing

Generation of private sound with a circular loudspeaker array and the Weighted Pressure Matching method

Ferdinando Olivieri, *Member, IEEE,* Filippo Maria Fazi, Simone Fontana, *Member, IEEE,* Dylan Menzies and Philip Arthur Nelson, *Member, IEEE,*

***Abstract*—In this work, we propose a system for private sound which is based on the Weighted Pressure Matching method (WPMM). The aim is to design the input signals to an array of loudspeakers which allow for the synthesis of a target field defined with large amplitude variations between the so-called dark points and the listener's position. The system enables listeners to control the trade-off between directivity performance and the accuracy of reproduction of the target field at the listening position when the input energy to the array is limited. This is achieved by calculating the WPMM weight in the dark zone based on a performance constraint on the characteristics of the sound field in the listening zone. The system is validated for a number of pre-defined use-case scenarios. The results of the experiments in an anechoic environment with a circular array prototype show that listeners can control the performance trade-off in a wide frequency range. In the second part of the paper, algorithms are presented for the fast update of the input signals when the user selects a new value of the performance constraint.**

I. Introduction

Loudspeaker arrays driven with purposefully-designed input signals can be used for the implementation of systems for private sound, a technology that allows listeners to privately listen to any selected audio material without affecting other people located in the same environment [1], [4], [5]. The input signals may be designed by means of beamforming methods, such as the Time-Reversal Mirror [3], sound focusing [18], the Minimum Variance Distortionless Response [2], as well as methods that aim at the control of the acoustic energy density in the control area [6], [7], [4], [8], hereafter also referred to as energy-based methods, and methods that aim at the synthesis of a target field defined at a set of control points, such as the Pressure Matching Method (PMM) [17], [8] and the Weighted Pressure Matching Method (WPMM) [13], [9]. In this work, we limit our analysis to the PMM and the WPMM, as they may provide better sound quality than that achieved with energy-based methods [8], [24], [9], [18].

In private sound applications, the listener's position is usually referred to as the "bright point" and the rest of the control zone where sound radiation is to be minimized is referred to as the "dark zone". The target field is typically chosen with a large difference in level between the bright point and the dark zone. However, for a given array configuration and a given wavelength, a target field of this kind may be difficult to be reproduced especially if the bright point and the dark zone are adjacent to each other [11]. Previous works reported that arrays driven with input signals designed with the PMM may suffer of limited ability to preserve the characteristics of the spectrum of the original audio material to be delivered to the listener [12], [10]. This may be related to the fact that the transfer functions between bright and dark points are similar at low frequencies [8], where the size of the array is small compared to the wavelength. In such scenarios, the system, in an attempt to provide the large level variations defined in the target field, achieves a reproduced field at the bright point which may substantially differ from the target field [10]. This is an undesired effect which may lead to increased distraction of the listener [32].

More generally, a performance trade-off exists between directivity performance, input energy required by the array to perform directional sound radiation, and accuracy of reproduction of the target field in the listening area [18], [12], [23], [24], hereafter succinctly referred to as "quality". For example, for a given level of the input energy, a given system may be able to provide high directivity performance at the expense of a reduced quality in the listening zone [9], [10], [12]. A number of PMM-based methods were proposed to control the performance trade-off by means of 1) parameters that are freely customizable by the users [13], [9], [12] or, 2) by calculating the input signals that satisfy given performance constraints [14], [15]. Some of these methods are based on the WPMM [13], [9], [14], [12], a generalized formulation of the PMM wherein the so-called WPMM weights are assigned to the control points where sound is to be controlled with the aim of prioritizing (or not prioritizing) accurate reproduction of the target field at given control points [14], [9]. We consider the problem of calculating the input signals that satisfy a performance constraint [14], [15]. In this work, we set a performance constraint on the accuracy of the reproduction of the target field at the bright point, which is, hereafter, succinctly referred to as "quality constraint". Since

Manuscript received xxxx; revised xxxx.

F. Olivieri was with the Institute of Sound and Vibration Research, University of Southampton, Southampton, SO17 1BJ, UK. (e-mail: f.olivieri@ieee.org).

F. M. Fazi, D. Menzies, and P. A. Nelson are with the Institute of Sound and Vibration Research, University of Southampton, Southampton, SO17 1BJ, UK. (e-mail: filippo.fazi@soton.ac.uk; D.Menzies@soton.ac.uk; P.A.Nelson@soton.ac.uk).

S. Fontana is with the Huawei European Research Center, Riesstraße 25, 80992, Munich, Germany. (e-mail: Simone.Fontana@huawei.com).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org and https://zenodo.org/badge/latestdoi/89870495, provided by the author. The material includes Matlab code and data files and its size is 80 MB. Contact F. Olivieri (f.olivieri@ieee.org) for further questions about this work.

2329-9290 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2017.2700945, IEEE/ACM Transactions on Audio, Speech, and Language Processing

quality and directivity performance are interrelated (when the input energy is limited), by setting the quality constraint, the user can control the directivity/quality trade-off.

We consider the following scenario: the listener is given the ability to customize the value of the quality constraint. Ideally, the system quickly updates the input signals that satisfy the new performance constraint. To this end, methods can be used that are based on iterative searches of the input signals [14], [15]. However, these methods may be time consuming and, hence, not be suitable for practical applications. In this work, we address this problem by proposing a system that 1) can be used to design the input signals that satisfy a desired quality constraint and 2) quickly updates the input signals (once the user sets a new value of the quality constraint). In order to accomplish that, we split this work in two parts.

1) We firstly design a method (that is based on the WPMM) to control the performance trade-off between quality and directivity when the input energy is limited. In the proposed formulation, the WPMM weight in the dark zone is calculated based on a performance constraint on the reproduced field in the bright zone. We validate the proposed method in a number of use-case scenarios where different configurations of the control zones (e.g., bright or dark) are considered. The optimal input signals are initially calculated by performing a full-search of the WPMM weight in the dark zone.

2) We then address the problem of the fast update of the input signals by proposing two methods: 1) an iterative algorithm and 2) a method based on the approximation of the input signals by means of a truncated Neumann series. This method is based on the calculation of the zeros of a polynomial and guidelines are proposed for the calculation of the order of truncation of the Neumann series. We show that this method can accurately estimate, over a wide frequency range, the value of the WPMM weight in the acoustically dark zone that satisfies the performance constraint. Furthermore, we discuss under which conditions the second method is more computationally efficient and less time-consuming than the iterative methods.

System performance is evaluated by means of numerical simulations and experiments in an anechoic environment with a circular array prototype mounted on a rigid cylinder. Furthermore, strategies for using the proposed system in a reverberant environment are discussed. The proposed methods can be readily extended to other array and control zone geometries, provided that the transfer functions are known.

This work is organized as follows: the theoretical background is reported in Section II. The system design is described in Section III. The experimental validation of the proposed system is in Section IV. The method for the update of the input signals is reported in V. Possible strategies to use the system in a reverberant environment are discussed in Section VI and the conclusions are drawn in Section VII. We provide a Matlab implementation of the methods presented in this work [51].

II. Theoretical Background

We define the vectors $\mathbf{0}_Y^T = [0, 0, \cdots, 0]$ and $\mathbf{1}_Y^T = [1, 1, \cdots, 1]$ of length Y where $[\cdot]^T$ indicates the operation of vector (or matrix) transpose. Given a square matrix $\mathbf{Y}$, $\mathbf{Y}^n$ indicates the multiplication of $\mathbf{Y}$ by itself n times, e.g. $\mathbf{Y}^3 = \mathbf{YYY}$. All the acoustical quantities introduced in this work have a time dependence of $e^{-i\omega t}$, where $i = \sqrt{-1}$ is the imaginary unit and t is time, and ω is a given angular frequency. We consider a L channel circular array prototype, $L = 32$, shown in Fig. 1, whose transducers are uniformly arranged on a circle of radius r, $r = 0.25\,\mathrm{m}$. The angular polar coordinate of the ℓ-th loudspeaker, $\ell \in [1, L]$, is $\phi_\ell = (2\pi/L)\ell$. The array is mounted on a rigid cylindrical body whose height is $0.35\,\mathrm{m}$. The control zone consists of M control points, $M = 72$, where virtual microphones are uniformly arranged on a circle of radius R, $R = 3\,\mathrm{m}$. The loudspeaker array and the control points lie on the same plane. The angular coordinate of the m-th control point is $\theta_m = (2\pi/M)m$, $m \in [1, M]$. A single bright point is considered that is located at the point of angular coordinate θ_B, $\theta_B = 90°$. The ℓ-th transducer of the array is driven by the input signal $q_\ell(\omega)$. The vector of the input signals is $\mathbf{q}^T(\omega) = [q_1(\omega), \ldots, q_L(\omega)]$. The output signal to a virtual microphone placed at the m-th control point to estimate the sound pressure generated by the array driven with $\mathbf{q}(\omega)$ is denoted by $p(\mathbf{x}_m, \omega)$. The control area consists of M control points and the vector of the virtual microphone signals is $\mathbf{p}^T(\omega) = [p(\mathbf{x}_1, \omega), \ldots, p(\mathbf{x}_M, \omega)]$. The relation between the vectors $\mathbf{p}(\omega)$ and $\mathbf{q}(\omega)$ is

$$\mathbf{p}(\omega) = \mathbf{Z}(\omega)\mathbf{q}(\omega), \tag{1}$$

where $\mathbf{Z}(\omega)$ is the matrix of Transfer Functions (TFs) between the virtual microphones and the input signals. Hereafter, the dependence on ω is omitted to simplify the notation.

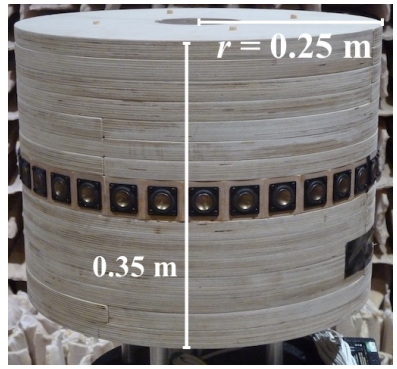


Fig. 1. Details of the circular array prototype mounted on a rigid cylinder.

The circular array is modeled as a discrete distribution of monopoles radiating in free-field and installed on an infinite rigid cylinder. The (m, ℓ)-th element $Z_{m,\ell}$ of $\mathbf{Z}$ is the far-field complex radiation pattern generated by this distribution of sources on the horizontal plane (where the distribution of monopoles lies) and it is calculated as [50], [16]

$$Z_{m,\ell} = \sum_{n=-K}^{K} \frac{2c\, i^{(1-n)}}{\pi\omega H_n'(r\omega/c)} e^{in\theta_m} e^{-in\phi_\ell}, \tag{2}$$

where c is the speed of sound, $c = 343\,\mathrm{m\,s^{-1}}$, $H_n'(r\omega/c)$ is the derivative of the Hankel function of the first kind of order n [16], and K is the maximum order (truncation). Here we choose $K = 150 \gg r\omega/c$. It is assumed that R is large enough

2329-9290 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2017.2700945, IEEE/ACM Transactions on Audio, Speech, and Language Processing

for the far-field assumption to be satisfied.

In private sound applications, the control points are usually divided into points where sound is desired or undesired, that are usually referred to as acoustically bright points and acoustically dark zone, respectively. Depending on the scenario, we also consider the "gray zone" [9], a particular case of dark zone that is not occupied by users. In the most general case, we partition $\mathbf{Z}$ and $\mathbf{p}$ as

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}_B^T \\ \mathbf{Z}_D \\ \mathbf{Z}_G \end{bmatrix}, \mathbf{p} = \begin{bmatrix} p_B \\ \mathbf{p}_D \\ \mathbf{p}_G \end{bmatrix} = \begin{bmatrix} \mathbf{z}_B^T \mathbf{q} \\ \mathbf{Z}_D \mathbf{q} \\ \mathbf{Z}_G \mathbf{q} \end{bmatrix}, \tag{3}$$

respectively, where p_B, $\mathbf{p}_D$, $\mathbf{p}_G$, $\mathbf{z}_B$, $\mathbf{Z}_D$, and $\mathbf{Z}_G$ are the microphone signals and transfer functions between the virtual microphones located at the bright, dark, and gray points and the input signals, respectively. The number of dark and gray points are M_D and M_G, respectively, and $M = M_D + M_G + 1$. A target field $\hat{\mathbf{p}}^T = [\hat{p}(\mathbf{x}_1), \ldots, \hat{p}(\mathbf{x}_M)]$ defined as [9]

$$\hat{\mathbf{p}} = \begin{bmatrix} \hat{p}_B \\ \hat{\mathbf{p}}_D \\ \hat{\mathbf{p}}_G \end{bmatrix} = \begin{bmatrix} 1 \\ \mathbf{0}_{M_D} \\ \mathbf{0}_{M_G} \end{bmatrix}, \tag{4}$$

where $\hat{p}_B$, $\hat{\mathbf{p}}_D$, and $\hat{\mathbf{p}}_G$, are the target fields defined at the bright, dark, and gray points, respectively. The target field $\hat{\mathbf{p}}$ can be reproduced by driving the array with input signals designed with the WPMM [9], [14]. The WPMM cost function is defined as [9], [14]

$$J(\mathbf{q}) = \psi_B |p_B - 1|^2 + \psi_D \|\mathbf{p}_D\|^2 + \psi_G \|\mathbf{p}_G\|^2 + \beta \|\mathbf{q}\|^2, \tag{5}$$

where $\|\cdot\|$ is the l_2-norm operator, $0 \leq \psi_B \leq 1$, $0 \leq \psi_D \leq 1$, $0 \leq \psi_G \leq 1$ are the WPMM weights for the reproduction errors at the bright, dark and gray points, respectively, and $\beta \in [0, \infty)$ is the Tikhonov regularization parameter [9] that serves to control the energy of the input signals [14], [9]. The terms $\|\mathbf{p}_D\|^2$ and $\|\mathbf{p}_G\|^2$ in Eq. (5) may be normalized by the number of control points in the dark and gray zone, respectively, and the normalization may be included in ψ_D and ψ_G. As reported in previous works [52], [53], the upper limit for the positive and real-valued WPMM weights is arbitrary. In this work, we restrict the range of the WPMM weights to values between 0 and 1 as in previous works on the WPMM for loudspeaker arrays [9], [13], [14]. The input signals that minimize $J(\mathbf{q})$ can be found by setting the partial derivative of $J(\mathbf{q})$ with respect to the real and the imaginary parts of $\mathbf{q}$ to zero and by solving with respect to $\mathbf{q}$, that is [9]

$$\mathbf{q} = \left(\psi_B \mathbf{z}_B^* \mathbf{z}_B^T + \psi_D \mathbf{Z}_D^H \mathbf{Z}_D + \psi_G \mathbf{Z}_G^H \mathbf{Z}_G + \beta \mathbf{I}\right)^{-1} \psi_B \mathbf{z}_B^*, \tag{6}$$

where $(\cdot)^*$, $(\cdot)^H$, and $(\cdot)^{-1}$ indicate the complex conjugate, complex conjugate transpose, and the inverse matrix, respectively, and $\mathbf{I}$ is the identity matrix. By defining the $M \times M$ diagonal matrix $\hat{\mathbf{\Psi}}$ whose elements on the main diagonal are $\left[\sqrt{\psi_B}, \sqrt{\psi_D}\mathbf{1}_{M_D}, \sqrt{\psi_G}\mathbf{1}_{M_G}\right]$, Eq. (6) can be also written as [9]

$$\mathbf{q} = \left(\mathbf{Z}^H \hat{\mathbf{\Psi}}^2 \mathbf{Z} + \beta \mathbf{I}\right)^{-1} \mathbf{Z}^H \hat{\mathbf{\Psi}}^2 \hat{\mathbf{p}}. \tag{7}$$

An appropriate modeling delay may be required to ensure that the input signals in Eq. (6) are causal [17].

III. System design

By using the WPMM and the TFs $\mathbf{Z}$, we design a system that can be used in a variety of settings and applications, hereafter referred to as use-case scenarios, which are defined by a given listener/control-zone configurations (i.e., changes in $\mathbf{z}_B$, $\mathbf{Z}_D$, and $\mathbf{Z}_G$). In order to design the input signals that provide a desired performance trade-off [13], [14], [9], appropriate WPMM weights are assigned to the control points where sound is to be controlled hence prioritizing (or not prioritizing) accurate reproduction at given control points [14], [9]. For example, higher values of ψ_B, ψ_D and ψ_G would result in higher accuracy of reproduction of the target field at the bright point, and at the dark and gray points, respectively. Thus, in order to ensure quality at the listener's position (that is, ideally, $p_B = \hat{p}_B$), maximum value of WPMM weight is given to the bright zone, i.e. $\psi_B = 1$. Conversely, a small value ψ_G is assigned to the WPMM weights in the gray zone, due to the fact that no accurate reproduction of the target field is required. By doing this, we also relax the constraint in the zones where no accurate reproduction is required [9]. Based on the desired performance trade-off between quality and directivity, we define the following use-case scenarios:

1) the Max-Quality Scenario (MQS), wherein the listener requires the highest possible quality (and no requirements for directivity performance). In this case, we consider only the bright point. The input signals $\mathbf{q}_{\text{MQS}}$ are calculated by means of a sound focusing technique [18], that is $\mathbf{q}_{\text{MQS}} = \mathbf{z}_B^* \big/ \left(\mathbf{z}_B^H \mathbf{z}_B\right)$. This scenario may be selected, for example, if the listener is alone in the environment.

2) In the Max-Directivity Scenario (MDS) the listener requires privacy and hence, highest directivity performance. All the control points but the bright point are set as dark points. The input signals are calculated by setting $\psi_D = 1$ in Eq. (6), thus yielding

$$\mathbf{q}_{\text{MDS}} = \left(\mathbf{z}_B^* \mathbf{z}_B^T + \mathbf{Z}_D^H \mathbf{Z}_D + \beta \mathbf{I}\right)^{-1} \mathbf{z}_B^*, \tag{8}$$

which corresponds to the PMM solution (also known as regularized least squares solution [9]). This has been, in turn, shown to be equivalent to a mode-matching approach, below the spatial aliasing frequency, as reported in reference [16]. The mode-matching approach for beamforming has also been used for microphone arrays [26]. This scenario may be suitable for listeners located in an environment where other people are present whose positions are likely to vary with time (e.g., the system is operating in a public space).

3) The Quality-Control Scenario (QCS) is similar to the MDS, but in this case the user can select a desired trade-off between quality and directivity by means of the parameter ψ_D. The input signals $\mathbf{q}_{\text{QCS}}$ are calculated as $\mathbf{q}_{\text{QCS}} = \left(\mathbf{z}_B^* \mathbf{z}_B^T + \psi_D \mathbf{Z}_D^H \mathbf{Z}_D + \beta \mathbf{I}\right)^{-1} \mathbf{z}_B^*$. Fig. 2a shows the arrangement of the control points for the QCS and the MDS.

4) The Hybrid-Scenario is similar to the QCS, the difference being that the zones that are not occupied by other people are labeled as gray zones [9] and as dark zones otherwise. In the context of this work, "Hybrid" is just a nomenclature that refers to a scenario where both dark and gray points are simultaneously present. Shin *et al.* named this scenario as

2329-9290 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2017.2700945, IEEE/ACM Transactions on Audio, Speech, and Language Processing

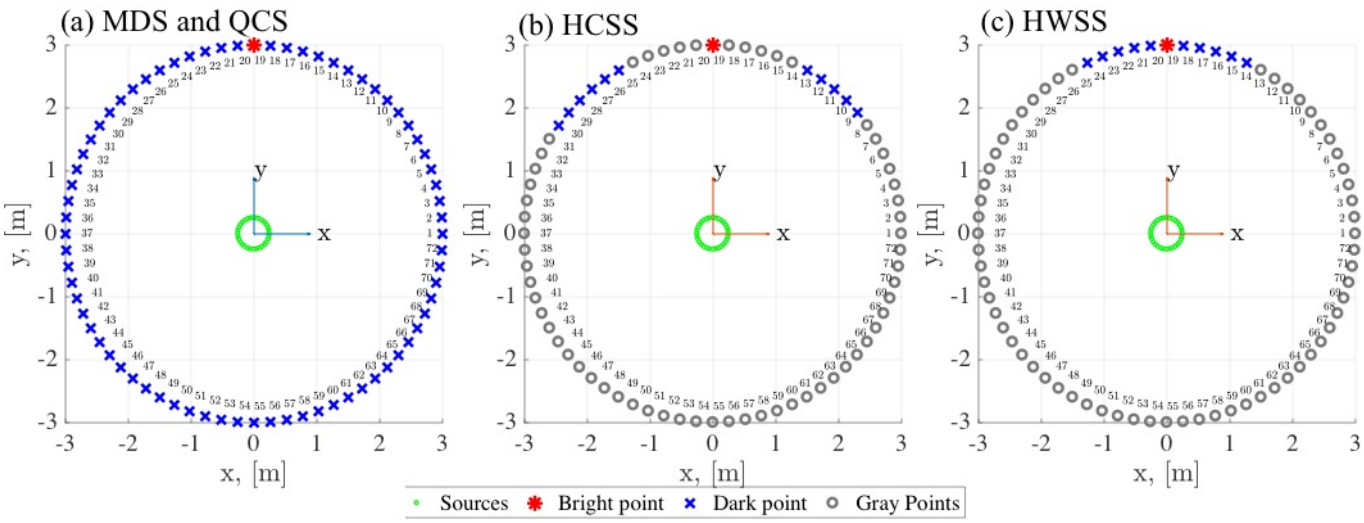


Fig. 2. Arrangement and numbering of the control points for the QCS and MDS (Fig. 2a) and for the two Hybrid-Scenarios (Fig. 2b and Fig. 2c, respectively).

"sound field control with a low-priority zone" [9]. The input signals are calculated with Eq. (6) and a low value of ψ_G. The formulation in Eq. (6) with $\psi_D = 1$ corresponds to that proposed in [9]. This scenario may be suitable when the positions of the other people are not likely to frequently vary with time (e.g., in a lounge room as opposed to a public space as in the MDS or in the QCS). In the case of the Hybrid scenario, system performance may depend on the relative distance between the bright and dark points. In this work, two different hybrid scenarios are analyzed. In the first one, shown in Fig. 2b, two sets of dark points are selected that are closely spaced (adjacent) to the bright point. In the second case, shown in Fig. 2c, the sets of dark points are widely spaced with respect to the bright point. The first and the second cases are hereafter referred to as Hybrid Closely-Spaced Scenario (HCSS) and Hybrid Widely-Spaced Scenario (HWSS), respectively.

A. Control of the directivity/quality performance trade-off

Previous works have addressed the problem of the control of the directivity/quality performance trade-off by means of the WPMM. In this work, the WPMM weight ψ_D is the parameter by means of which the user controls the performance trade-off between quality and directivity (when the array effort is limited and for given values of β and ψ_G) in the QCS and the Hybrid scenarios. In general, by increasing ψ_D, we increase the priority of the reproduction error in the dark zone (i.e., $\|\mathbf{p}_D\|^2$ in Eq. (5)), which would in turn provide higher attenuation of the Sound Pressure Level (SPL) in the dark zone and thus higher directional performance. However, this may come at the expenses of the quality at the listening position. In this work, we calculate ψ_D so that the input signals $\mathbf{q}$ satisfy (for given values of ψ_G and β) the following performance constraint, hereafter succinctly and loosely referred to as the "quality" constraint, that is

$$\mathbf{z}_B^T \mathbf{q}(\psi_D, \psi_G, \beta) = p_B \geq p_{B,\min}, \tag{9}$$

where $p_{B,\min} \leq \hat{p}_B$ is a real-valued parameter set by the user that indicates the minimum SPL that the user allows in the bright zone. In this work, $p_{B,\min}$ is chosen to be frequency independent. However, other criteria may be considered for the setting of $p_{B,\min}$ in a frequency-dependent fashion, which are, for example, based on psychoacoustical considerations or on the type of audio program (e.g., speech, music, etc). A thorough analysis of this topic lays beyond the scope of this paper and may be the object of future investigation.

The optimal value of ψ_D (in the sense of Eq. (9)), i.e. $\overline{\psi_D}$, is calculated as the maximum value of ψ_D that satisfies the quality constraint in Eq. (9). This choice of $\overline{\psi_D}$ ensures that we get the best directivity performance for a given $p_{B,\min}$. By setting the value of $p_{B,\min}$ (and, in turn, of $\overline{\psi_D}$), the user controls the desired level of "quality", expressed in dB as $20\log_{10}(p_{B,\min}/\hat{p}_B)$. We choose a real-valued $p_{B,\min}$ because $\mathbf{z}_B^T\mathbf{q}$ is real-valued and the proof is as follows. Let us define $\tilde{\mathbf{Z}}^{-1} = \left(\mathbf{z}_B^H\mathbf{z}_B + \psi_D\mathbf{Z}_D^H\mathbf{Z}_D + \psi_G\mathbf{Z}_G^H\mathbf{Z}_G + \beta\mathbf{I}\right)^{-1}$. By substituting Eq. (6) into Eq. (9), and by noting that $\mathbf{z}_B^T = (\mathbf{z}_B^*)^H$, we can write $\mathbf{z}_B^T\mathbf{q} = (\mathbf{z}_B^*)^H\tilde{\mathbf{Z}}^{-1}\mathbf{z}_B^*$. Since matrix $\tilde{\mathbf{Z}}^{-1}$ is Hermitian, then it follows that $\mathbf{z}_B^T\mathbf{q}$ is real valued [19]. As a consequence of that, the phase of the reproduced field at the bright point is equal to zero regardless of the values of ψ_D, ψ_G, and β. For a single frequency, $\overline{\psi_D}$ may be calculated by means of an algorithm that performs a full search of $\overline{\psi_D}$ which can be summarized as follows. For each frequency, we start with $\psi_D = 1$. If the quality constraint in Eq. (9) is not satisfied (that is, $\mathbf{z}_B^T\mathbf{q}(\psi_D) < p_{B,\min}$) and $\psi_D > 0$, we iteratively decrease ψ_D by a small decrement δ_ψ until the optimal value $\overline{\psi_D}$ is reached. The update rule of the value of $\overline{\psi_D}$ at the j-th iteration, say $\overline{\psi_D}^{(j)}$, is $\overline{\psi_D}^{(j)} = \overline{\psi_D}^{(j-1)} - \delta_\psi$. A Matlab implementation of the full search of $\overline{\psi_D}$ is available in the code provided in [51]. In addition to ψ_D, system performance also depends on other parameters (e.g., β and ψ_G) whose setting is addressed in the next sections.

B. Choice of the regularization parameter

The regularization factor β, which was introduced to limit the input energy, has significant impact on system performance [4], [8], [9], [20] and various techniques are available for the

2329-9290 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2017.2700945, IEEE/ACM Transactions on Audio, Speech, and Language Processing

calculation of β [8], [9], [20]. In this work, we use the so-called Normalized Tikhonov regularization (NTR) method, wherein the regularization factor is calculated as [9], [21]

$$\beta = \beta_0 \sigma_1^2, \tag{10}$$

where σ_1 is the largest singular value of matrix $\mathbf{Z}$ and β_0 is a positive real-valued factor set by the user [9]. In Appendix A, we discuss the relations between $|p_B|$, $\|\mathbf{q}\|$ and β. In this work, we calculate β for the MDS (see Eq. (8)), we store it in the system, and we use it for all the other scenarios. This reduces the system complexity for the calculation of ψ_D (when the scenario changes) and, hence, for the calculation of the input signals. However, due to this choice, we expect the input energy to vary between different scenarios. An optimal choice for β to satisfy a constraint on a desired level of maximum input energy allowed to the sources [20], [8], [22] within the framework proposed in this work is left to future work. The parameter β_0 may also be used to ensure system robustness against mismatches between assumed $\mathbf{Z}$ and actual responses. More specifically, higher values of β_0 (and, hence, higher values of β) result in higher system robustness [4], [23], [24], [9]. This aspect is important and more details on the calculation of β_0 are reported in the next section.

IV. Experimental validation

The aim of the experiments described in this section is to validate the ability of the system to control the directivity/quality performance trade-off by means of the method proposed in Section III-A. With reference to Fig. 3a, the array was installed in the ISVR Anechoic chamber [25] and placed on top of a turntable. An omni-directional microphone was placed at $R = 3\,\mathrm{m}$ from the geometrical center of the array. Both the loudspeaker array and the microphone were placed at a height of 1.5 m from the floor.

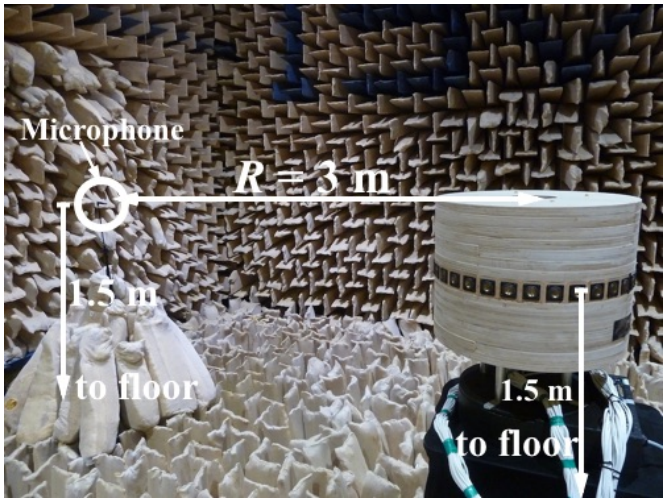
We firstly measured the matrix $\mathbf{H}$ of the TFs between the sources of the circular array prototype and the control points with the aim of calculating the value of β_0 that ensures a good trade-off between directivity performance and system robustness (more details are given in the following paragraphs). The elements of matrix $\mathbf{H}$ were estimated using a system identification method based on the cross-correlation of the signals captured by the microphone (due to the sound generated by the sources of the array) and a white noise signal of 6s duration (that was fed in input to the array) [27], [33]. This procedure was repeated 72 times by rotating the turntable every $\Delta\theta$, $\Delta\theta = 5^\circ$, from 0° to 355°. In order to estimate the value of β_0, we calculated the input signals $\mathbf{q}_{\mathrm{MDS}}$ (see Eq. (8)) for increasing values of β_0 and we evaluated the directivity performance in terms of the acoustic contrast [6] $\mathrm{AC}_{\mathrm{MDS}}$, $\mathrm{AC}_{\mathrm{MDS}} = 10\log_{10}\left(M_D \left\|\mathbf{h}_B^T \mathbf{q}_{\mathrm{MDS}}\right\|^2 \Big/ \|\mathbf{H}_D \mathbf{q}_{\mathrm{MDS}}\|^2\right)$, where $\mathbf{h}_B$ and $\mathbf{H}_D$ are the vector and the matrix of the measured transfer functions between the bright and dark points and the sources of the array, respectively. The $\mathrm{AC}_{\mathrm{MDS}}$ in Fig. 3b shows that, as β_0 increases, the $\mathrm{AC}_{\mathrm{MDS}}$ becomes smoother (as a function of frequency) and this indicates higher system robustness. More specifically, we see that $\beta_0 = 10^{-2}$ provides a higher $\mathrm{AC}_{\mathrm{MDS}}$ at low frequencies than that provided by $\beta_0 = 10^{-1}$. For values of $\beta_0 < 10^{-2}$ the $\mathrm{AC}_{\mathrm{MDS}}$ is more erratic. Due to these reasons, we chose $\beta_0 = 10^{-2}$, and we calculated β using Eq. (10).

After the calculation of β, we measured the response of the input signals designed with the transfer function $\mathbf{Z}$ for each scenario. We set a frequency independent value of $p_{B,\min}$, $p_{B,\min}\,\mathrm{dB} = -3\,\mathrm{dB}$ (for the QCS and the Hybrid Scenarios), and $\delta_\psi = 10^{-3}$ (see Section III-A). The WPMM weight ψ_G in the low-priority zone was chosen as $\psi_G = 10^{-2}$, that is two order of magnitude lower than the maximum value of ψ_G, i.e., $\psi_G = 1$ [9], [12]. The input signals for the five scenarios were calculated by using the full search of ψ_D on a per-frequency basis in the frequency range $[0, f_s/2]\,\mathrm{Hz}$, where $f_s = 48\,\mathrm{kHz}$ is the sampling frequency. The frequency range was divided into $N_{\mathrm{FFT}}/2 + 1$ frequency bins with uniform frequency spacing, where $N_{\mathrm{FFT}} = 8192$. The input signals in the discrete time domain were obtained by inverse Fourier transform of the frequency-domain input signals [28]. A modeling delay of $\Delta t = N_{\mathrm{FFT}}/(2f_s)$ was applied to all the input signals to ensure their causality. Fig. 4a shows the values of $\overline{\psi_D}$ calculated with the full-search of ψ_D for each scenario. As previously stated, due to the choice of β, the input energy varies as the use-case scenario varies. This is shown in Fig. 4b. More specifically, whilst the input energy for the MDS is lower than that of the MQS throughout the frequency range, the input energy of the QCS, HWSS, and HCSS is comparable to that of the MQS. This seems to indicate that the input energy increases as $\overline{\psi_D}$ decreases.

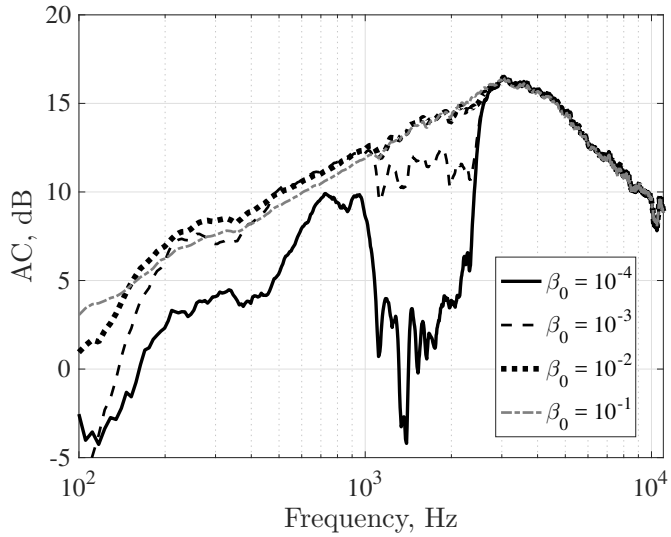
System performance was evaluated in the anechoic chamber by measuring the responses $\mathbf{d}$ of the array driven with the input signals designed for each of the use-case scenarios (see Section III). The measurement setup and procedure is almost identical to that used for the measurement of the TFs $\mathbf{H}$. Fig. 5 shows the magnitude of the Frequency Response Function (FRF) of the reproduced field d_B measured at the bright point due to input signals designed for the QCS, HCSS, HWSS and MDS. All the FRF were normalized by the measured response at the bright point d_B^{MQS} of the input signals designed for the MQS, which is taken as a reference. Whilst the MDS shows the problem of the energy loss at low frequencies [10], the quality constraint set in the other use-case scenarios limits the energy loss. At higher frequencies, all the input signals have similar performance. The input signals designed for the QCS, HCSS, HWSS show a reduction of the SPL that is almost equal to the selected $p_{B,\min}\,\mathrm{dB} = -3\,\mathrm{dB}$ (see Fig. 5) hence indicating that the performance quality constraint is satisfied. However, the results of the HWSS could be potentially improved by reducing the value of the regularization factor at low frequencies. This, however, may reduce the robustness of the system against errors such as inaccuracies in the positions of the drivers [4] or reverberation [40]. Fig. 6 shows the measured acoustic contrast AC for the QCS, MQS, and MDS, where AC is defined as $\mathrm{AC} = 10\log_{10}\left(M_D |d_B|^2 \Big/ \|\mathbf{d}_D\|^2\right)$ [6], and $\mathbf{d}_D$ is the vector of sound pressures measured in the anechoic chamber at the dark points. As expected, at frequencies up to 1 kHz, the input signals for the QCS scenario provide an AC that is between the MQS (the lowest) and the MDS (the highest). Above 1kHz, the AC of the QCS and MDS are identical. However,

2329-9290 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2017.2700945, IEEE/ACM Transactions on Audio, Speech, and Language Processing

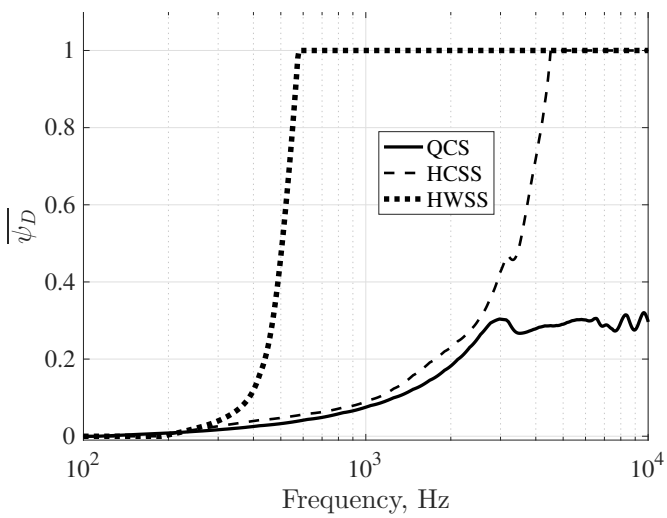


(a) Details of the experiments.

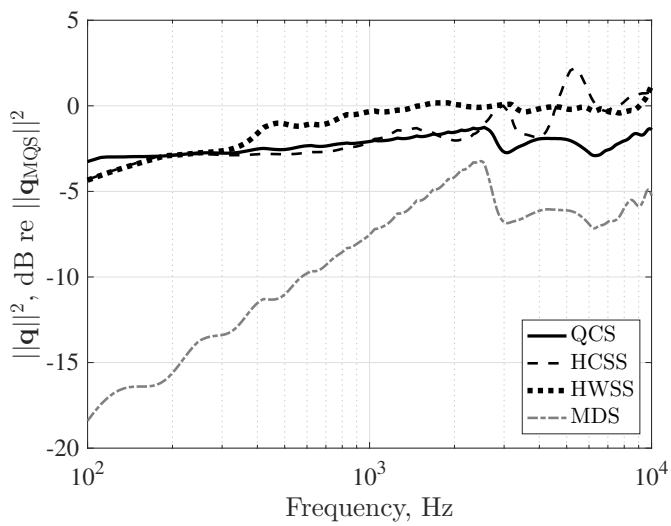


(b) Acoustic contrast for the MDS as a function of β_0.

Fig. 3. (Color online) Details of the experiments (Fig. 3a) showing the circular array prototype and the microphone installed in the ISVR Anechoic Chamber. Calculation of β_0 (Fig. 3b) through the prediction of the Acoustic Contrast of the input signals designed for the MDS scenario.



(a) $\overline{\psi_D}$



(b) Input energy

Fig. 4. Values of $\overline{\psi_D}$ calculated by means of the full-search of ψ_D (see Section III-A) for three use-case scenarios (Fig. 4a). Energy (Fig. 4b) of the input signals designed for the use-case scenarios.

the highest contrast of the MDS corresponds to lower quality performance than in the other scenarios (see Fig. 5). In this case, the advantage of using the proposed method over the MDS is that we can control the SPL at the listening point to reflect the constraint set in Eq. (9). The directivity performance of the HCSS and HWSS are evaluated by inspection of the sound field plots in Fig. 7a and Fig. 7b, respectively. As previously stated, in the case of the Hybrid Scenario, performance depends on the relative position of the bright and the dark points. In fact, in the HWSS (Fig. 7b) the system provides higher attenuation of the SPL in the dark zone at lower frequencies than that provided for the HCSS (Fig. 7a).

V. Methods for the Update of the Input Signals

In the previous Sections, we have used ψ_D to control the directivity/quality trade-off and the input signals were calculated by means of a full search of ψ_D (see Section III-A), which may be time-consuming (e.g., as a function of δ_ψ) and thus not suitable for practical applications. In this section, we devise two methods for the estimation of the optimal ψ_D once the user sets a new value of $p_{B,\min}$ and the system needs to update the input signals so that they satisfy Eq. (9). The results of the full search of ψ_D are used as a reference to compare the accuracy of the ψ_D estimated by the two methods proposed in the following subsections.

2329-9290 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2017.2700945, IEEE/ACM Transactions on Audio, Speech, and Language Processing

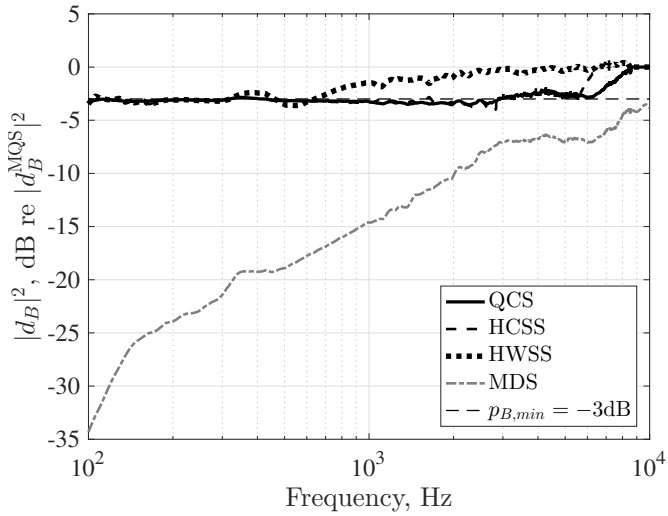


Fig. 5. The magnitude responses of the sound pressure measured at the bright point for each use-case scenario.

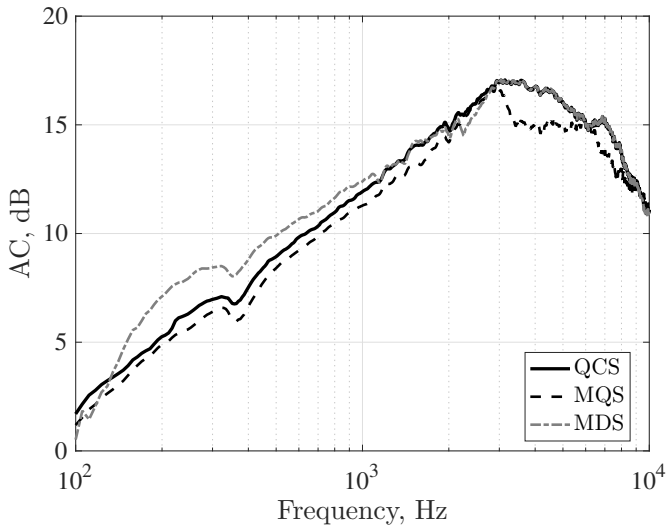


Fig. 6. Measured Acoustic Contrast provided by input signals designed for some of the use-case scenarios.

A. Method based on Neumann series

Let us consider a given scenario and a set of input signals $\mathbf{q}(\widehat{\psi_D})$ calculated with Eq. (6) by setting $\widehat{\psi_D} = 1/2$. At this point, the listener sets a desired directivity/quality trade-off (i.e., by setting a value for $p_{B,\min}$) but the input signals $\mathbf{q}(\widehat{\psi_D})$ may not satisfy the performance constraint on $p_{B,\min}$. If that is the case, then we need to find a new set of input signals $\tilde{\mathbf{q}} = \mathbf{q}(\widehat{\psi_D} + \Delta\psi_D)$ that satisfies the performance constraint, where

$$-0.5 \leq \Delta\psi_D \leq 0.5 \tag{11}$$

is the value that we need to add to $\widehat{\psi_D}$ in order to ensure that $\mathbf{z}_B^T\tilde{\mathbf{q}} \geq p_{B,\min}$ (see Eq. (9)). Let us define $\mathbf{A} = \mathbf{z}_B^*\mathbf{z}_B^T + \widehat{\psi_D}\mathbf{Z}_D^H\mathbf{Z}_D + \psi_G\mathbf{Z}_G^H\mathbf{Z}_G + \beta\mathbf{I}$, so that

$$\mathbf{q}(\widehat{\psi_D}) = \mathbf{A}^{-1}\mathbf{z}_B^*, \tag{12}$$

and

$$\tilde{\mathbf{q}} = \mathbf{q}(\widehat{\psi_D} + \Delta\psi_D) = \left(\mathbf{A} + \Delta\psi_D\mathbf{Z}_D^H\mathbf{Z}_D\right)^{-1}\mathbf{z}_B^*. \tag{13}$$

If the entries of matrix $\Delta\psi_D\mathbf{Z}_D^H\mathbf{Z}_D$ are small enough (compared to those of $\mathbf{A}$) and the following condition is satisfied [19]

$$\lim_{n\to\infty}\left(-\Delta\psi_D\mathbf{A}^{-1}\mathbf{Z}_D^H\mathbf{Z}_D\right)^n = 0\mathbf{I}, \tag{14}$$

then we can write [19], [29]

$$\left(\mathbf{A} + \Delta\psi_D\mathbf{Z}_D^H\mathbf{Z}_D\right)^{-1} = \sum_{n=0}^{\infty}\left(-\Delta\psi_D\mathbf{A}^{-1}\mathbf{Z}_D^H\mathbf{Z}_D\right)^n\mathbf{A}^{-1}, \tag{15}$$

that is also known in the literature as a Neumann Series expansion [19], [29]. The condition in Eq. (14) is true if and only if [30]

$$\rho\left(-\Delta\psi_D\mathbf{A}^{-1}\mathbf{Z}_D^H\mathbf{Z}_D\right) < 1, \tag{16}$$

where $\rho\left(-\Delta\psi_D\mathbf{A}^{-1}\mathbf{Z}_D^H\mathbf{Z}_D\right)$ indicates the spectral radius of matrix $-\Delta\psi_D\mathbf{A}^{-1}\mathbf{Z}_D^H\mathbf{Z}_D$ [19]. The condition in Eq. (16) depends on multiple factors (e.g., $\mathbf{Z}$, ψ_D, ψ_G, and β). In Appendix B we show that

$$\rho\left(-\Delta\psi_D\mathbf{A}^{-1}\mathbf{Z}_D^H\mathbf{Z}_D\right) \leq |\Delta\psi_D|\left\|\mathbf{Z}_D^H\mathbf{Z}_D\right\| / \left(\lambda_L^{(0)} + \beta\right), \tag{17}$$

and $\lambda_L^{(0)}$ is the smallest singular value of $\mathbf{Z}^H\hat{\mathbf{\Psi}}_0^2\mathbf{Z}$, where the diagonal matrix $\hat{\mathbf{\Psi}}_0$ has elements $\left[1, \sqrt{\widehat{\psi_D}}\mathbf{1}_{M_D}, \sqrt{\psi_G}\mathbf{1}_{M_G}\right]$ on its main diagonal (see Appendix B). This suggests that there is an upper bound to $\rho\left(-\Delta\psi_D\mathbf{A}^{-1}\mathbf{Z}_D^H\mathbf{Z}_D\right)$ and this is controlled by β and $\Delta\psi_D$, among other factors. Therefore, for a given $\Delta\psi_D$, it is always possible to select a value β so that Eq. (16) is satisfied.

We assume that Eq. (16) holds, which is the case in all the use-cases under analysis (QCS, HWSS, and HCSS) as shown in Fig. 8a. By substituting Eq. (15) into Eq. (13), and by using Eq. (12), we can write

$$\tilde{\mathbf{q}} = \sum_{n=0}^{\infty}\Delta\psi_D^n(-1)^n\left(\mathbf{A}^{-1}\mathbf{Z}_D^H\mathbf{Z}_D\right)^n\mathbf{q}(\widehat{\psi_D}). \tag{18}$$

The equation above suggests that the updated set of input signals $\tilde{\mathbf{q}} = \mathbf{q}(\widehat{\psi_D} + \Delta\psi_D)$ can be calculated using the reference set $\mathbf{q}(\widehat{\psi_D})$ and, most noticeably, no more matrix inversions are required for the calculation of $\tilde{\mathbf{q}}$. In fact, $\mathbf{A}^{-1}$ and $\mathbf{Z}_D^H\mathbf{Z}_D$ are computed for $\mathbf{q}(\widehat{\psi_D})$ and then stored in the system.

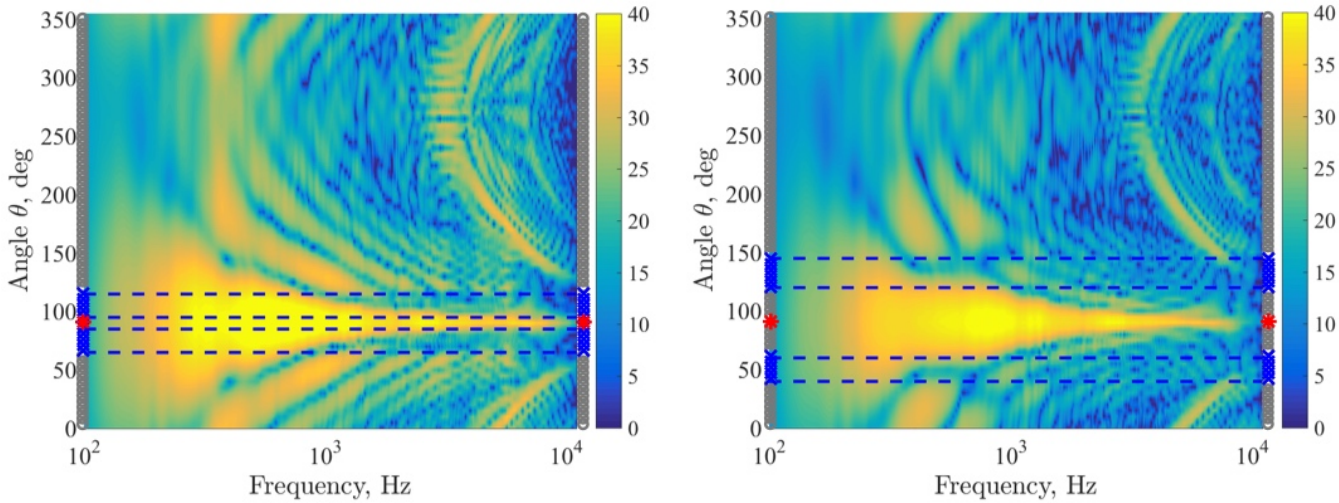
The Neumann series in Eq. (18) consists of an infinite series of terms. Since the terms $\Delta\psi_D^n$ and $\left\|\left(\mathbf{A}^{-1}\mathbf{Z}_D^H\mathbf{Z}_D\right)^n\right\|$ decrease monotonically as N increases, it is reasonable to approximate the series in Eq. (15) by truncating it to an order N, thus yielding the approximated input signals $\tilde{\mathbf{q}}_N$ that are defined as

$$\tilde{\mathbf{q}}_N = \sum_{n=0}^{N}\Delta\psi_D^n(-1)^n\left(\mathbf{A}^{-1}\mathbf{Z}_D^H\mathbf{Z}_D\right)^n\mathbf{q}(\widehat{\psi_D}). \tag{19}$$

Let us assume that N is large enough so that $\tilde{\mathbf{q}}_N$ approximates $\tilde{\mathbf{q}}$ well, that is $\tilde{\mathbf{q}}_N \approx \tilde{\mathbf{q}}$. In principle, if $\tilde{\mathbf{q}}$ satisfies the quality constraint in Eq. (9), and if $\tilde{\mathbf{q}}_N$ well approximates $\tilde{\mathbf{q}}$, then we may infer that $\tilde{\mathbf{q}}_N$ may (approximately) satisfy the quality constraint. The objective is now to determine the value of $\Delta\psi_D$ (that appears as a multiplication factor in Eq. (19)) such that $\tilde{\mathbf{q}}_N$ satisfies the constraint in Eq. (9). By multiplying both

2329-9290 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2017.2700945, IEEE/ACM Transactions on Audio, Speech, and Language Processing



(a) HCSS. (b) HWSS.

Fig. 7. (Color online) Responses of the soundfield measured in anechoic environment due to the input signals designed for the HCSS (Fig. 7a) and the HWSS (Fig. 7b). The color scale is in dB re 1V. The dashed blue linear indicate the limits of the dark zones for each use-case scenario. The distribution of the control points follows the same specifications as in the legend in Fig. 2.

members of Eq. (19) by $\mathbf{z}_B^T$ and by setting

$$\tilde{p}_{B,n} = (-1)^n \mathbf{z}_B^T \left(\mathbf{A}^{-1}\mathbf{Z}_D^H\mathbf{Z}_D\right)^n \mathbf{q}(\widehat{\psi_D}), \tag{20}$$

we can write the quality constraint as

$$\mathbf{z}_B^T\tilde{\mathbf{q}}_N = \sum_{n=0}^{N} \Delta\psi_D{}^n \tilde{p}_{B,n} \leq p_{B,\min}, \tag{21}$$

where we recall that $-0.5 \leq \Delta\psi_D \leq 0.5$ (see Eq. (11)). In Appendix C we show that the terms $\tilde{p}_{B,n}$ are real-valued. Hence, Eq. (21) is a real-valued polynomial of degree N in $\Delta\psi_D$. If that is the case, we are able to estimate the $\overline{\Delta\psi_D}$ that satisfies the constraint on $p_{B,\min}$ (as an equality) by solving

$$\overline{\Delta\psi_D} = \Delta\psi_D \in \mathbb{R} : \sum_{n=0}^{N} \tilde{p}_{B,n}\Delta\psi_D{}^n - p_{B,\min} = 0, \tag{22}$$

where $\mathbb{R}$ denotes the set of the real numbers. If multiple real roots exist, $\overline{\Delta\psi_D}$ is chosen as the highest among the roots that satisfy Eq. (11). The final ψ_D is calculated as

$$\psi_D = \begin{cases} 1, \text{ if } \widehat{\psi_D} + \overline{\Delta\psi_D} > 1, \\ 0, \text{ if } \widehat{\psi_D} + \overline{\Delta\psi_D} < 0, \\ \widehat{\psi_D} + \overline{\Delta\psi_D}, \text{ otherwise.} \end{cases} \tag{23}$$

The algorithm for the update of the input signals is summarized in Alg. 1.

Algorithm 1 Algorithm for the update of the input signal (based on the Neumann series).

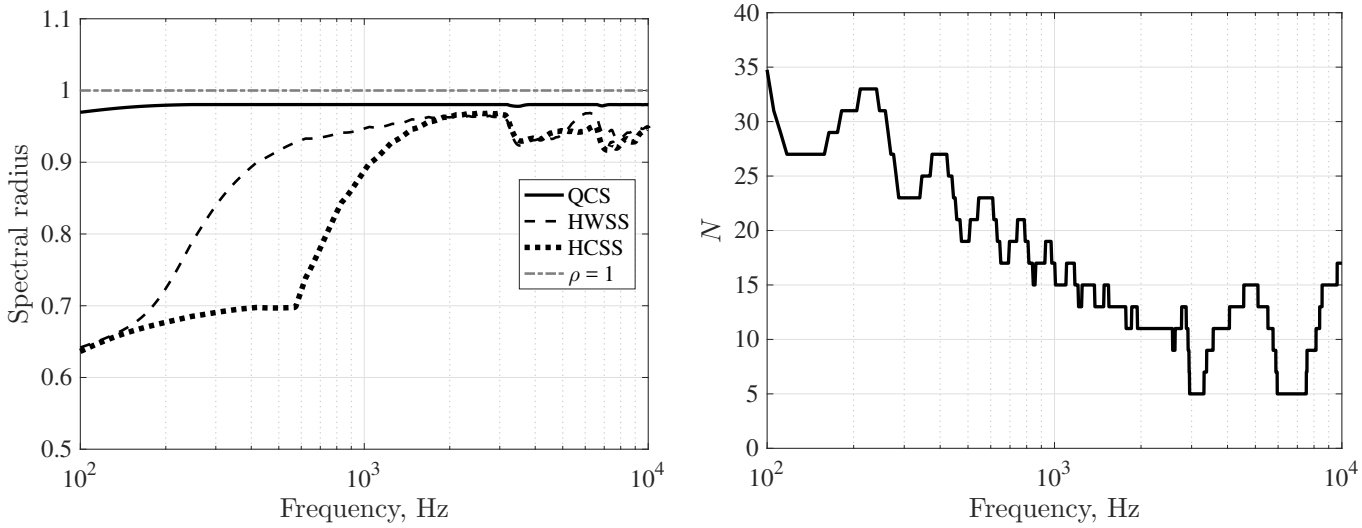
INPUTS: $p_{B,\min}$.
STORED IN THE SYSTEM: $\tilde{p}_{B,n}$, $\mathbf{A}$, $\mathbf{Z}_D$, N, $\mathbf{q}(\widehat{\psi_D})$.
ASSUMPTION: Eq. (16) is satisfied for $\Delta\psi_D = 0.5$.
#1) Calculate ψ_D with Eq. (23) and Eq. (22).
#2) Calculate $\tilde{\mathbf{q}}$ (or $\tilde{\mathbf{q}}_N$) with ψ_D and Eq. (6) (or Eq. (19)).
OUTPUTS: $\tilde{\mathbf{q}}$ (or $\tilde{\mathbf{q}}_N$), ψ_D.

By truncating the series to a given order N, we are introducing errors between the $\tilde{\mathbf{q}}$ (calculated with the full-search of ψ_D and considered as "true" values) and $\tilde{\mathbf{q}}_N$ (calculated with the Neumann series in Eq. (19)). These errors depend on N, as well as on the values of $\Delta\psi_D$ and on frequency. The error $\varepsilon(N)$ between the two sets of input signals is defined as $\varepsilon(N) = 10\log_{10}\left(\|\tilde{\mathbf{q}}_N - \tilde{\mathbf{q}}\|^2 \big/ \|\tilde{\mathbf{q}}\|^2\right)$ dB. If, on the one hand, it may be possible to accurately approximate the input signals $\tilde{\mathbf{q}}$ by setting high values of N, on the other hand it may be beneficial to define it as a frequency-dependent parameter in order to reduce the computational load. More specifically, the chosen N decreases as frequency increases. By considering the MDS configuration of control points, we calculate the exact input signals $\tilde{\mathbf{q}}$ with $\psi_D = \widehat{\psi_D} + \Delta\psi_D$ and $\Delta\psi_D = 0.5$ using Eq. (6). Instead, the approximated input signals are calculated with Eq. (19) and $\Delta\psi_D = 0.5$. The selected value of N (for a given frequency) is $N = \min_N\{\varepsilon(N) \leq \varepsilon_{\text{MAX}}\}$, where ε_{MAX} is an error threshold (in dB) set by the user, typically to a very low value, that we set to $\varepsilon_{\text{MAX}} = -80$ dB. Furthermore, given that the coefficients $\tilde{p}_{B,n}$ of the polynomial are real valued, N can be chosen to be odd to ensure the existence of at least one real valued root in the polynomial in Eq. (22) (note that if $\mathbf{z}_B^T\mathbf{q}(\widehat{\psi_D}) - p_{B,\min} < 0$, the value of the polynomial for $\Delta\psi_D = 0$ is negative). The algorithm for the calculation of N can be summarized as follows. For each frequency, we start with $N = 0$. At each iteration, if $\varepsilon(N) > \varepsilon_{\text{MAX}}$, we perform $N = N + 1$ until $\varepsilon(N) \leq \varepsilon_{\text{MAX}}$. If the calculated value of N is even, we then increase N by 1 and we proceed to the next frequency. By following a similar reasoning as that adopted for β (see Section III-B), the values of N calculated with the above-mentioned method and shown in Fig. 8b are stored in the system and used for all the various scenarios.

As an example, Fig. 9a shows the polynomial in Eq. (22) as a function of $\overline{\psi_D}$ at four frequencies for the HCSS problem (see Fig. 2b). The intersection of each polynomial (at each frequency) with the dashed horizontal line in the range $-0.5 \leq$

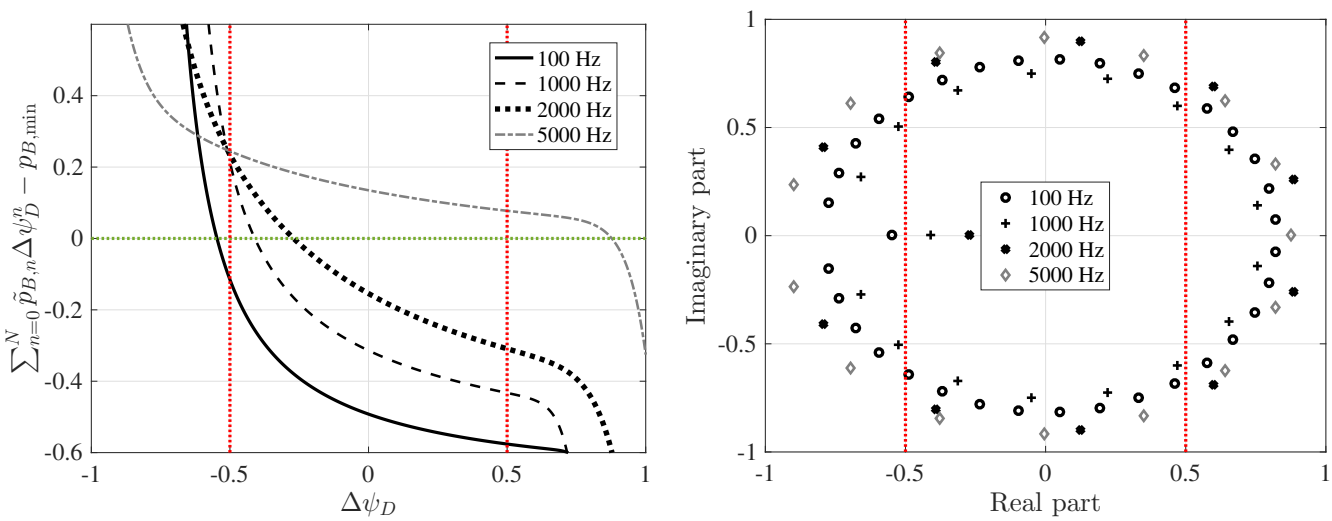
2329-9290 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2017.2700945, IEEE/ACM Transactions on Audio, Speech, and Language Processing



(a) Spectral radius. (b) N.

Fig. 8. Spectral radius of matrix $-\Delta\psi_D \mathbf{A}^{-1}\mathbf{Z}_D^H\mathbf{Z}_D$ (for $\Delta\psi_D = 0.5$) for some use-case scenarios (Fig. 8a). Values of N calculated with the method introduced in Section V-A and $\varepsilon_{\text{MAX}} = -80$ dB (Fig. 8b).



(a) Polynomial. (b) Roots.

Fig. 9. Plots of the values (Fig. 9a) and of the roots (Fig. 9b) of the polynomial in Eq. (22) at four frequencies for the HCSS (whose control problem is shown in Fig. 2b). The selected values of $\Delta\psi_D$ in Fig. 9a and Fig. 9b are the ones that intersect the x-axis. The number N of the roots in Fig. 9b is shown in Fig. 8b. The vertical dashed lines represent the domain of $\Delta\psi_D$.

$\overline{\Delta\psi_D} \leq 0.5$ is the selected value $\overline{\Delta\psi_D}$. This value corresponds to the real root of the polynomial shown in Fig. 9b. At 100 Hz, $\overline{\Delta\psi_D} < -0.5$ (see Eq. (11)) and hence, the solution $\overline{\Delta\psi_D}$ is set to $\overline{\Delta\psi_D} = -0.5$ (see Eq. (23)). As frequency increases (e.g., at 1 kHz and 2 kHz in Fig. 9b), the solutions $\overline{\Delta\psi_D}$ are in the domain of $\Delta\psi_D$. At higher frequencies (e.g., at 5 kHz in Fig. 9b), the solution $\overline{\Delta\psi_D} > 0.5$ so we set $\overline{\Delta\psi_D} = 0.5$.

The error ε_{ψ_D} between the values of ψ_D calculated with the full-search (that are taken as a reference) and those calculated with Eq. (23) is defined as $\varepsilon_{\psi_D} = 10\log_{10}\left(\left|(\widehat{\psi_D} + \overline{\Delta\psi_D}) - \overline{\psi_D}\right|^2 / |\overline{\psi_D}|^2\right)$ dB. The values of ε_{ψ_D} in dB as a function of frequency for each scenario are shown in Fig. 10. Whilst the errors for the HCSS and HWSS are in the range between -30dB and -70dB, that of the QCS is high at very low frequencies (from 100 Hz to 300 Hz). This corresponds to small differences between the values of $\widehat{\psi_D} + \overline{\Delta\psi_D}$ and $\overline{\psi_D}$ (as shown in the subplot in Fig. 10), where it is observed that Alg. 1 underestimates the "true" values of ψ_D (i.e., $\overline{\psi_D}$). In Fig. 11, we show the error $\varepsilon_{\mathbf{d}}$, $\varepsilon_{\mathbf{d}} = 10\log_{10}\left(\left\|\mathbf{H}\left(\mathbf{q}(\widehat{\psi_D} + \overline{\Delta\psi_D}) - \mathbf{q}(\overline{\psi_D})\right)\right\|^2 / \|\mathbf{H}\mathbf{q}(\overline{\psi_D})\|^2\right)$ dB, between the predicted radiation patterns of the input signals calculated with $\overline{\psi_D}$ and $\widehat{\psi_D} + \overline{\Delta\psi_D}$. The differences between $\widehat{\psi_D} + \overline{\Delta\psi_D}$ and $\overline{\psi_D}$ have no impact on the HCSS and HWSS throughout the whole frequency range and on the QCS at frequencies above 1 kHz. At frequencies below 1 kHz, the $\varepsilon_{\mathbf{d}}$ for the QCS decreases as frequency increases and the maximum value is about -18 dB (at around 150 Hz).

2329-9290 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2017.2700945, IEEE/ACM Transactions on Audio, Speech, and Language Processing

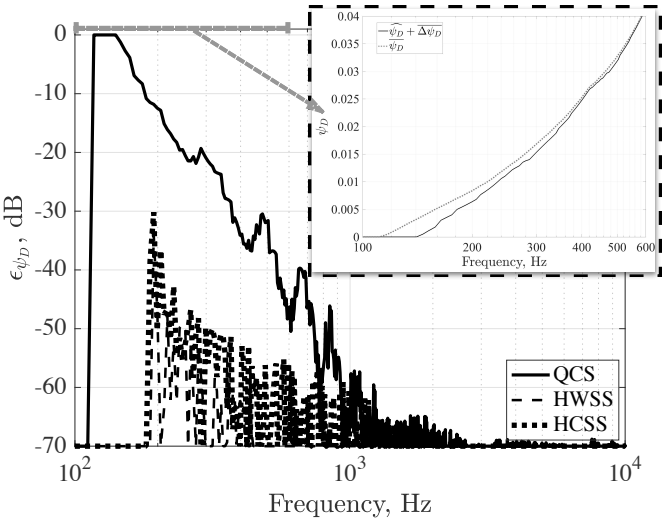


Fig. 10. Error between optimal ψ_D calculated with the full-search of ψ_D and Alg. 1. The subplot shows the values of ψ_D calculated for the QCS with the full-search and Alg. 1 between 100 Hz and 600 Hz.

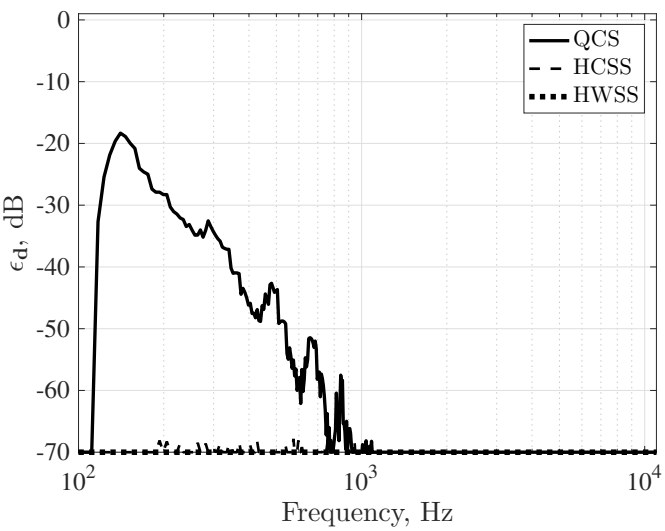


Fig. 11. Error $\varepsilon_{\mathbf{d}}$ between the radiation patterns (predicted using the measured transfer functions $\mathbf{H}$) generated by the input signals calculated with $\overline{\psi_D}$ and $\widehat{\psi_D} + \overline{\Delta\psi_D}$.

B. Iterative method

The estimation of the optimal ψ_D can be also performed with the pseudo-code in Alg. 2. This algorithm, which uses a bisection-like method for the estimation of the optimal ψ_D, is significantly faster than the full-search of ψ_D and, in our implementation [51], it returns identical results as that provided by the full-search up to the 5-th decimal digit of ψ_D (if the threshold γ_ψ is set to $\gamma_\psi = 10^{-6}$).

C. Computational advantages of Alg. 1 versus Alg. 2

Once the user sets a new value for $p_{B,\min}$, for a single frequency, Alg. 2 requires, at each step, a matrix inversion and a matrix-vector multiplication. The operation of matrix inversion can be performed in $\mathscr{O}(L^3)$ operations [35]. For a single frequency, the time it takes for Alg. 2 to calculate the updated input signals is proportional to the number of iterations times the time required to perform the matrix inversion.

Algorithm 2 Algorithm for the calculation of $\overline{\psi_D}$ (bisection-like method).

INPUTS: $\mathbf{Z}$, $\hat{\mathbf{p}}$, $p_{B,\min}$, β, ψ_G, γ_ψ, $\psi_{\text{MAX}} = 1$, and $\psi_{\text{MIN}} = 0$.
1) $\psi_D = (\psi_{\text{MAX}} - \psi_{\text{MIN}})/2 + \psi_{\text{MIN}}$.
2) Calculate $\mathbf{q}(\psi_D)$ and $p_B = \mathbf{z}_B^T \mathbf{q}(\psi_D)$.
if $\left|p_B - p_{B,\min}\right| \leq \gamma_\psi$ OR $\psi_{\text{MAX}} \leq \gamma_\psi$ **then** 3) Return ψ_D.
else
  if $p_B < p_{B,\min}$ **then** $\psi_{\text{MAX}} = \psi_D$ **else** $\psi_{\text{MIN}} = \psi_D$
  end if
  Go to Step 1).
end if
OUTPUTS: ψ_D, $\mathbf{q}$.

Instead, Alg. 1 requires the calculation of the roots of a polynomial of order N (Step #1 of Alg. 1) and then the update of the input signals (Step #2 of Alg. 1). Step #1 of Alg. 1 can be implemented by calculating the eigenvalues of the companion matrix of the polynomial [34], which is, for example, the implementation provided in Matlab [49]. This operation can be performed in $\mathscr{O}(N^2)$ operations [34], [36], [37]. Whilst L does not depend on frequency and hence the computational complexity of Alg. 2 is frequency-independent, the complexity of Alg. 1 decreases as frequency increases. This is due to the choice of N (see Section V-A and Fig. 8b). Step #2 of Alg. 1 can be implemented 1) by means of a matrix inversion (Eq. (6)) with $\mathscr{O}(L^3)$ operations [35] or, 2) by summing the terms $\left(\mathbf{A}^{-1}\mathbf{Z}_D^H\mathbf{Z}_D\right)^n$ weighted by $\overline{\Delta\psi_D}$ Eq. (19)) with a time complexity of $\mathscr{O}(L^2)$ (if the terms $\left(\mathbf{A}^{-1}\mathbf{Z}_D^H\mathbf{Z}_D\right)^n$ are calculated offline and stored in the system). We consider option one (the matrix inversion) for Step #2 of Alg. 1. By using the Matlab implementations of Alg. 1 and Alg. 2 provided in [51], we compare the time taken by each algorithm (averaged over 1000 repetitions) to compute the input signals for the QCS on a personal computer (2.6 GHz Intel Core i5 and 16 GB RAM) with Matlab 2016b. The average time required by Alg. 1 is 1.49s whilst that of Alg. 2 is 5.92s.

To summarize, the complexities of a single iteration of Alg. 2 (that is proportional to $\mathscr{O}(L^3)$) and that of Alg. 1 (Step #2, which is $\mathscr{O}(L^3)$) are comparable. However, in general, the iterative algorithms may require a number of iterations to complete, whilst Alg. 1 does not. Hence Alg. 1 is likely to be faster than the iterative algorithms (e.g., the full-search and Alg. 2). However, Alg. 1 requires more steps than Alg. 2 in the setup phase (e.g., the calculation of $\mathbf{q}(\widehat{\psi_D})$ in Step #1 of Alg. 1). These steps have to be performed once (for a given scenario) and do not depend on the value of $p_{B,\min}$.

VI. Possible strategies for the reverberant environment

In a practical scenario, the proposed system is likely to be used in a reverberant environment. Previous works have investigated the design of systems for sound field control in reverberant environments [38], [39] and the negative effects of reverberation on system performance [40], [43], [47]. A number

2329-9290 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2017.2700945, IEEE/ACM Transactions on Audio, Speech, and Language Processing

of strategies have been devised to mitigate such negative effects by means, for example, of higher order sources [41], [42] or by exploiting first order reflections [45], [46], [41], [44]. In case of diffuse sound fields, Simon-Galvez *et al.* [39], [40] have highlighted that the level of the input sources plays a crucial role in system performance. Overall, previous works emphasize the importance of the problems that arise due to the effects of reverberation. For example, as reported in the work of Simon-Galvez *et al.* [40], low values of the regularization factor may be used to preserve the directivity performance of the system in a reverberant environment. However, this may lead to large values of input energy to the array, thus making the system difficult to be implemented in practice.

In this section, we provide suggestions for using the proposed system in a reverberant environment that are based on previous works. Due to the complex nature of the problem, a thorough study of the performance of the proposed methods in a reverberant environment requires further work and experimental validation. This may be the object of further investigation.

Two options are considered for using the system in a reverberant environment. 1) To measure the transfer functions in the reverberant environment [40], [46], and then use that instead of $\mathbf{Z}$ for the calculation of the input signals with the WPMM. In this case, the proposed algorithms for the control of the performance trade-off would still apply. However, this strategy may lead to audio quality issues in the listening zone [47] and it may be less effective at higher frequencies. 2) If the environment is characterized by strong early reflections, one option is to reduce the radiated sound towards the reflectors. This can be done by imposing a dark region in the directions of the reflectors, a strategy which was previously proposed in the context of cross-talk cancellation as shown in reference [48]. 3) Another option is to increase the overall directivity of the array (e.g., by increasing the value of ψ_D), thus decreasing the quality/directivity ratio.

VII. Conclusions

In this work, we have presented an implementation of a system for private sound with a circular array and the WPMM. The proposed system accounts for different use-case scenarios and the user is allowed to control the quality/directivity performance trade-off (when the input energy is limited) by setting the desired characteristics of the frequency response of the sound field in the listening zone. The control of the trade-off is then performed by setting the WPMM weight in the dark zone so that the input signals satisfy the desired performance constraint. The results of the experiments with a circular array prototype have shown that this is effective in a wide frequency range. Three algorithms were proposed for the calculation of the WPMM weight in the dark zone that are based on 1) iterative search of the optimal value of the WPMM weight in the dark zone (Alg. 2) and 2) on an approximation of the inverse matrix (by means of the Neumann series) involved in the calculation of the input signals (Alg. 1). The main idea behind Alg. 1 is to use a given set of input signals (stored in the system) and update it. The results of the validation Alg. 1 have shown that the accuracy of the filter update depends on frequency and on the use case considered. In general, it is very accurate at high frequencies and less accurate at low frequencies. The results of the simulations have shown that Alg. 1 is faster than the iterative algorithm (Alg. 2). Further research could aim at improving performance at low-frequencies by means of a joint optimal tuning of the regularization factor and the WPMM weight in the dark zone.

References

[1] W. Druyvesteyn and J. Garas, "Personal sound," *J. Audio Eng. Soc.*, vol. 45, no. 9, pp. 685–701, 1997.

[2] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," Proc. IEEE, vol. 57, no. 8, pp. 14081418, 1969.

[3] M. Tanter, J. L. Thomas, and M. Fink, "Time reversal and the inverse filter.," J. Acoust. Soc. Am., vol. 108, no. 1, pp. 223234, Jul. 2000.

[4] S. J. Elliott, J. Cheer, J. W. Choi, and Y. Kim, "Robustness and Regularization of Personal Audio Systems," *IEEE Trans. Audio. Speech. Lang. Processing*, vol. 20, no. 7, pp. 2123–2133, sep 2012.

[5] T. Betlehem, W. Zhang, M. A. Poletti, and T. D. Abhayapala, "Personal Sound Zones: Delivering interface-free audio to multiple listeners," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 81–91, mar 2015.

[6] J. W. Choi and Y. H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," *J. Acoust. Soc. Am.*, vol. 111, no. 4, p. 1695, 2002.

[7] M. Shin, S. Q. Lee, F. M. Fazi, P. A. Nelson, D. Kim, S. Wang, K. H. Park, and J. Seo, "Maximization of acoustic energy difference between two spaces," *J. Acoust. Soc. Am.*, vol. 128, no. 1, pp. 121–31, jul 2010.

[8] S. Galvez, F. Marcos, S. J. Elliott, J. Cheer, "A superdirective array of phase shift sources," *J. Acoust. Soc. Am.*, vol. 132, no. May 2013.

[9] M. Shin, F. M. Fazi, P. A. Nelson, and F. C. Hirono, "Controlled sound field with a dual layer loudspeaker array," *J. Sound Vib.*, vol. 333, no. 16, pp. 3794–3817, Aug 2014.

[10] F. M. Fazi, M. Shin, F. Olivieri, S. Fontana, "Low frequency performance of circular loudspeaker arrays," in *Audio Eng. Soc. Conv. 138*, May 2015, pp. 1–4.

[11] F. Olivieri, F. Fazi, M. Shin, and P. Nelson, "Pressure-Matching beamforming method for loudspeaker arrays with frequency dependent selection of control points," in *AES 138th Conv.*, Warsaw, Poland, 2015.

[12] F. Olivieri, F. M. Fazi, P. A. Nelson, and S. Fontana, "Comparison of Strategies for Accurate Reproduction of a target field with Compact Arrays of Loudspeakers for the Generation of Zones of Private Sound and Silence," *J. Audio Eng. Soc.*, vol. 64, no. 11, pp. 905917, 2016.

[13] J. H. Chang and F. Jacobsen, "Sound field control with a circular double-layer array of loudspeakers," *J. Acoust. Soc. Am.*, vol. 131, no. 6, p. 4518, jun 2012.

[14] T. Betlehem and P. D. Teal, "A constrained optimization approach for multi-zone surround sound," in *2011 IEEE Int. Conf. Acoust. Speech Signal Process.*, vol. 1. May 2011, pp. 437–440.

[15] Y. Cai, M. Wu, and J. Yang, "Sound reproduction in personal audio systems using the least-squares approach with acoustic contrast control constraint," *J. Acoust. Soc. Am.*, vol. 135, no. 2, pp. 734–741, Feb 2014.

[16] F. M. Fazi, M. Shin, F. Olivieri, S. Fontana, Y. Lang, and L. Yue, "Comparison of Pressure-Matching and Mode-Matching Beamforming for Methods for Circular Loudspeaker Arrays," in *Audio Eng. Soc. Conv. 137*, Oct 2014, pp. 1–14.

[17] O. Kirkeby and P. A. Nelson, "Reproduction of plane wave sound fields," *J. Acoust. Soc. Am.*, vol. 94, no. November 1993, pp. 2992–3000, 1993.

[18] M. R. Bai, J.-C. Wen, H. Hsu, Y.-H. Hua, and Y.-H. Hsieh, "Investigation on the reproduction performance versus acoustic contrast control in sound field synthesis." *J. Acoust. Soc. Am.*, vol. 136, no. 4, pp. 1591–600, Oct 2014.

[19] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*. Society for Industrial and Applied Mathematics (SIAM), 2000.

[20] T. Betlehem and C. Withers, "Sound Field Reproduction With Energy Constraint on Loudspeaker Weights," *IEEE Trans. Audio. Speech. Lang. Processing*, vol. 20, no. 8, pp. 2388–2392, oct 2012.

[21] P.-A. Gauthier, C. Camier, O. Gauthier, Y. Pasco, and A. Berry, "Sound field reproduction of real flight recordings in cabin mock-up," in *AES 52th Int. Conf.*, 2013, pp. 1–10.

[22] F. Olivieri, F. M. Fazi, P. A. Nelson, M. Shin, S. Fontana, and L. Yue, "Theoretical and experimental comparative analysis of beamforming methods for loudspeaker arrays under given performance constraints," *Journal of Sound and Vibration*, vol. 373, pp. 302324, Jul. 2016.

2329-9290 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2017.2700945, IEEE/ACM Transactions on Audio, Speech, and Language Processing

[23] P. Coleman, P. J. Jackson, M. Olik, M. Olsen, M. Møller, and J. Abildgaard Pedersen, "The influence of regularization on anechoic performance and robustness of sound zone methods," in *Proc. Meet. Acoust.*, vol. 19, no. 1, 2013.

[24] P. Coleman, P. J. B. Jackson, M. Olik, M. Møller, M. Olsen, and J. Abildgaard Pedersen, "Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array," *J. Acoust. Soc. Am.*, vol. 135, no. 4, pp. 1929–1940, 2014.

[25] ISVR, "ISVR Anechoic Chamber Web site," 2012. [Online] Available: http://www.isvr.co.uk/faciliti/lg_anech.htm

[26] G. W. Elko, "Superdirectional microphone arrays," in Acoustic Signal Processing for Telecommunication, S. L. Gay and J. Benesty, Eds., Kluwer Academic, 2000.

[27] K. Shin and J. Hammond, *Fundamentals of signal processing for sound and vibration engineers*. Wiley, 2008.

[28] O. Kirkeby, P. Nelson, H. Hamada, and F. Orduna-Bustamante, "Fast deconvolution of multichannel systems using regularization," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 2, pp. 189–194, Mar 1998.

[29] M. Wu, B. Yin, A. Vosoughi, C. Studer, J. R. Cavallaro, and C. Dick, "Approximate matrix inversion for high-throughput data detection in the large-scale MIMO uplink," *Proc. - IEEE Int. Symp. Circuits Syst.*, pp. 2155–2158, 2013.

[30] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.

[31] J. T. Schwartz, *Introduction to Matrices and Vectors*, Dover Books on Mathematics. Dover Publications, 2012.

[32] K. Baykaner, P. Coleman, M. Olik, and S. Bech, "The Relationship Between Target Quality and Interference in Sound Zone", *J. Audio Eng. Soc.*, 63(1), 2015.

[33] L. Ljung and T. Söderström, "Theory and practice of recursive identification." The MIT Press, 1985.

[34] D. A. Bini, P. Boito, Y. Eidelman, L. Gemignani, and I. Gohberg, "A fast implicit QR eigenvalue algorithm for companion matrices," *Linear Algebra and its Applications*, vol. 432, no. 8, pp. 2006-2031, Apr. 2010.

[35] T. H. Cormen and C. E. Leiserson, "Introduction to algorithms," 3rd edition. MIT Press, 2009.

[36] A. Edelman and H. Murakami, "Polynomial roots from companion matrix eigenvalues," *Mathematics of Computation*, vol. 64, no. 210, pp. 763-776, 1995.

[37] S. Chandrasekaran, M. Gu, J. Xia, and J. Zhu, "A Fast QR Algorithm for Companion Matrices," in Recent Advances in Matrix and Operator Theory, vol. 179, no. 7, Basel: Birkhäuser Basel, 2007, pp. 111-143.

[38] T. Betlehem and T. D. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms," *J. Acoust. Soc. Am.*, vol. 117, no. 4, pp. 2100-2111, Apr. 2005.

[39] M. F. Simon Galvez and S. J. Elliott, "The Design of a Personal Audio Superdirective Array in a Room," presented at the AES 52nd International Conference Sound Field Control - Engineering and Perception, 2013, pp. 240-247.

[40] M. F. Simon Galvez, S. J. Elliott, and J. Cheer, "The effect of reverberation on personal audio devices," *J. Acoust. Soc. Am.*, vol. 135, no. 5, pp. 2654-2663, May 2014.

[41] T. Betlehem and M. A. Poletti, "Two dimensional sound field reproduction using higher order sources to exploit room reflections," *J. Acoust. Soc. Am.*, vol. 135, no. 4, pp. 1820-1833, Apr. 2014.

[42] M. A. Poletti, T. Betlehem, and T. D. Abhayapala, "Higher-Order Loudspeakers and Active Compensation for Improved 2D Sound Field Reproduction in Rooms," *J. Audio Eng. Soc.*, vol. 63, no. 1, pp. 31-45, Jan. 2015.

[43] M. Olik, P. J. Jackson, P. Coleman, M. Olsen, M. Mo/ller, and S. Bech"Influence of low-order room reflections on sound zone system performance," *J. Acoust. Soc. Am.*, 133, 3349–3349 (2013).

[44] Kolundzija, Mihailo and Faller, Christof and Vetterli, Martin, "Design of a Compact Cylindrical Loudspeaker Array for Spatial Sound Reproduction," Audio Engineering Society Convention 130, May 2011.

[45] T. Hooley, "Single box surround sound," *Acoustical Science and Technology*, 27(6), pp. 354-360, 2006.

[46] A. Canclini, D. Markovic, F. Antonacci, A. Sarti, and S. Tubaro, "A room-compensated virtual surround system exploiting early reflections in a reverberant room," in 20th European Signal Processing Conference (EUSIPCO), pp. 1029–1033, 2012.

[47] F. Olivieri, M. Shin, F. Fazi, P. Nelson, and P. Otto, "Loudspeaker array processing for multi-zone audio reproduction based on analytical and measured electroacoustical transfer functions," Audio Eng. Soc. Conf. 52nd Int. Conf. Sound F. Control - Eng. Percept., pp. 1–10, 2013.

[48] M. F. Simon Galvez and F. M. Fazi, "Room Compensation for Binaural Reproduction with Loudspeaker Arrays," presented at the EuroRegio 2016, Porto, Portugal.

[49] The MathWorks, Inc., "Matlab 2016b Polynomial roots" [Online]. Available: https://uk.mathworks.com/help/matlab/ref/roots.html [Accessed: 22-Jan-2017].

[50] E. G. Williams, Fourier Acoustics. Academic Press, 1999.

[51] F. Olivieri, Matlab implementations of the algorithms presented in this paper. Available for download at http://ieeexplore.ieee.org and https://zenodo.org/badge/latestdoi/89870495.

[52] B. D. Carlson and D. Willner, "Antenna pattern synthesis using weighted least squares," IEE Proceedings H Microwaves, Antennas and Propagation, vol. 139, no. 1, p. 11, 1992.

[53] J. Eriksson and M. Viberg, "Asymptotic Properties of Nonlinear Weighted Least Squares in Radar Array Processing," IEEE Trans. Signal Process., vol. 52, no. 11, pp. 30833095, Nov. 2004.

Appendix A
Relations between $|p_B|$, $\|\mathbf{q}\|$, and β_0

In this Appendix, we prove that the relations between $|p_B|$, $\|\mathbf{q}\|$ and β are

$$|p_B| \leq \|\mathbf{z}_B\| \, \|\mathbf{q}\| \, , \text{and} \, \|\mathbf{q}\| = \|\mathbf{z}_B\| \Big/ \left(\lambda_L + \beta_0 \sigma_1^2\right), \qquad (24)$$

where λ_L is the smallest eigenvalue of matrix $\mathbf{Z}^H \mathbf{\Psi}^2 \mathbf{Z}$. The proof of Eq. (24) is as follows. By using the Cauchy-Schwarz inequality on $|p_B| = \left|\mathbf{z}_B^T \mathbf{q}\right|$ and Eq. (7), we can write

$$|p_B| \leq \|\mathbf{z}_B\|^2 \left\| \left(\mathbf{Z}^H \hat{\mathbf{\Psi}}^2 \mathbf{Z} + \beta \mathbf{I}\right)^{-1} \right\|. \qquad (25)$$

We apply the eigenvalue decomposition to matrix $\mathbf{Z}^H \hat{\mathbf{\Psi}}^2 \mathbf{Z} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^H$, where $\mathbf{\Lambda}$ is the matrix of eigenvalues $\lambda_1, \cdots, \lambda_L$ (in decreasing order) and $\mathbf{U}$ is that of the eigenvectors. It follows that $\mathbf{Z}^H \hat{\mathbf{\Psi}}^2 \mathbf{Z} + \beta \mathbf{I} = \mathbf{U} \left(\mathbf{\Lambda} + \beta \mathbf{I}\right) \mathbf{U}^H$. By recalling that $\|\mathbf{U}\| = \left\|\mathbf{U}^H\right\| = 1$, we can write $|p_B| \leq \|\mathbf{z}_B\|^2 \left\| \left(\mathbf{\Lambda} + \beta \mathbf{I}\right)^{-1} \right\|$. We recall that $\left\| \left(\mathbf{\Lambda} + \beta \mathbf{I}\right)^{-1} \right\|$ is equal to the maximum value among the elements in the matrix, that is

$$\left\| \left(\mathbf{\Lambda} + \beta \mathbf{I}\right)^{-1} \right\| = 1 / \left(\lambda_L + \beta\right). \qquad (26)$$

Substituting Eq. (26) and Eq. (10) in Eq. (25) yields the result in the left-hand side of Eq. (24). Similarly, by writing $\|\mathbf{q}\| \leq \left\| \left(\mathbf{Z}^H \hat{\mathbf{\Psi}}^2 \mathbf{Z} + \beta_0 \sigma_1^2 \mathbf{I}\right)^{-1} \right\| \|\mathbf{z}_B^*\|$ and by using Eq. (26) and Eq. (10), we obtain the result in the right-hand side of Eq. (24). Eq. (24) suggests that increasing values of β_0 (and, hence, β) may lead to a reduction of the pressure field in the listening zone and of the input energy.

Appendix B
Convergence of the Neumann series

We prove that Eq. (17) holds. The following relations holds [19] $\rho\left(-\Delta\psi_D \mathbf{A}^{-1} \mathbf{Z}_D^H \mathbf{Z}_D\right) \leq |\Delta\psi_D| \left\|-\mathbf{A}^{-1} \mathbf{Z}_D^H \mathbf{Z}_D\right\|$ and $\left\|-\mathbf{A}^{-1} \mathbf{Z}_D^H \mathbf{Z}_D\right\| \leq \left\|\mathbf{A}^{-1}\right\| \left\|\mathbf{Z}_D^H \mathbf{Z}_D\right\|$. By considering $\mathbf{A}$ in its general form, i.e. $\mathbf{A} = \mathbf{Z}^H \hat{\mathbf{\Psi}}_0^2 \mathbf{Z} + \beta \mathbf{I}$, we take the Singular Value Decomposition (SVD) of $\mathbf{A} = \mathbf{U}_0 \left(\mathbf{\Lambda}_0 + \beta \mathbf{I}\right) \mathbf{U}_0^H$. By following a similar procedure as that in Appendix A, we write $\left\|\mathbf{A}^{-1}\right\| \leq \left\| \left(\mathbf{\Lambda}_0 + \beta \mathbf{I}\right)^{-1} \right\| = 1 / \left(\lambda_L^{(0)} + \beta\right)$. By using the results above, we obtain the result in Eq. (17).

2329-9290 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TASLP.2017.2700945, IEEE/ACM Transactions on Audio, Speech, and Language Processing

APPENDIX C
PROOF OF REAL VALUED $\tilde{p}_{B,n}$

In this section, we prove that the coefficients $\tilde{p}_{B,n}$ in Eq. (20) are real valued $\forall n$. Let us define $\mathbf{L} = \mathbf{z}_B^* \mathbf{z}_B^T + \psi_G \mathbf{Z}_G^H \mathbf{Z}_G + \beta \mathbf{I}$, which gives $\widehat{\psi_D} \mathbf{Z}_D^H \mathbf{Z}_D = \mathbf{A} - \mathbf{L}$. Hence, the term $\left(\mathbf{A}^{-1}\mathbf{Z}_D^H\mathbf{Z}_D\right)^n$ in Eq. (20) can be also written as

$$\left(\mathbf{A}^{-1}\mathbf{Z}_D^H\mathbf{Z}_D\right)^n = \left(\widehat{\psi_D}^{-1}\mathbf{A}^{-1}(\mathbf{A}-\mathbf{L})\right)^n = \widehat{\psi_D}^{-n}\left(\mathbf{I}-\mathbf{A}^{-1}\mathbf{L}\right)^n. \tag{27}$$

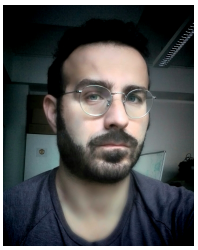
By using Eq. (27) and Eq. (12), $\tilde{p}_{B,n}$ in Eq. (20) can then be written as $\tilde{p}_{B,n} = (-1)^n \widehat{\psi_D}^{-n} \mathbf{z}_B^T \left(\mathbf{I}-\mathbf{A}^{-1}\mathbf{L}\right)^n \mathbf{A}^{-1}\mathbf{z}_B^*$. The term $\left(\mathbf{I}-\mathbf{A}^{-1}\mathbf{L}\right)^n = (-1)^n\left(\mathbf{A}^{-1}\mathbf{L}-\mathbf{I}\right)^n$ can be expressed in terms of a binomial expansion [31] $(-1)^n\left(\mathbf{A}^{-1}\mathbf{L}-\mathbf{I}\right)^n = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!}\left(\mathbf{A}^{-1}\mathbf{L}\right)^k(-1)^k$, where $n!$ indicates the factorial of n, thus yielding an alternative expression for $\tilde{p}_{B,n}$, that is

$$\tilde{p}_{B,n} = \sum_{n=0}^{N}(-2)^n \sum_{k=0}^{n} \frac{n!}{k!(n-k)!}(-1)^k \mathbf{z}_B^T\left(\mathbf{A}^{-1}\mathbf{L}\right)^k \mathbf{A}^{-1}\mathbf{z}_B^*. \tag{28}$$

By noting that $\left(\mathbf{A}^{-1}\mathbf{L}\right)^k = \mathbf{A}^{-1}\mathbf{L}\left(\mathbf{A}^{-1}\mathbf{L}\right)^{k-1}$ and $\mathbf{z}_B^T\mathbf{A}^{-1} = \mathbf{q}^H$, the term $\mathbf{z}_B^T\left(\mathbf{A}^{-1}\mathbf{L}\right)^k\mathbf{A}^{-1}\mathbf{z}_B^*$ in Eq. (28) can be also rewritten as $\mathbf{q}^H\mathbf{L}\left(\mathbf{A}^{-1}\mathbf{L}\right)^{k-1}\mathbf{q}$. We now show that $\mathbf{q}^H\mathbf{L}\left(\mathbf{A}^{-1}\mathbf{L}\right)^{k-1}\mathbf{q}$ is real-valued for all k. This is possible thanks to the fact that $\mathbf{q}^H\mathbf{Y}\mathbf{q}$ is real valued for any complex vector $\mathbf{q}$ if $\mathbf{Y}$ is Hermitian [19]. We prove that matrix $\mathbf{L}\left(\mathbf{A}^{-1}\mathbf{L}\right)^{k-1}$ is Hermitian in two cases. 1) If $k-1$ is odd, that is $k-1 = 2k'+1$, then Let us write $\mathbf{q}' = \mathbf{L}\left(\mathbf{A}^{-1}\mathbf{L}\right)^{k'}\mathbf{q}$. It follows that $(\mathbf{q}')^H = \mathbf{q}^H\mathbf{L}\left(\mathbf{A}^{-1}\mathbf{L}\right)^{k'}$. Since both $\mathbf{L}$ and $\mathbf{A}^{-1}$ are Hermitian matrices (i.e., $\mathbf{A}^{-1} = (\mathbf{A}^{-1})^H$ and $\mathbf{L} = \mathbf{L}^H$), we can find a complex vector $\mathbf{q}' = \mathbf{L}\left(\mathbf{A}^{-1}\mathbf{L}\right)^{k'}\mathbf{q}$ so that $\mathbf{q}^H\mathbf{L}\left(\mathbf{A}^{-1}\mathbf{L}\right)^{k-1}\mathbf{q} = (\mathbf{q}')^H\mathbf{A}^{-1}\mathbf{q}'$. Hence, because $\mathbf{A}^{-1}$ is Hermitian, the quantity $(\mathbf{q}')^H\mathbf{A}^{-1}\mathbf{q}'$ is real valued. 2) If $k-1$ is even, that is $k-1 = 2k'$, then $\left(\mathbf{A}^{-1}\mathbf{L}\right)^{2k'} = \left(\mathbf{A}^{-1}\mathbf{L}\right)^{k'}\left(\mathbf{A}^{-1}\mathbf{L}\right)^{k'}$ and $\mathbf{q}^H\mathbf{L}\left(\mathbf{A}^{-1}\mathbf{L}\right)^{k-1}\mathbf{q} = \mathbf{q}^H\mathbf{L}\left(\mathbf{A}^{-1}\mathbf{L}\right)^{k'}\left(\mathbf{A}^{-1}\mathbf{L}\right)^{k'}\mathbf{q}$. Two vectors $\mathbf{q}' = \left(\mathbf{A}^{-1}\mathbf{L}\right)^{k'}\mathbf{q}$ and $(\mathbf{q}')^H = \mathbf{q}^H\left(\mathbf{L}\mathbf{A}^{-1}\right)^{k'}$ exist, so that $\mathbf{q}^H\mathbf{L}\left(\mathbf{A}^{-1}\mathbf{L}\right)^{k-1}\mathbf{q} = (\mathbf{q}')^H\mathbf{L}\mathbf{q}'$. Hence, since $\mathbf{L}$ is Hermitian, the quantity $(\mathbf{q}')^H\mathbf{L}\mathbf{q}'$ is real valued.

Ferdinando Olivieri received his MSc in Telecommunications Engineering from the University of Florence (Italy) in 2010 and his MSc in Acoustics from the Institute of Sound and Vibration Research (ISVR), University of Southampton (UK), in 2012. He received his PhD in Audio Signal Processing from ISVR with a thesis on signal processing for loudspeaker arrays and personalized audio. He is currently working at Qualcomm as a Research Engineer in 3D Audio. His research interests include array signal processing, inverse problems for acoustics, beamforming, and sound field capturing and reproduction. He is a member of the Audio Engineering Society and of the IEEE Signal Processing Society.

Filippo Maria Fazi graduated in Mechanical Engineering from the University of Brescia (Italy) in 2005. He obtained his PhD in acoustics from the Institute of Sound and Vibration Research (ISVR) of the University of Southampton, UK, in 2010, with a thesis on sound field reproduction. In the same year, he was awarded a fellowship by the Royal Academy of Engineering and by the Engineering and Physical Sciences Research Council. He is currently an Associate Professor at the University of Southampton. Dr Fazi s research interests include Audio technologies, Electroacoustics and Digital Signal Processing, with special focus on acoustical inverse problems, multi-channel systems, virtual acoustics, microphone and loudspeaker arrays.



Simone Fontana received the M.Sc. degree in telecommunication engineering from Politecnico di Torino and Telecom ParisTech, and the Ph.D. degree in audio signal processing and technical physics from Telecom ParisTech and University of Parma. Since 2013 he works in Huawei European Research Center as Senior Research Fellow and Technical Team Leader. His current research interests include spatial audio processing, speech enhancement and new transducers for communication devices.

Dylan Menzies is a Senior Research Fellow in the Institute of Sound and Vibration, at the University of Southampton. Areas of interest include spatial audio synthesis and reproduction, sound synthesis for virtual environments, and musical synthesis and interfaces. He holds a PhD in electronics from the University of York, an MA in mathematics from Cambridge University, and has worked as a research engineer for several companies including Sony Professional Audio.

Philip Nelson holds the post of Professor of Acoustics in the Institute of Sound and Vibration Research at the University of Southampton. He has personal research interests in the fields of acoustics, vibrations, signal processing, control systems and fluid dynamics and is the author or co-author of 2 books, over 120 papers in refereed journals, 30 granted patents, and over 200 other technical publications. He served from 2005-2013 as Pro Vice-Chancellor of the University of Southampton, with particular responsibility for Research and Enterprise. He previously served as Director of the Institute of Sound and Vibration Research and as Director of the Rolls-Royce University Technology Centre in Gas Turbine Noise. Professor Nelson is a Fellow of the Royal Academy of Engineering, a Chartered Engineer, a Fellow of the Institution of Mechanical Engineers, a Fellow of the Institute of Acoustics, and a Fellow of the Acoustical Society of America. He is the recipient of both the Tyndall and Rayleigh Medals of the Institute of Acoustics, and was President of the International Commission for Acoustics from 2004-2007. He also served as the Chair of the Sub-Panel for General Engineering in REF 2014. From 2014 he has been Chief Executive of the Engineering and Physical Sciences Research Council and took on the role of Chair of the RCUK Executive Group from October 2015.

2329-9290 (c) 2016 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.