Personal Sound Zones by Subband Filtering and Time Domain Optimization

Vicent Molés-Cases, Gema Piñero, *Senior Member, IEEE*, Maria de Diego, *Senior Member, IEEE*, and Alberto Gonzalez, *Senior Member, IEEE*

***Abstract*—Personal Sound Zones (PSZ) systems aim to render independent sound signals to multiple listeners within a room by using arrays of loudspeakers. One of the algorithms used to provide PSZ is Weighted Pressure Matching (wPM), which computes the filters required to render a desired response in the listening zones while reducing the acoustic energy arriving to the quiet zones. This algorithm can be formulated in time and frequency domains. In general, the time-domain formulation (wPM-TD) can obtain good performance with shorter filters and delays than the frequency-domain formulation (wPM-FD). However, wPM-TD requires higher computation for obtaining the optimal filters. In this article, we propose a novel approach to the wPM algorithm named Weighted Pressure Matching with Subband Decomposition (wPM-SD), which formulates an independent time-domain optimization problem for each of the subbands of a Generalized Discrete Fourier Transform (GDFT) filter bank. Solving the optimization independently for each subband has two main advantages: 1) lower computational complexity than wPM-TD to compute the optimal filters; 2) higher versatility than the classic wPM algorithms, as it allows different configurations (sets of loudspeakers, filter lengths, etc.) in each subband. Moreover, filtering the input signals with a GDFT filter bank (as in wPM-SD) requires lower computational effort than broadband filtering (as in wPM-TD and wPM-FD), which is beneficial for practical PSZ systems. We present experimental evaluations showing that wPM-SD offers very similar performance to wPM-TD. In addition, two cases where the versatility of wPM-SD is beneficial for a PSZ system are presented and experimentally validated.**

***Index Terms*—Personal audio, sound zones, subband filtering.**

I. Introduction

PERSONAL Sound Zones (PSZ) systems aim to render different audio signals to different users within a room with minimum leakage between them [1], [2]. To achieve it, an array of loudspeakers and a set of filters (one per loudspeaker and user) are used. The goal of PSZ algorithms is to find the optimal filters for the system. Usually, the term bright zone denotes the region where an audio signal is rendered, and dark zone the region where the sound level is minimized [3].

Manuscript received February 26, 2020; revised June 11, 2020 and July 31, 2020; accepted September 3, 2020. Date of publication September 11, 2020; date of current version September 30, 2020. This work was supported by Grants RTI2018-098085-B-C41 (MCIU/AEI/FEDER, UE), RED2018-102668-T and PROMETEO/2019/109. The work of Vicent Molés-Cases has been supported by Spanish Ministry of Education under Grant FPU17/01288. The associate editor coordinating the review of this article and approving it for publication was Prof. Stefania Cecchi. *(Corresponding author: Vicent Molés-Cases.)*

The authors are with the Institute of Telecommunications and Multimedia Applications, Universitat Politècnica de València, Valencia 46022, Spain (e-mail: vimoca3@iteam.upv.es; gpinyero@iteam.upv.es; mdediego@dcom.upv.es; agonzal@dcom.upv.es).

Digital Object Identifier 10.1109/TASLP.2020.3023628

Several techniques can be used to obtain the filters for the system, such as beamforming [4], [5], soundfield synthesis [6], [7], energy cancellation approaches [8], [9], or hybrid approaches [10], [11]. Among these, the hybrid technique Weighted Pressure Matching (wPM) [12] exhibits a compromise between the performance obtained with two different techniques, namely, Pressure Matching (PM) [13] and Acoustic Contrast Control (ACC) [3]. On the one hand, PM is a soundfield synthesis approach that aims to render a target response in the bright zone, but does not offer control over the energy in the dark zone [13]. On the other hand, ACC is an energy cancellation approach whose goal is to minimize the energy in the dark zone, but it can not synthesize a target response in the bright zone [3]. Instead, wPM offers the possibility to both render a target response in the bright zone while keeping control over the energy in the dark zone. To do so, the authors in [12] proposed a novel cost function in which a weighting parameter is used to balance the components of the cost functions for PM and ACC.

Originally, wPM was proposed as a frequency-domain approach (wPM-FD), i.e., the optimal filter coefficients are computed independently for a set of control frequencies, and then, the time-domain filters are obtained by windowing the Inverse Discrete Fourier Transform (IDFT) of the optimal filter coefficients [14]. Later, the authors in [15] proposed a formulation of wPM in time-domain (wPM-TD), in which the time-domain filters are directly obtained by solving a single optimization problem. In general, the performance of both techniques is very similar when long filters are used, however, wPM-TD outperforms wPM-FD for short filters [15]. Moreover, wPM-FD requires a modelling delay of about half the filter length to assure the causality of the filters, while shorter delays can be selected for wPM-TD [15]. This is specially important for applications where low latency is required, e.g., two-way telecommunication applications. However, the computational demands for solving the optimization are significantly higher with wPM-TD, as it involves the inverse of a large matrix. The higher complexity is not an important limitation for static PSZ systems, where the filters can be computed off-line. However, it is an important aspect for dynamic PSZ systems where the location of the zones may change over time [16] or when the characteristics of the input signal are taken into account [17], as the filters must be often recomputed. Consequently, an algorithm offering similar performance to wPM-TD but with a lower computational complexity could be very useful for certain applications. Recently, the authors in [18] proposed an iterative algorithm to efficiently obtain the solution for wPM-TD. However, its convergence rate is related to a set of parameters, which can lead to a high

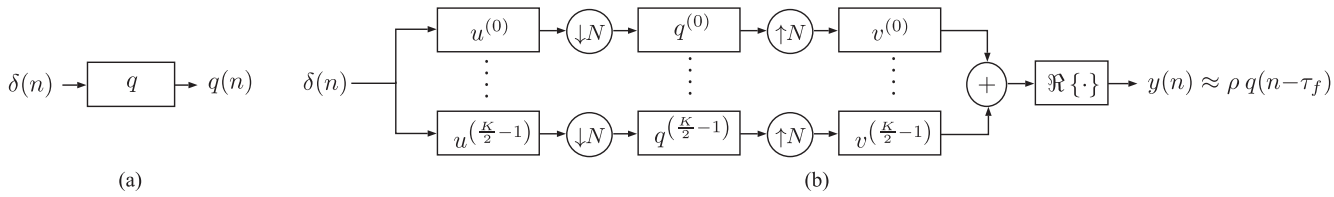

Fig. 1. (a) Diagram of a FIR filter q. (b) Subband approximation of the FIR filter q using a GDFT filter bank.

number of iterations if not selected properly (and there is few intuition in how to select them). Thus, we propose in this article a non-iterative algorithm with lower computational demands than wPM-TD.

An important aspect of PSZ systems in real-world scenarios is that long filters must be used to compensate for the room reverberation [19]. Such long filters require high computational complexity to filter the input signals, which in some cases can be unaffordable for real-time applications [20]. The computational demands can be lowered using filter-bank structures [21], as the filtering can be performed efficiently for each subband at a lower sampling rate. The authors in [22] proposed to use Quadrature Mirror Filters (QMFs) with non-uniform subbands, where a frequency-domain approach is used to obtain the filters for each subband. Particularly, the authors formulate the optimization problem for each subband by using the Room Impulse Response (RIR) filtered with the analysis filters of the filter bank. This approach is limited to frequency-domain optimizations, as the effect of the analysis filters in the RIR would produce a frequency weighting in a time-domain optimization, which would degrade the performance in the edges of the subbands. Then, this approach can be used to reduce the computational complexity of the filtering operation, but still presents the same limitations as wPM-FD (i.e., longer filters and delays) because the optimization for each subband is solved in the frequency-domain.

In this article, we propose to use a Generalized Discrete Fourier Transform (GDFT) filter bank [23] to perform the filtering operation using a set of subband filters. To obtain the filters, we propose the Weighted Pressure Matching with Subband Decomposition (wPM-SD) algorithm, which formulates an independent time-domain optimization problem for each subband. To avoid the undesired frequency weighting produced by the analysis filters of the filter bank, we propose to formulate the optimization using the subband components of the RIR obtained with the algorithm presented in [24]. Using a time-domain optimization for each subband offers several advantages with respect to the approach presented in [22], as shorter filters and delays. We show by experimental validation that the performance obtained by wPM-SD can be approximately equal to the one showed by wPM-TD (the validity of the approximation depends on the prototype filter selected for the filter bank). Solving the optimization independently for each subband can significantly decrease the required computational complexity, as the matrix that must be inverted for each subband is significantly smaller than with wPM-TD. Also, the computational demands to filter the input signals are lower than with wPM-TD, as the filtering for each subband is performed at a lower sampling rate. Moreover, as the optimization is independently solved for each subband, the proposed algorithm offers a versatility that wPM-TD lacks, since different sets of loudspeakers, microphones, and filter lengths can be selected for each subband.

The article is structured as follows. Section II studies the subband decomposition of a Finite Impulse Response (FIR) in its subband components for a GDFT filter bank. Section III reviews the wPM-TD algorithm. Section IV presents the novel wPM-SD algorithm. Section V presents experimental results to show the performance of the proposed algorithm. Finally, Section VI summarizes the main conclusions.

Notation: Throughout this article matrices and vectors are represented by upper and lower case boldface letters, respectively, $(\cdot)^T$ stands for transpose, $(\cdot)^H$ stands for conjugate transpose, $\|\cdot\|$ for vector 2-norm, subindex $\downarrow N$ denotes downsampling by a factor N, $*$ denotes the discrete convolution operation, $\mathbb{Z}$ denotes the set of all integers, $\mathbf{I}$ denotes the identity matrix, $\delta(n)$ denotes the unit impulse, and $\lceil\cdot\rceil$ denotes the ceiling function.

II. Subband Decomposition

In this section, we study the decompostion of a FIR filter in its subband components using a GDFT filter bank. The main advantage of GDFT filter banks is that a computationally efficient implementation is possible for fractional oversampling [23], i.e., for $0 < (N/K) \leq 1$ (being N the resampling factor and K the number of subbands).

Fig. 1(b) illustrates a GDFT filter bank with K subbands, where $u^{(k)}$ and $v^{(k)}$ denote the analysis and synthesis filters for the k-th subband, respectively. In the following, we focus on the case with even number of subbands.[1] The analysis and synthesis filters are single-sided responses (i.e., with complex coefficients), and are obtained by modulating a low pass prototype filter $p(n)$ of length L_p and bandwidth $2\pi/K$ as

$$u^{(k)}(n) = p(n)e^{j\frac{2\pi}{K}\left(k+\frac{1}{2}\right)n}, \tag{1}$$

$$v^{(k)}(n) = p(L_p-1-n)e^{-j\frac{2\pi}{K}\left(k+\frac{1}{2}\right)n}, \tag{2}$$

for $0 \leq n \leq L_p - 1$ and $0 \leq k \leq K-1$. The analysis and synthesis filters in the positive and negative spectrum are complex conjugates, i.e., $u^{(k)}(n) = (u^{(K-1-k)}(n))^* \,\forall k$ and $v^{(k)}(n) = (v^{(K-1-k)}(n))^* \,\forall k$. Then, only $K/2$ subbands must be processed, as the other subbands are their complex conjugates. From (2), it follows that the synthesis filters are time-reversed and conjugated versions of the analysis filters.

[1]For a GDFT filter bank, K can be either even or odd. For K even, the $K/2$ subbands in the positive and negative spectrum are complex conjugates. For K odd, the $(K-1)/2$ subbands in the positive and negative spectrum are complex conjugates and the subband in the center of the spectrum is unique.

Given a FIR filter denoted by q with length L_q, the main goal of the subband decomposition is to find the set of subband components $q^{(k)}$ for a GDFT filter bank such that the systems in Fig. 1(a) and Fig. 1(b) are equivalent. We consider that both systems are equivalent when $S(z) = \rho Q(z) z^{-\tau_f}$, where $Q(z)$ is the z-transform of q, $S(z)$ is the z-transform of the overall response of the filter bank including the subband components $q^{(k)}$, τ_f is a delay produced by the filter bank, and ρ is an arbitrary amplitude scaling. Equality between both systems is only achieved when an ideal low pass filter is used as prototype filter [24]. For the case with non-ideal prototype filter, the equality can only be assumed approximately. However, it is shown in [24] that the approximation error can be negligible if the following conditions are fulfilled:

1) Negligible reconstruction error, i.e.,

$$\sum_{k=0}^{K/2-1} \Re\left\{U^{(k)}(z)V^{(k)}(z)\right\} \approx \rho z^{-\tau_f}, \tag{3}$$

2) Negligible aliasing level in the subbands, i.e.,

$$\sum_{i=1}^{N-1} U^{(k)}(z\Phi_N^i)V^{(k)}(z) \approx 0, \quad \forall k, \tag{4}$$

where $z = e^{j\omega}$ for $0 \leq \omega < 2\pi$, $\Phi_N^i = e^{-j\frac{2\pi i}{N}}$, and $U^{(k)}(z)$ and $V^{(k)}(z)$ are the z-transforms of $u^{(k)}$ and $v^{(k)}$, respectively. For a GDFT filter bank, (3) is achieved if the prototype filter is a root-Nyquist filter [24]. Regarding (4), it must be noticed that low values of (N/K) and high values of L_p assure low aliasing levels, but require a higher computational effort. Therefore, a compromise between (N/K) and L_p must be reached such that (4) is fulfilled while still benefiting from a reduction of the computational complexity (as we will show in Section V-B).

The authors in [24] show that for a GDFT filter bank fulfilling (3) and (4), the subband components $q^{(k)}$ that make the systems in Fig. 1(a) and Fig. 1(b) approximately equivalent are given by

$$q^{(k)}(n) = \min_{q^{(k)}(n)} \left\{ \sum_{n\in\mathbb{Z}} \left| u_{\downarrow N}^{(k)}(n) * q^{(k)}(n) - \check{q}^{(k)}(n) \right|^2 \right\}, \tag{5}$$

where $u_{\downarrow N}^{(k)}$ is the downsampled analysis filter of length $L_{p\downarrow} = \lceil L_p/N \rceil$ for the k-th subband, and $\check{q}^{(k)}(n)$ is the analysis component of $q(n)$ of length $L_{\check{q}} = \lceil (L_q + L_p - 1)/N \rceil$ for the k-th subband, which is defined as

$$\check{q}^{(k)}(n) = \left(u^{(k)}(n) * q(n) \right)_{\downarrow N}. \tag{6}$$

Then, the length of the subband component $q^{(k)}$ is given by

$$L'_q = L_{\check{q}} - L_{p\downarrow} + 1 = \left\lceil \frac{L_q + L_p - 1}{N} \right\rceil - \left\lceil \frac{L_p}{N} \right\rceil + 1. \tag{7}$$

From (5), we can see that the optimal subband component minimizes the mean square error between the analysis component (6) and the subband output when feeding the filter bank with $\delta(n)$. Also, it is important to highlight that the subband components have complex coefficients because the analysis and synthesis filters are single-sided. Now, if we define

$$\mathbf{q}^{(k)} = \left[q^{(k)}(0) \ \dots \ q^{(k)}(L'_q - 1) \right]^T,$$

$$\check{\mathbf{q}}^{(k)} = \left[\check{q}^{(k)}(0) \ \dots \ \check{q}^{(k)}(L_{\check{q}} - 1) \right]^T,$$

$$\mathbf{U}^{(k)} = \begin{bmatrix} u_{\downarrow N}^{(k)}(0) & \dots & u_{\downarrow N}^{(k)}(L_{p\downarrow}-1) & 0 & \dots & 0 \\ 0 & & & & & \\ \vdots & & Toeplitz & & & \\ 0 & & & & & \end{bmatrix}^T,$$

we can re-write (5) as

$$\begin{aligned} \mathbf{q}^{(k)} &= \min_{\mathbf{q}^{(k)}} \left\{ \left\| \mathbf{U}^{(k)}\mathbf{q}^{(k)} - \check{\mathbf{q}}^{(k)} \right\|^2 \right\} \\ &= ((\mathbf{U}^{(k)})^H \mathbf{U}^{(k)})^{-1} (\mathbf{U}^{(k)})^H \check{\mathbf{q}}^{(k)}. \end{aligned} \tag{8}$$

In (8), the term $((\mathbf{U}^{(k)})^H \mathbf{U}^{(k)})^{-1}(\mathbf{U}^{(k)})^H$ performs the deconvolution of the downsampled analysis filter and the analysis component for the k-th subband. Then, this term removes from the subband component any frequency weighting produced by the analysis filter. Thus, using the subband components obtained with (8) for the formulation of an inverse problem in time-domain does not lead to any frequency weighting. This is an important advantage with respect to the approach presented in [22], where the analysis components (6) are used to define the optimization problem, and consequently, an undesired frequency weighting that degrades the performance is produced if the optimization is formulated in time-domain.

Finally, it is important to notice that (8) requires to compute $K/2$ different matrix inversions to obtain the $K/2$ subband components. Next, we derive a novel solution that only requires one matrix inversion. We start by re-writing $\mathbf{U}^{(k)}$ as

$$\mathbf{U}^{(k)} = \mathbf{F}_1^{(k)} \mathbf{P} \mathbf{F}_2^{(k)}, \tag{9}$$

where $\mathbf{F}_1^{(k)}$ and $\mathbf{F}_2^{(k)}$ are diagonal matrices of size $L_{\check{q}}$ and L'_q, respectively, which are defined as

$$\mathbf{F}_1^{(k)} = \text{Diag}\left(\left[e^{j\frac{2\pi N}{K}\left(k+\frac{1}{2}\right)0}, \dots, e^{j\frac{2\pi N}{K}\left(k+\frac{1}{2}\right)(L_{\check{q}}-1)} \right] \right),$$

$$\mathbf{F}_2^{(k)} = \text{Diag}\left(\left[e^{-j\frac{2\pi N}{K}\left(k+\frac{1}{2}\right)0}, \dots, e^{-j\frac{2\pi N}{K}\left(k+\frac{1}{2}\right)(L'_q-1)} \right] \right),$$

and $\mathbf{P}$ is the convolution matrix of the downsampled prototype filter of size $L_{\check{q}} \times L'_q$, which is defined as

$$\mathbf{P} = \begin{bmatrix} p_{\downarrow N}(0) & \dots & p_{\downarrow N}(L_{p\downarrow}-1) & 0 & \dots & 0 \\ 0 & & & & & \\ \vdots & & Toeplitz & & & \\ 0 & & & & & \end{bmatrix}^T.$$

Now, we can use (9) to re-write (8) as

$$\begin{aligned} \mathbf{q}^{(k)} &= \left(\mathbf{F}_2^{(k)^H} \mathbf{P}^T \mathbf{F}_1^{(k)^H} \mathbf{F}_1^{(k)} \mathbf{P} \mathbf{F}_2^{(k)} \right)^{-1} \mathbf{F}_2^{(k)^H} \mathbf{P}^T \mathbf{F}_1^{(k)^H} \check{\mathbf{q}}^{(k)} \\ &= \mathbf{F}_2^{(k)^H} \left(\mathbf{P}^T \mathbf{P} \right)^{-1} \mathbf{P}^T \mathbf{F}_1^{(k)^H} \check{\mathbf{q}}^{(k)}. \end{aligned} \tag{10}$$

Equation (10) has the key property that the real matrix $(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T$ is common to all subbands, so only one matrix must be inverted for computing all the subband components. In Section IV, we will use (10) to obtain the subband components of the RIR required by the proposed wPM-SD algorithm.

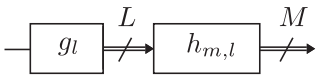

Fig. 2. Model of a PSZ system using FIR broadband filters.

III. Weighted Pressure Matching

Let us consider a PSZ system formed by an array of L loudspeakers and $M = M_d + M_b$ control points, where M_b and M_d are the number of control points used to spatially sample the bright and dark zones, respectively. The system aims to render a given audio signal in the bright zone while keeping the interference in the dark zone as low as possible. To achieve it, the audio signal is filtered through a FIR filter denoted by g_l prior to be fed to the l-th loudspeaker. Fig. 2 shows the system model. From now on, we denote $\mathcal{B}$ and $\mathcal{D}$ as the sets of control points for the bright and dark zones, respectively.

Next, we describe the model used for the time-domain formulation of wPM. Usually, the design of PSZ algorithms for static systems does not consider the characteristics of the audio signal [25]. Therefore, the following model deals with impulse responses rather than signals. Thus, let us define $h_{m,l}$ as the L_h-length Room Impulse Response (RIR) between the l-th loudspeaker and the m-th control point, and g_l as the L_g-length FIR filter for the l-th loudspeaker. We define the cascade impulse response for the m-th control point as

$$x_m(n) = \sum_{l=0}^{L-1} h_{m,l}(n) * g_l(n), \tag{11}$$

and the vector with the non-zero samples of $x_m(n)$ as

$$\mathbf{x}_m = \left[x_m(0) \ \dots \ x_m(L_g + L_h - 2)\right]^T = \sum_{l=0}^{L-1} \mathbf{H}_{m,l}\mathbf{g}_l, \tag{12}$$

where $\mathbf{g}_l = [g_l(0), \dots, g_l(L_g-1)]^T$, and

$$\mathbf{H}_{m,l} = \begin{bmatrix} h_{m,l}(0) & \dots & h_{m,l}(L_h-1) & 0 & \dots & 0 \\ 0 & & & & & \\ \vdots & & Toeplitz & & & \\ 0 & & & & & \end{bmatrix}^T,$$

with dimensions $(L_h + L_g - 1) \times L_g$. Furthermore, a vector containing the non-zero samples of the cascade impulse response in all the control points is defined as

$$\mathbf{x} = \left[\mathbf{x}_0^T, \ \dots \ , \mathbf{x}_{M-1}^T\right]^T = \mathbf{H}\mathbf{g}, \tag{13}$$

where $\mathbf{g} = [\mathbf{g}_0^T, \dots, \mathbf{g}_{L-1}^T]^T$, and

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{0,0} & \dots & \mathbf{H}_{0,L-1} \\ \vdots & \vdots & \vdots \\ \mathbf{H}_{M-1,0} & \dots & \mathbf{H}_{M-1,L-1} \end{bmatrix}.$$

Similarly, the vector containing the target impulse response for all control points is given by

$$\mathbf{d} = \left[\mathbf{d}_0^T \ \dots \ \mathbf{d}_{M-1}^T\right]^T, \tag{14}$$

where

$$\mathbf{d}_m = \left[d_m(0) \ \dots \ d_m(L_h + L_g - 2)\right]^T, \tag{15}$$

and $d_m(n)$ is the target response for the m-th control point.

Once the model is presented, we describe the Weighted Pressure Matching algorithm with time-domain formulation (wPM-TD) proposed by [15]. This algorithm aims to find the optimal filters $\mathbf{g}$ that minimize the following cost function

$$J(\mathbf{g}) = \|\mathbf{W}(\mathbf{x} - \mathbf{d})\|^2 + \lambda \|\mathbf{g}\|^2, \tag{16}$$

where $\mathbf{d}$ is usually selected such that the target in the dark zone is zero, i.e., $\mathbf{d}_m = \mathbf{0} \ \forall m \in \mathcal{D}$. Also, λ is a regularization parameter that constraints the energy of the filters [26]. The weighting matrix $\mathbf{W}$ can include the effects of the time, frequency, and spatial weighting [18], but we focus in the case with spatial weighting only for the sake of simplicity. Then, we define $\mathbf{W}$ as a block diagonal matrix, i.e., $\mathbf{W} = \text{Diag}\{\mathbf{W}_0, \dots, \mathbf{W}_{M-1}\}$, formed by diagonal submatrices of size $L_h + L_g - 1$, which are defined as

$$\mathbf{W}_m = \begin{cases} \sqrt{\frac{\mu}{M_d}}\,\mathbf{I} & m \in \mathcal{D} \\ \sqrt{\frac{1-\mu}{M_b}}\,\mathbf{I} & m \in \mathcal{B} \end{cases},$$

where parameter μ is a weighting term that is used to balance the solution, e.g., high values of μ put more effort in minimizing the mean energy in the dark zone whereas low values of μ put more effort in minimizing the Mean Square Error (MSE) with respect to the target response in the bright zone. Then, (16) offers a balanced solution between minimizing the mean energy in the dark zone and the MSE in the bright zone. Using (13), we can reduce the cost function to

$$J(\mathbf{g}) = \mathbf{g}^T\mathbf{H}^T\mathbf{W}^T\mathbf{W}\mathbf{H}\mathbf{g} - 2\mathbf{g}^T\mathbf{H}^T\mathbf{W}^T\mathbf{W}\mathbf{d} + \lambda\mathbf{g}^T\mathbf{g}, \tag{17}$$

where the term $\mathbf{d}^T\mathbf{W}^T\mathbf{W}\mathbf{d}$ has been omitted because has no effect on the optimization. Finally, the optimal solution that minimizes (17) is

$$\mathbf{g} = \left(\mathbf{H}^T\mathbf{W}^T\mathbf{W}\mathbf{H} + \lambda\mathbf{I}\right)^{-1}\mathbf{H}^T\mathbf{W}^T\mathbf{W}\mathbf{d}. \tag{18}$$

Equation (18) presents two main drawbacks: 1) it requires the inversion of a matrix of size $(L{\cdot}L_g) \times (L{\cdot}L_g)$, which for some cases can be computationally unfeasible; 2) the optimization is broadband, so does not offer the versatility to select different configurations for different frequency bands (which in some cases could be beneficial). Next, we propose a novel algorithm that overcomes these limitations.

IV. Weighted Pressure Matching With Subband Decomposition

In this section, we propose to use a GDFT filter bank with K subbands to perform the filtering operation for each of the loudspeakers in the PSZ system. Then, rather than using a broadband filter g_l for each loudspeaker, we use a set of subband filters $g_l^{(k)}$, $0 \le k \le K-1$. Additionally, we propose a novel algorithm to compute the optimal subband filters. We show in Fig. 3(a) the system model used, where the filtering required by the PSZ system is performed in the subbands of the filter bank and the output of the filter bank is convolved with the RIR. It

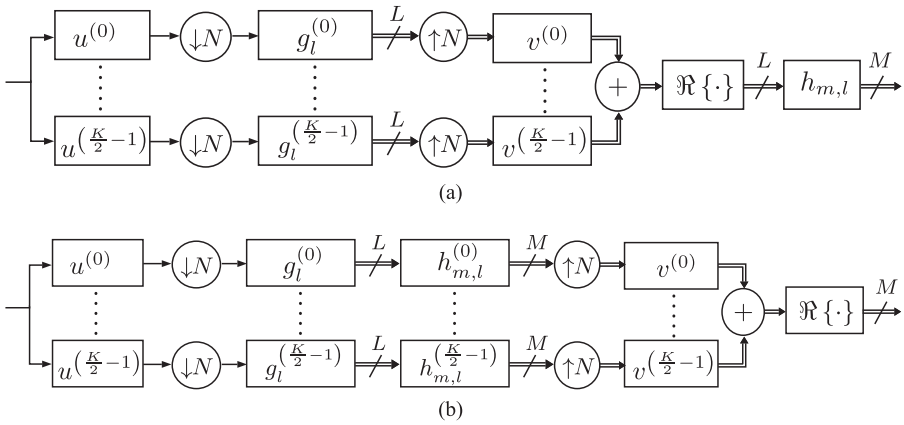

Fig. 3. (a) Model of a PSZ system using a GDFT filter bank. (b) Equivalent model using the subband components of the RIR.

is important to mention that the subband filter $g_l^{(k)}$ has complex coefficients (contrary to the broadband case, where g_l has real coefficients), but only $K/2$ subband filters must be computed for each loudspeaker due to the hermitian symmetry (assuming that K is even[2]).

Now, let us denote $h_{m,l}^{(k)}$ as the k-th subband component of $h_{m,l}$ obtained with the subband decomposition procedure described in Section II. Specifically, we need to replace the generic response q used in Section II by response $h_{m,l}$, and use (10) to obtain the subband component $h_{m,l}^{(k)}$ of length $L'_h = \lceil (L_h + L_p - 1)/N \rceil - \lceil L_p/N \rceil + 1$. Then, the authors in [24] show that the system models in Fig. 3(a) and (b) are approximately equivalent if conditions (3) and (4) are fulfilled. The main advantage of the system in Fig. 3(b) is that the subsystems for each subband can be considered independent, so the optimization problem can be solved independently for each subband. This fact is important because not only provides computational improvements to solve the optimization, but also offers higher versatility than the broadband case (18) because it allows us to use different configurations for each subband. For example, different sets of loudspeakers and different filter lengths can be used in each subband.

First, we start by formulating the time-domain model for each subband of Fig. 3(b). For the sake of simplicity, we assume in this section that all subbands use the same set of loudspeakers, and that all the subband components $g_l^{(k)}$ have the same length (denoted by L'_g). Then, the cascade impulse response for the k-th subband is defined similarly to (13) as

$$\mathbf{x}^{(k)} = \mathbf{H}^{(k)} \mathbf{g}^{(k)}, \tag{19}$$

where $\mathbf{g}^{(k)} = [(\mathbf{g}_0^{(k)})^T, \ldots, (\mathbf{g}_{L-1}^{(k)})^T]^T$ with $\mathbf{g}_l^{(k)} = [g_l^{(k)}(0), \ldots, g_l^{(k)}(L'_g - 1)]^T$, and

$$\mathbf{H}^{(k)} = \begin{bmatrix} \mathbf{H}_{0,0}^{(k)} & \cdots & \mathbf{H}_{0,L-1}^{(k)} \\ \vdots & \vdots & \vdots \\ \mathbf{H}_{M-1,0}^{(k)} & \cdots & \mathbf{H}_{M-1,L-1}^{(k)} \end{bmatrix},$$

[2]For the odd case $((K-1)/2) + 1$ subband filters must be computed.

with

$$\mathbf{H}_{m,l}^{(k)} = \begin{bmatrix} h_{m,l}^{(k)}(0) & \cdots & h_{m,l}^{(k)}(L'_h - 1) & 0 & \cdots & 0 \\ 0 & & & & & \\ \vdots & & Toeplitz & & & \\ 0 & & & & & \end{bmatrix}^T.$$

Similarly, we can define the target impulse response for the k-th subband as

$$\mathbf{d}^{(k)} = \left[(\mathbf{d}_0^{(k)})^T \quad \cdots \quad (\mathbf{d}_{M-1}^{(k)})^T \right]^T, \tag{20}$$

where

$$\mathbf{d}_m = \left[d_m^{(k)}(0) \quad \cdots \quad d_m^{(k)}(L'_h + L'_g - 2) \right]^T, \tag{21}$$

and $d_m^{(k)}$ is the k-th subband component of d_m, which is obtained with the subband decomposition described in Section II.

Now, we propose the Weighted Pressure Matching with Subband Decompositon (wPM-SD) algorithm to obtain the optimal subband filters $g_l^{(k)}$. For each subband k, the optimal coefficients $\mathbf{g}^{(k)}$ minimize the following cost function

$$J(\mathbf{g}^{(k)}) = \|\mathbf{W}(\mathbf{x}^{(k)} - \mathbf{d}^{(k)})\|^2 + \lambda \|\mathbf{g}^{(k)}\|^2. \tag{22}$$

Similarly to the wPM-TD case, the weighting matrix $\mathbf{W}$ can include the effects of the time, frequency, and spatial weighting, but only the spatial weighting is considered in the following. Then, we define $\mathbf{W}$ as a block diagonal matrix, i.e., $\mathbf{W} = diag\{\mathbf{W}_0, \ldots, \mathbf{W}_{M-1}\}$, formed by diagonal submatrices of size $L'_h + L'_g - 1$, which are defined as

$$\mathbf{W}_m = \begin{cases} \sqrt{\frac{\mu}{M_d}}\, \mathbf{I} & m \in \mathcal{D} \\ \sqrt{\frac{1-\mu}{M_b}}\, \mathbf{I} & m \in \mathcal{B} \end{cases}.$$

Finally, the optimal solution for the k-th subband is given by

$$\mathbf{g}^{(k)} = (\mathbf{H}^{(k)H} \mathbf{W}^T \mathbf{W} \mathbf{H}^{(k)} + \lambda \mathbf{I})^{-1} \mathbf{H}^{(k)H} \mathbf{W}^T \mathbf{W} \mathbf{d}^{(k)}. \tag{23}$$

Solution (23) requires the inversion of a matrix of size $(L \cdot L'_g) \times (L \cdot L'_g)$, which is approximately N times smaller than the matrix that needs to be inverted with wPM-TD, leading to important computational savings (as we will show in Section V-B). It is important to mention that wPM-SD requires the subband

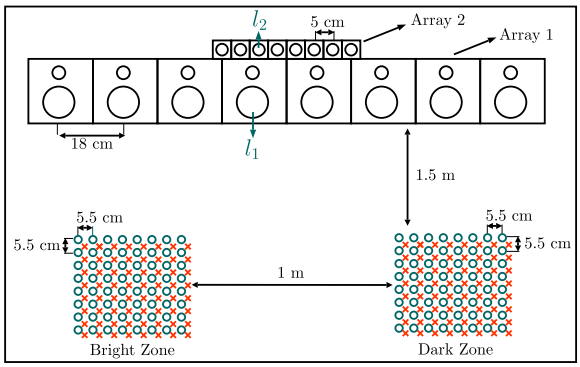

Fig. 4. Layout used for the experimental validation, where ○ and denote × control and validation points, respectively.

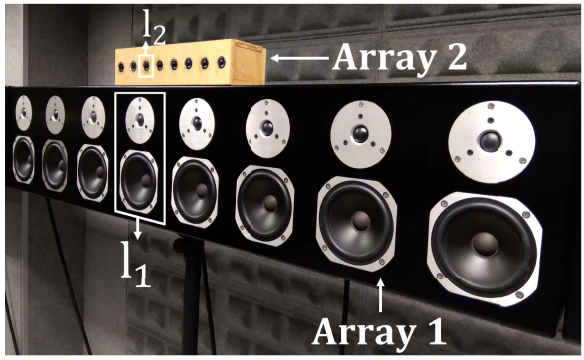

Fig. 5. Arrays of loudspeakers used for the experimental validation.

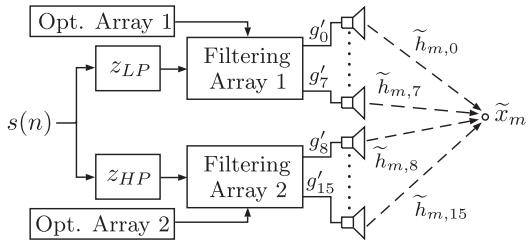

Fig. 6. Block diagram of the signal processing used in the PSZ system.

decomposition of the RIR using (10), which could require additional computational load. However, only one matrix inversion is required to compute the subband components for all the microphones, loudspeakers, and subbands, so its computational cost is not significant. Moreover, although we have assumed that the same weighting matrix $\mathbf{W}$ and the same regularization factor λ are used for all subbands, different values could be selected for each subband. Also, different number of loudspeakers and filter lengths can be used for each subband. Finally, we have assumed that the same algorithm is used to obtain the optimal filters for all subbands (in our case wPM), however, the subband formulation allows us to use different algorithms for different subbands (as for example ACC, which could be beneficial for very low frequencies [27]). In conclusion, wPM-SD can not only offer a reduction of the computational cost with respect to wPM-TD, but also greater versatility, as it allows to use different setups for different subbands.

V. Results

In this section, the proposed algorithm is evaluated and compared with wPM-FD and wPM-TD. First, we define the experimental setup and the metrics used for the evaluations. Next, the algorithms are compared using the defined metrics. Finally, we present two practical examples showing how to benefit from the versatility of wPM-SD.

A. Setup and Metrics

The experimental evaluations have been carried out in the laboratory of Audio and Communications Signal Processing of the Institute of Telecommunications and Multimedia Applications (iTEAM). The room presents a reverberation time of $T_{60} = 0.18$ s and its dimensions can be found in [28]. Fig. 4 shows the layout used for the evaluations, which includes two arrays of speakers (shown in Fig. 5). Array 1 is formed by 8 two-way speakers with frequency range 100–20000 Hz and an inter-element distance of 18 cm. Therefore, the highest frequency that can be rendered without spatial aliasing is around 1.9 kHz [29], which is too low for most applications. Array 2 is formed by 8 speakers with frequency range 800–20000 Hz and an inter-element distance of 5 cm. Thus, its highest aliasing-free frequency is around 5 kHz, but the size of its speakers and its aperture limits its performance for low frequencies. Thus, using both arrays can assure a frequency range of 100–5000 Hz without aliasing and with good performance. The system considers two zones, where the left and right zones are the bright and dark zones, respectively (as shown in Fig. 4). The zones are spatially sampled by using a grid of 8×8 microphones with an inter-element distance of 5.5 cm. Different sets of RIRs are used to compute and to evaluate the filters, i.e., the RIRs in the control points ($h_{m,l}$) and in the validation points ($\widetilde{h}_{m,l}$). The RIRs were measured using the swept-sine technique [30] with a sampling frequency of 44100 Hz, and then downsampled to 14700 Hz. In all cases, equal effort is used to minimize the error in the bright zone and the energy in the dark zone, i.e., $\mu = 0.5$.

The block diagram of the system used for the evaluations is shown in Fig. 6. The diagram includes the FIR cross-over filters z_{LP} and z_{HP} with cut-off frequency $f_c = 1.9$ kHz, length 61 samples, and delay $\tau_c = 30$ samples (which have been designed using Matlab's *fir1()* function with a Kaiser window with shape factor 1.7). These filters avoid driving the speakers at frequencies where either spatial aliasing or inefficient radiation occurs. Thus, array 1 operates at frequencies below f_c, and array 2 at frequencies above f_c. Then, we can compute the optimal filters for each array independently. We use $h_{m,l}$ for $l = 0, \ldots, 7$ in the optimization for array 1, and $h_{m,l}$ for $l = 8, \ldots, 15$ for array 2. The target response in the optimization for array i is selected as

$$d_m(n) = \begin{cases} h_{m,l_i}(n - \tau_d) & m \in \mathcal{B} \\ 0 & m \in \mathcal{D} \end{cases}, \tag{24}$$

i.e., null response for the dark zone, and the RIR of the l_i-th loudspeaker for the bright zone, where l_i is the reference loudspeaker for array i (with $i = 1, 2$). Specifically, l_1 and l_2 are the indices of the fourth loudspeaker in array 1 and the third loudspeaker in array 2, respectively (as shown in Fig. 5). For

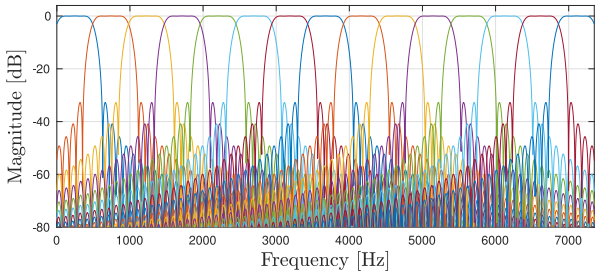

Fig. 7. Magnitude response of the analysis filters in the positive spectrum for the selected GDFT filter bank.

wPM-TD and wPM-SD, the target includes a modelling delay of $\tau_d = 350$ samples (i.e., 23.8 ms). For wPM-FD, a modeling delay of $\tau_d = L_g/2$ is required to assure causality. The optimal filters for array i are applied to the input signal in the "Filtering Array i" block in Fig. 6, which performs broadband filtering for wPM-TD and wPM-FD, and subband filtering for wPM-SD. It is important to mention that both arrays could be optimized together in the band 1–1.9 kHz (as the aperture of array 2 is long enough to operate in such frequencies). However, we decided not to use array 2 in this range because we observed that it requires high effort levels, which could damage its speakers.

For wPM-SD, we use a GDFT filter bank with $K = 30$ subbands, resampling factor $N = 22$, and prototype filter of length $L_p = 150$ (obtained using the method proposed in [31] with $\gamma = 0.005$). The magnitude of the analysis filters in the positive spectrum is shown in Fig. 7. The filter bank used presents a Reconstruction Error [23] of -77.04 dB and a Signal-To-Aliasing Ratio [23] of 46.45 dB. Then, we can assume that conditions (3) and (4) are fulfilled. The selected filter bank produces a delay of $\tau_f = 149$ samples.

For wPM-FD, the optimal filter coefficients are computed for a set of $L_h + L_g - 1$ control frequencies using the optimization proposed in [12]. After that, an IDFT of size $L_h + L_g - 1$ is used to obtain the time-domain responses, which are truncated using a Hanning window of size L_g to obtain the filters (similarly to the procedure described in [32]).

It is worth mentioning that we do not solve the optimization for all control frequencies/subbands for wPM-FD/wPM-SD. For wPM-FD, the optimization is only solved for frequencies below f_c for array 1, and above f_c for array 2. For wPM-SD, it is only solved for subbands 0-3 for array 1, and for subbands 4-14 for array 2. In the other control frequencies/subbands, the filter coefficients are set to 0. Thus, we avoid driving the arrays at frequencies outside their operation bandwidth. Then, the cross-over filters are not really needed for wPM-SD and wPM-FD, as the sound zones filters already attenuate the energy at frequencies outside the operation bandwidth of the arrays. This is an advantage of wPM-SD and wPM-FD with respect to wPM-TD for PSZ systems using multiple arrays, as the cross-over filters add further delay to the system. However, in the following we use the cross-over filters for evaluating all the algorithms in order to make a fair comparison between them (as we assure that the performance differences are not caused by the effects of the cross-over filters).

Next, we present the metrics used for evaluating the algorithms. First, the mean energy in the validation points of the bright and dark zones [25] at frequency f is

$$P_b(f) = \frac{1}{M_b} \sum_{m \in \mathcal{B}} \left| \widetilde{X}_m(f) \right|^2, \tag{25}$$

$$P_d(f) = \frac{1}{M_d} \sum_{m \in \mathcal{D}} \left| \widetilde{X}_m(f) \right|^2, \tag{26}$$

respectively, where $\widetilde{X}_m$ is the Discrete Time Fourier Transform (DTFT) of the cascade impulse response for the m-th validation point ($\widetilde{x}_m$). The cascade impulse response includes the contributions of both arrays and it is obtained by feeding the system in Fig. 6 with $\delta(n)$. The second metric is the acoustic contrast (AC) [3], which is related to the level of acoustic isolation between zones and it is defined as

$$C(f) = \frac{P_b(f)}{P_d(f)}. \tag{27}$$

The third metric is the Normalized Mean Square Error (NMSE) for the validation points in the bright zone [11], which is defined as

$$\epsilon(f) = \frac{\sum_{m \in \mathcal{B}} \left| \widetilde{X}_m(f) - \widetilde{D}_m(f) \right|^2}{\sum_{m \in \mathcal{B}} \left| \widetilde{D}_m(f) \right|^2}, \tag{28}$$

where $\widetilde{D}_m(f)$ is the target transfer function for the m-th validation point, and it is defined as

$$\widetilde{D}_m(f) = \begin{cases} e^{-j2\pi\tau f} \widetilde{H}_{m,l_1}(f) \text{ if } f \leq f_c \\ e^{-j2\pi\tau f} \widetilde{H}_{m,l_2}(f) \text{ if } f > f_c \end{cases}, \tag{29}$$

in which $\widetilde{H}_{m,l}(f)$ is the DTFT of $\widetilde{h}_{m,l}$, and l_1 and l_2 are the indices of the reference speakers for array 1 and 2, respectively. Also, a term that considers the delay is included in (29). For wPM-TD and wPM-FD, this delay includes the modelling delay used in the optimization and the delay of the cross-over filters, i.e., $\tau = \tau_c + \tau_d$. For wPM-SD, the delay of the filter bank is also considered, i.e., $\tau = \tau_c + \tau_d + \tau_f$. The last metric is the Array Effort (AE) [26], which is defined as

$$AE(f) = \frac{\sum_{l=0}^{15} |G'_l(f)|^2}{E_r(f)}, \tag{30}$$

where $G'_l(f)$ is the DTFT of the signal fed to loudspeaker l when driving the system in Fig. 6 with $\delta(n)$, and $E_r(f)$ is the energy required by the reference loudspeaker to produce the same energy in the bright zone as the PSZ system. Specifically, l_1 is the reference loudspeaker for $f \leq f_c$, and l_2 for $f > f_c$. To improve the readability of the results, we use a third-octave averaging [33] for all frequency-domain plots.

B. Comparison of wPM-TD, wPM-FD, and wPM-SD

Next, we compare wPM-TD, wPM-FD, and wPM-SD using the setup and the metrics presented in Section V-A. For wPM-TD and wPM-SD, we use $\lambda = 2{\cdot}10^{-3}$ and $\lambda = 2{\cdot}10^{-4}$ for arrays 1 and 2, respectively. This selection produces an array effort lower than 2 dB in the range 100–5000 Hz. For wPM-FD, a search has been carried out to find the regularization factor for each control frequency that leads to the same array effort as wPM-TD. In the following, we assume that when L_g is selected as the filter length

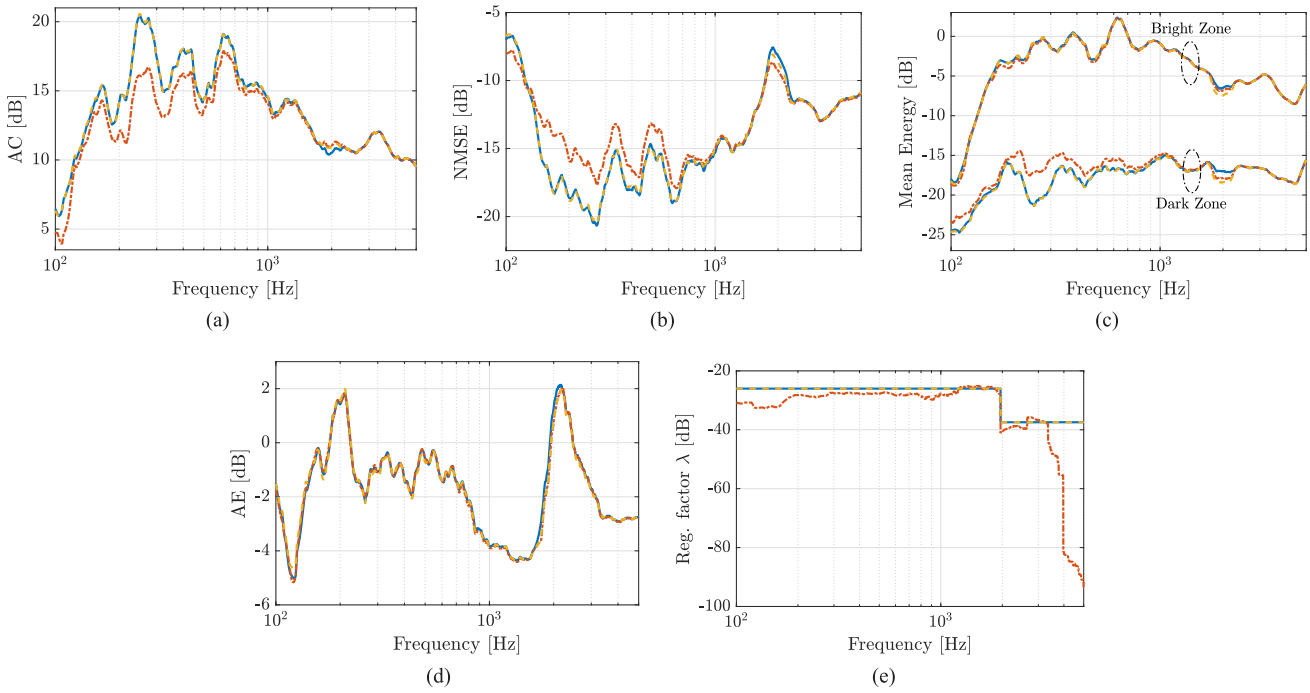

Fig. 8. Comparison of the performance of wPM-TD (——), wPM-FD (-·-·-), and wPM-SD (- - -) in terms of: (a) acoustic contrast, (b) normalized mean square error, (c) mean energy of the cascade transfer function for the bright and dark zones, (d) array effort, and (e) regularization factor.

for the broadband algorithms (i.e., wPM-TD and wPM-FD), the following relation is used to obtain the length of the subband filters for wPM-SD:

$$L_g' = \left\lceil \frac{L_g + L_p - 1}{N} \right\rceil - \left\lceil \frac{L_p}{N} \right\rceil + 1. \tag{31}$$

This selection is motivated by the fact that the length of the subband components of a FIR signal with length L_g is given by (31), as studied in Section II.

We start by comparing the algorithms with $L_g = 1500$ for wPM-FD and wPM-TD, and $L_g' = 69$ for wPM-SD. The Acoustic Contrast, the NMSE, the mean energy in each zone, the array effort, and the regularization factor are shown in Fig. 8 as a function of frequency. First, we would like to highlight that the array effort (Fig. 8(d)) is almost identical for all the algorithms. This means that the differences between the algorithms in terms of acoustic contrast or NMSE are not achieved at the cost of higher effort, and then, the comparison between algorithms is fair. From Fig. 8, it is clear that the performance of the algorithms is equal for frequencies above 2.2 kHz (where array 2 operates). The explanation is that long filters are not required in these frequencies because the reverberation time of the room in these frequencies is low (as we will see in Section V-D). Then, the effective length of the filters in these frequencies is shorter than the window used with wPM-FD, and so, the truncation does not degrade the performance. At frequencies near the cut-off frequency (i.e., 1.9 kHz), wPM-FD an wPM-SD perform better than wPM-TD, being wPM-FD slightly superior than wPM-SD. This phenomenon occurs because for wPM-FD and wPM-SD the filter coefficients are set to 0 in the control frequencies/subbands outside the frequency range of the arrays, an consequently, the interference between arrays at frequencies around 1.9 kHz is lower than with wPM-TD. The performance of wPM-TD in these frequencies could be improved by increasing the length of the cross-over filters. Also, we can see that wPM-TD and wPM-SD offer a better performance than wPM-FD for frequencies below 1 kHz. At certain frequencies, wPM-FD offers 4 dB lower acoustic contrast and 3 dB higher NMSE than wPM-TD and wPM-SD.

Now, we further study the performance in the range 100–1000 Hz for array 1 (where greater differences were found in Fig. 8). In particular, we study the influence of the filter length in the performance of each algorithm. We plot in Fig. 9 the average acoustic contrast, NMSE, and array effort in the band 100–1000 Hz as a function of the filter length. Also, we plot in Fig. 9(d) the delay of the system. For wPM-TD and wPM-SD, the performance converges at $L_g = 1500$, where an average acoustic contrast of 16.1 dB and NMSE of -16.5 dB are achieved. For wPM-FD, the performance converges to the same values, but $L_g = 3000$ is required to achieve the convergence (twice the length needed for wPM-TD and wPM-SD). Then, we can see that the performance of the studied algorithms is nearly the same if long enough filters are used. However, wPM-TD and wPM-SD can achieve good performance even if shorter filters are selected. Moreover, we can see in Fig. 9(d) that the delay for wPM-TD and wPM-SD is constant, while it increases with the filter length for wPM-FD. Then, wPM-TD and wPM-SD not only can obtain good performance with shorter filters than wPM-FD, but also with shorter delays. We can see from Fig. 8 that wPM-TD and wPM-SD are nearly identical in terms of acoustic contrast, NMSE, and array effort, but wPM-SD requires higher delay. Specifically, we have a delay $\tau = 36.0$ ms for wPM-SD and $\tau = 25.9$ ms for wPM-TD. Still, the delay for wPM-SD is significantly lower than for wPM-FD, which is $\tau = 53.1$ ms for $L_g = 1500$ and $\tau = 104.1$ ms for $L_g = 3000$.

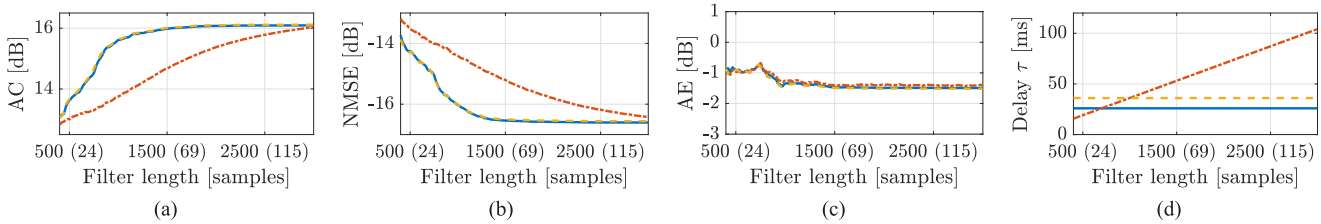

Fig. 9. Comparison of the average performance of wPM-TD (——), wPM-FD (-·-·-), and wPM-SD (- - -) in the frequency range 100–1000 Hz as a function of the filter length in terms of: (a) acoustic contrast, (b) normalized mean square error, (c) array effort, and (d) system delay. In the x-axis, the number within the brackets indicates the length of the equivalent subband filters (obtained with (31)).

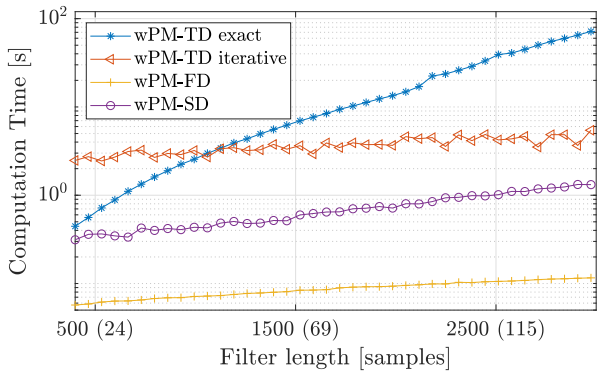

Fig. 10. Measured computation time for different algorithms as a function of the filter length. In the x-axis, the number within the brackets indicates the length of the equivalent subband filters for wPM-SD (obtained with (31)).

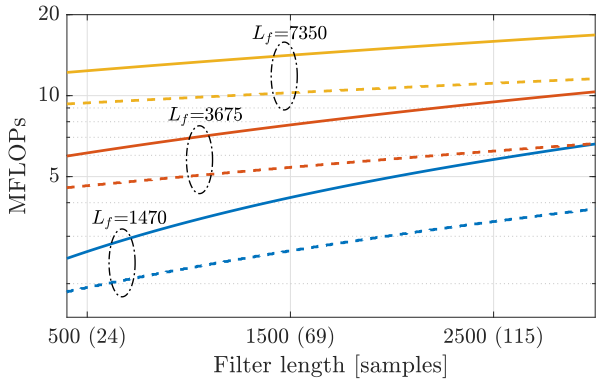

Fig. 11. Mega FLOPs required for broadband filtering (——), and subband filtering (- - -) as a function of the filter length for three different frame lengths (L_f). In the x-axis, the number within the brackets indicates the length of the equivalent subband filters (obtained with (31)).

Next, we compare the computational demands of the algorithms. For the comparison, we consider a single array (i.e., 8 speakers) where the optimization and the filtering are carried out for all control frequencies/subbands for wPM-FD/wPM-SD. We start by comparing the complexity of each algorithm for obtaining the optimal filters. Two different approaches have been used for wPM-TD: 1) computing the exact solution (18); 2) using the iterative algorithm proposed in [18] with $\gamma_0 = 10^{-8}$, $\gamma' = 10$, $s_{th} = 0.1$, and 300 iterations. For the iterative algorithm [18], an exact number of operations can not be calculated because each iteration requires different computations depending on the scenario. Then, we measure the computation time of the algorithms running in Matlab R2018a in an Intel Core i7-7700 processor at 3.60 GHz. The measurement has been computed as the mean for 150 repetitions. For wPM-TD with exact solution, wPM-FD, and wPM-SD, the function *linsolve()* has been used to obtain the optimal solutions. The iterative algorithm in [18] has been implemented using C language, and ran from Matlab using a *mex* function. The main reason is that iterative algorithms are computationally inefficient in Matlab [34]. We show in Fig. 10 the computation time of each algorithm as a function of the filter length. The results for wPM-SD include the computation time required to obtain the subband components of the RIR. From these results, we can see how obtaining the exact wPM-TD solution is computationally very demanding, while wPM-FD requires very low computation. The iterative algorithm [18] effectively reduces the complexity for filter lengths longer than 1000 samples. However, its computation time is still significantly higher than the one for wPM-SD, which has lower computation time than the exact wPM-TD solution for all the studied filter lengths. As we discussed previously, the computational demands for solving the optimization are not a key factor for static PSZ systems, because the filters are computed offline. However, the lower computational complexity of wPM-SD can be beneficial in dynamic systems, where the filters are recomputed online to take into account the mobility of the zones or the changing statistics of the input signal.

A key aspect of all PSZ systems is the computational complexity needed for filtering the audio signals, as it requires real-time performance. Next, we compare the computational demands of broadband filtering (for wPM-TD and wPM-FD) and subband filtering with a GDFT filter bank (for wPM-SD). In Fig. 11, we show the number of Floating Point Operations (FLOPs) required to filter signals of 1470, 3675, and 7350 samples (i.e., frames of 100, 250 and 500 ms) as a function of the filter length. We considered an efficient broadband filtering using the Fast Fourier Transform (FFT), and the efficient polyphase implementation of the GDFT filter bank [23]. The results show that subband filtering requires significantly lower number of FLOPS than the equivalent broadband filtering. Thus, wPM-SD requires lower computational demands than wPM-TD and wPM-FD. For example, wPM-SD with $L'_g = 69$ (with performance nearly equal to wPM-TD with $L_g = 1500$) requires lower number of FLOPs than wPM-TD with $L_g = 500$. Then, wPM-SD can use longer filters than wPM-FD and wPM-TD with lower computational complexity. This is a great advantage for PSZ systems, as it

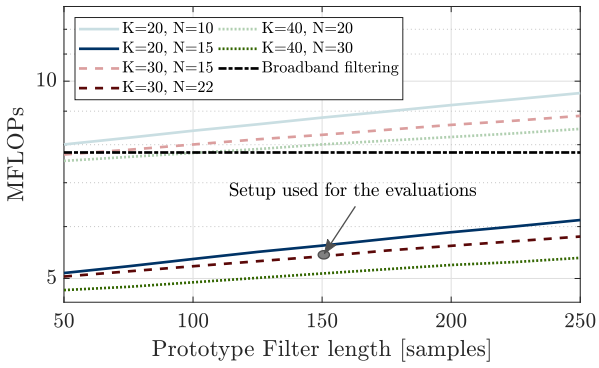

Fig. 12. Mega FLOPs required to perform subband filtering of a signal of 3675 samples as a funtion of the prototype filter length for different filter bank configurations. For each configuration, the length of the subband filters is obtained with (31) (assuming that $L_g = 1500$). Also, the Mega FLOPS required for the equivalent broadband filtering with $L_g = 1500$ are plotted.

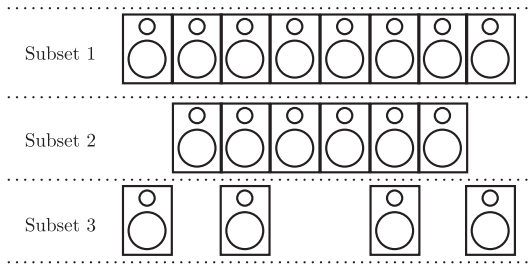

Fig. 13. Different subsets of loudspeakers for array 1.

TABLE I
INDICES OF THE SUBBANDS WHERE EACH SUBSET OF SPEAKERS IS USED FOR EACH CONFIGURATION

	Config. 1	**Config. 2**	**Config. 3**	**Config. 4**
Subset 1	0 – 3	–	–	–
Subset 2	–	0 – 3	–	1 – 3
Subset 3	–	–	0 – 3	0

allows to use very long filters with affordable computational demands.

Actually, the computational benefits of subband filtering depend on the selected filter bank configuration. To study this dependence, we show in Fig. 12 the number of FLOPs required by different configurations to filter a signal of 3675 samples. For each configuration, the length of the subband filters L'_g is obtained with (31) assuming that $L_g = 1500$. Results for $K = 20$, 30, and 40 subbands are included. For each of these values, two different resampling factors (N) are used, offering a fractional oversampling (N/K) of approximately 0.5 and 0.75. Also, the number of FLOPs required for broadband filtering with $L_g = 1500$ are included for comparison purposes. First, we can see that a fractional oversampling of 0.5 does not offer computational benefits with respect to broadband filtering. However, increasing the fractional oversampling to 0.75 significantly reduces the computational complexity. The main reason is that, for the same broadband filter length, a fractional oversampling factor of 0.75 requires shorter subband filters than 0.5 (because higher fractional oversampling factor implies higher N, and so, smaller L'_g, as it can be seen in (31)). Also, we can see that increasing the number of subbands reduces the computational complexity when using the same fractional oversampling and prototype filter length. Nevertheless, it is important to note that the higher the number of subbands, the longer the prototype filter required to fulfil condition (4) (leading to increased computational demands). Then, a compromise between number of subbands, resampling factor, and prototype filter length is required to obtain computation reductions. The selected configuration in Section V-A with $K = 30$, $N = 22$, and $L_p = 150$ seems a good choice, as leads to good performance with wPM-SD and significantly lower computational demands than broadband filtering.

From the presented results, we can conclude that the studied algorithms can offer similar performance, but wPM-TD and wPM-SD require shorter filters and delays than wPM-FD. Also, we showed that the performance of wPM-SD and wPM-TD is almost identical. However, wPM-SD leads to higher delay than wPM-TD in benefit of lower computational demands (both for solving the optimization and for filtering the input signals).

C. wPM-SD with Different Loudspeaker Setup in Each Subband

As we mentioned in Section IV, the versatility of wPM-SD allows us to use different loudspeaker sets in each subband. Next, we consider 3 different subsets of speakers for array 1 (shown in Fig. 13), where subsets 1, 2, and 3 are formed by 8, 6 and 4 speakers, respectively. We evaluate the performance of four different configurations for wPM-SD (shown in Table I), where we limit the study to the range 100–1900 Hz (i.e., subbands 0–3). Specifically, configurations 1, 2, and 3 use subsets 1, 2, and 3 in all subbands, respectively. Finally, configuration 4 is a subband-dependent configuration where subset 2 is used in subbands 1–3 and subset 3 in subband 0.

In Fig. 14, we compare the performance of the four configurations in Table I with $L'_g = 69$. We used $\lambda = 2{\cdot}10^{-3}$ for configuration 1, and searched for the regularization factors that lead to approximately equal filter energy than configuration 1 for the other configurations. From the results, it is obvious that the best performance is achieved for configuration 1, where all the available speakers in the array are used. The performance of configuration 2 is quite similar for frequencies above 500 Hz, with differences lower than 1 dB for the acoustic contrast and NMSE with respect to configuration 1. However, this configuration offers a noticeable worst performance for frequencies below 500 Hz. Probably, this degradation is produced because of the shorter aperture of the array for configuration 2. On the contrary, configuration 3 offers similar performance to configuration 1 for frequencies below 500 Hz (even though only 4 speakers are used) but very bad performance for higher frequencies. This bad performance is produced by the spatial aliasing, which appears at lower frequencies for configuration 3 due to its higher inter-element distance. Finally, it can be seen how using configuration 4 (with 4 speakers in subband 0 and 6 speakers in subbands 1–3) we can achieve a performance which is quite close to configuration 1. Moreover, reducing the number of speakers used in each subband allows us to further reduce the computational demands. Concretely, configuration 4 requires 31.25% less filter taps than configuration 1 for subbands 0–3. Then, selecting different loudspeaker sets for different subbands

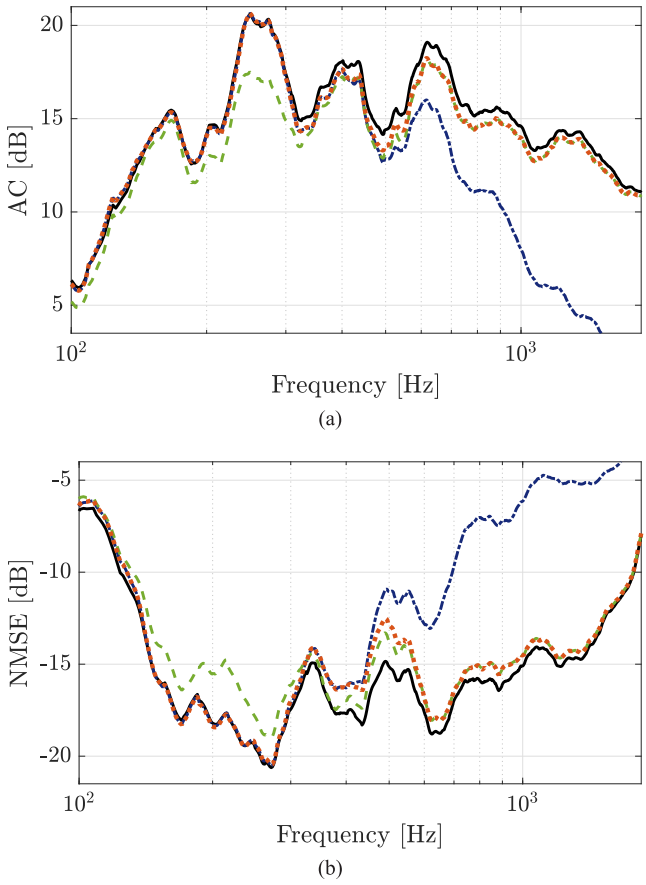

Fig. 14. Performance comparison for wPM-SD with loudspeaker configurations 1 (——), 2 (– – –), 3 (–·–·–), and 4 (·····) in terms of: (a) acoustic contrast, and (b) normalized mean square error.

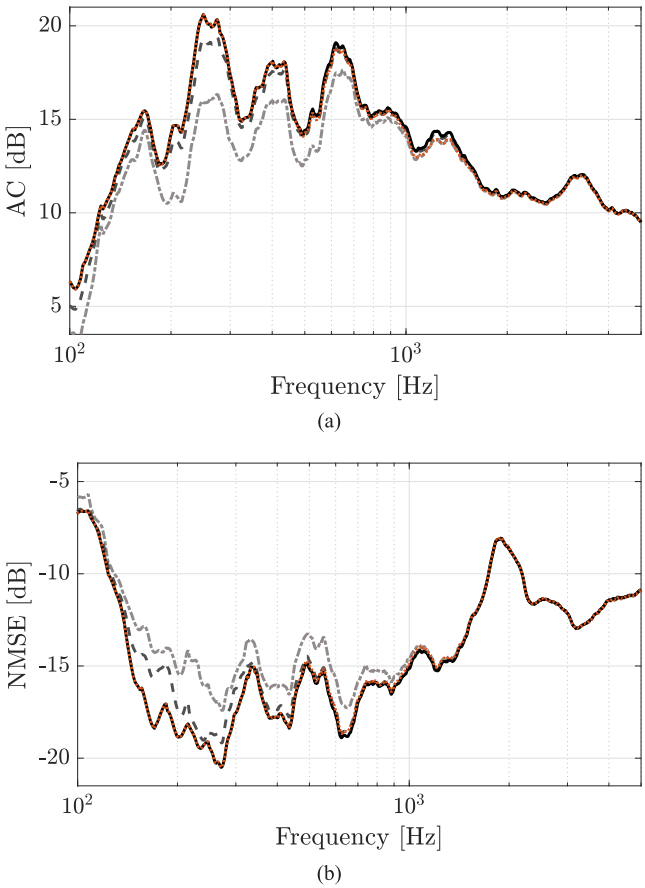

Fig. 15. Performance comparison for wPM-SD with $L'_g = 34$ (·····), $L'_g = 47$ (– – –), $L'_g = 69$ (——), and subband-dependent length (·····) in terms of: (a) acoustic contrast, and (b) normalized mean square error.

can be further used to reduce the computational demands in cases where the computational resources are limited (at the cost of worsening a little the performance in certain frequency bands). Thus, we can conclude that, according to the desired performance and to the available processing capabilities, the flexibility of wPM-SD allows to select the most convenient loudspeaker set for each subband.

D. wPM-SD with Different Filter Length in Each Subband

In PSZ systems, the length of the filters is related with the reverberation time of the environment, which is frequency dependent (being usually higher for lower frequencies [35]). For wPM-TD and wPM-FD, the frequencies with higher reverberation time determine the length of the filter that must be used for all the operation bandwidth. However, for wPM-SD the length of the filters can be selected for each subband according to the reverberation properties in that subband.

Next, we compare in Fig. 15 the performance of four different configurations for wPM-SD. We include three configurations with the same filter length for all subbands, with $L'_g = 34, 47, 69$. Also, we show the results for a configuration with subband-dependent filter length, whose details are shown in Table II. For the configuration with $L'_g = 69$, we used $\lambda = 2{\cdot}10^{-3}$ for subbands 0–3 (i.e., for array 1) and $\lambda = 2{\cdot}10^{-4}$ for subbands 4–14 (i.e., for array 2). For the other configurations, we searched for the regularization factors that lead to approximately equal filter energy than with $L'_g = 69$. From Fig. 15, we can notice that the best acoustic contrast and NMSE are achieved with $L'_g = 69$, but $L'_g = 34, 47$ offer identical performance above 1 kHz. This fact indicates that long filters are not required above 1 kHz for the studied scenario. For frequencies between 500 Hz and 1 kHz, $L'_g = 34$ offers worse performance than $L'_g = 69$, but $L'_g = 47$ leads to very similar results. Then, there is not benefit from selecting a higher value than $L'_g = 47$ for this frequency range. Also, the results show that long filters (as $L'_g = 69$) are beneficial for frequencies below 500 Hz. Finally, we can see that the subband-dependent configuration offers a performance approximately equal to $L'_g = 69$ in the range 100–5000 Hz.

TABLE II
FILTER LENGTH FOR EACH SUBBAND FOR THE SUBBAND-DEPENDENT CONFIGURATION

	L'_g=69	L'_g=47	L'_g=34
Subband indices	0	1	2-14

At this point, we study how wPM-SD with subband-dependent filter length can reduce the computational demands. We show in Table III the percentage of FLOPs required for filtering the input signals using the configuration in Table II with respect to subband filtering with $L'_g = 69$ and broadband filtering with $L_g = 1500$. For an input signal of 1470 samples, the

TABLE III
PERCENTAGE OF FLOPS REQUIRED BY SUBBAND FILTERING WITH SUBBAND-DEPENDENT FILTER LENGTH WITH RESPECT TO SUBBAND FILTERING WITH $L'_g = 69$ AND BROADBAND FILTERING WITH $L_g = 1500$

	L_f=1470	L_f=3675	L_f=7350
Subband filtering, L'_g=69	80.93 %	89.77 %	94.15 %
Broadband filtering, L_g=1500	51.14 %	62.32 %	68.22 %

computational savings of the subband-dependent configuration are 19.07% and 48.86% with respect to subband filtering with $L'_g = 69$ and broadband filtering with $L_g = 1500$, respectively. It is worth mentioning that these improvements are not obtained at the cost of worse performance, as from the results presented in this section and in Section V-B we know that wPM-SD with subband-dependent length, wPM-SD with $L'_g = 69$, and wPM-TD with $L_g = 1500$ offer almost identical performance. Then, the versatility of wPM-SD to select the filter length for each subband according to its reverberation properties can be used to further reduce the computational complexity of a PSZ system, with negligible performance degradation.

VI. CONCLUSION

In this article, we proposed the novel wPM-SD algorithm to obtain the optimal subband filters for a PSZ system using a Generalized Discrete Fourier Transform filter bank. The algorithm is based on formulating an independent time-domain optimization problem for each subband of the filter bank. To do so, the algorithm makes use of the subband decomposition algorithm proposed in [24], which is further optimized in this article to reduce the required number of matrix inversions. We presented experimental evaluations that confirm that wPM-SD can perform approximately equal to wPM-TD, but with a small increase in the system delay (produced by the filter bank). However, the computational efforts required for computing the optimal filters are significantly lower with wPM-SD. Also, the evaluation results showed that wPM-SD and wPM-TD require shorter filters and delays than wPM-FD. Moreover, one of the advantages of wPM-SD with respect to wPM-TD and wPM-FD is that it requires substantially lower computational complexity to filter the input audio signals. Then, wPM-SD is preferred over wPM-TD and wPM-FD for practical PSZ systems, where usually limited computation resources are available and long filters are needed. In addition, we showed that formulating and independent optimization problem for each subband leads to great versatility for a PSZ system, as different configurations can be used for each subband. First, we presented experimental results where wPM-SD is used with different sets of loudspeakers for each subband. The results showed that reducing the number of loudspeakers in certain subbands can be useful to further reduce the computational complexity with only a slight worsening of the performance. Finally, the versatility offered by wPM-SD was further validated with experimental results where subband-dependent filter length is used. The results showed that the performance is not degraded if short filters are used for the subbands with short reverberation time. Then, wPM-SD with a proper selection of the filter lengths for the different subbands can contribute to reduce even more the computational complexity of the PSZ system without performance degradation.

REFERENCES

[1] W. F. Druyvesteyn and R. M. Aarts, "Personal sound," *The J. Acoust. Soc. Amer.*, vol. 45, no. 9, pp. 685–701, 1997.

[2] T. Betlehem *et al.*, "Personal sound zones: Delivering interface-free audio to multiple listeners," *IEEE Signal Process. Mag.*, vol. 32, no. 2, pp. 81–91, Mar. 2015.

[3] J.-W. Choi and Y.-H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," *The J. Acoust. Soc. Amer.*, vol. 111, no. 4, pp. 1695–1700, 2002.

[4] B. Van Veen and K. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[5] E. Mabande and W. Kellermann, "Towards superdirective beamforming with loudspeaker arrays," in *Proc. 19th Int. Congr. Acoust.*, 2007.

[6] T. Betlehem and P. D. Teal, "A constrained optimization approach for multi-zone surround sound," *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2011.

[7] Y. J. Wu and T. Abhayapala, "Spatial multizone soundfield reproduction: Theory and design," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no. 6, pp. 1711–1720, Aug. 2011.

[8] M. Shin *et al.*, "Maximization of acoustic energy difference between two spaces," *The J. Acoust. Soc. Amer.*, vol. 128, p. 121, 2010.

[9] Y. Cai *et al.*, "Time-domain acoustic contrast control design with response differential constraint in personal audio systems," *The J. Acoust. Soc. Amer.*, vol. 135, p. 252, 2014.

[10] F. Olivieri *et al.*, "Generation of private sound with a circular loudspeaker array and the weighted pressure matching method," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 8, pp. 1579–1591, Aug. 2017.

[11] A. Canclini *et al.*, "A weighted least squares beam shaping technique for sound field control," *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2018.

[12] J.-H. Chang and F. Jacobsen, "Sound field control with a circular double-layer array of loudspeakers," *The J. Acoust. Soc. Amer.*, vol. 131, p. 4518, 2012.

[13] M. A. Poletti, "An investigation of 2D multizone surround sound systems," *Proc. 125th AES Conv.*, 2008.

[14] J. Cheer *et al.*, "Practical implementation of personal audio in a mobile device," *J. Audio Eng. Soc.*, vol. 61, no. 5, 2013.

[15] M. F. Simón Gálvez, S. J. Elliott, and J. Cheer, "Time domain optimization of filters used in a loudspeaker array for personal audio," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 11, pp. 1869–1878, Nov. 2015.

[16] M. B. Møller and J. Ostergaard, "A moving horizon framework for sound zones," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 28, pp. 256–265, 2020.

[17] T. Lee, J. K. Nielsen, and M. G. Christensen, "Towards perceptually optimized sound zones: A proof-of-concept study," *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2019.

[18] M. Schneider and E. A. P. Habets, "Iterative DFT-domain inverse filter optimization using a weighted least-squares criterion," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 12, pp. 1957–1969, 2019.

[19] H. Hacihabibouglu and Z. Cvetkovic, "Multichannel dereverberation theorems and robustness issues," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 2, pp. 676–689, Feb. 2012.

[20] S. M. Kuo, B. H. Lee, and W. Tian, *Real-Time Digital. Sig. Process.: Implementations Appl.*, 3rd ed., Wiley, 2013.

[21] P. P. Vaidyanathanm, *Multirate. Syst. and Filter. Banks.*. Prentice-Hall, 1993.

[22] H. So and J. W. Choi, "Subband Optimization and filtering technique for practical personal audio systems," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2019.

[23] S. Weiss, "On adaptive filtering in oversampled subbands," Ph.D. dissertation, 1998.

[24] J. P. Reilly *et al.*, "The complex subband decomposition and its application to the decimation of large adaptive filtering problems," *IEEE Trans. Signal Process.*, vol. 50, no. 11, pp. 2730–2743, Nov. 2002.

[25] M. B. Møller and M. Olsen, "Sound zones: On performance prediction of contrast control methods," in *Proc. AES Int. Conf. Sound Field Control*, 2016.

[26] S. J. Elliott *et al.*, "Robustness and regularization of personal audio systems," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, no. 7, pp. 2123–2133, Sep. 2012.
[27] D. H. M. Schellekens, M. Møller, and M. Olsen, "Time domain acoustic contrast control implementation of sound zones for low-frequency input signals," *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2016.
[28] GTAC, "Audio room description," 2018. [Online]. Available: http://www.gtac.upv.es/room.php
[29] F. Fahy and J. Walker, *Advanced Applications in Acoustics, Noise and Vibration*. Spon Press, 2004.
[30] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Proc. 108th AES Conv.*, 2000.
[31] S. Weiss, M. Harteneck, and R. W. Stewart, "On implementation and design of filter banks for subband adaptive systems," in *Proc. IEEE Workshop Signal Process. Syst., SiPS: Des. Implementation*, 1998.
[32] J. Cheer and S. Elliott, "Design and implementation of a personal audio system in a car cabin," in *Proc. Meetings Acoust.*, 2013.
[33] P. D. Hatziantoniou and J. N. Mourjopoulos, "Generalized fractional-octave smoothing of audio and acoustic responses," *J. Audio Eng. Soc.*, vol. 48, no. 4, pp. 259–280, 2000.
[34] "Accelerating MATLAB Algorithms and Applications." [Online]. Available: https://es.mathworks.com/company/newsletters/articles/accelerating-matlab-algorithms-and-applications.html
[35] H. Kuttruff, *Room Acoustics*, 5th ed., Abingdon, U.K.: Spon Press, 2009.

Vicent Molés-Cases received the M.Sc. degree in telecommunication engineering from Universitat Politècnica de València (UPV), Spain, in 2016. He is currently a Ph.D. Student at the Institute of Telecommunications and Multimedia Applications (iTEAM) of UPV. From 2016 to 2017, he was a Research Engineer at Ericsson, Sweden, where he worked on concept development and standardization of 5G NR. In 2016, he received the award for the best M.Sc. thesis in Innovation for the Networked Society by the chair of the Official College of Telecommunications Engineers of Spain. His research interests include sound field control, loudspeaker array processing, and physical layer design for digital communications.

Gema Piñero (Senior Member, IEEE) received the M.Sc. degree from the Technical University of Madrid (UPM), Spain, in 1990, and the Ph.D. degree from the Technical University of Valencia (UPV), Spain, in 1997, both in electrical engineering. She is currently a Full Professor at the UPV. Since 1995 she has led or contributed to 30 private and public funded projects on array signal processing for acoustics and communications, active noise control, sound perception and acoustic sensor networks. She has co-authored more than 80 papers in international journals and conferences, serving in the Technical Committee in some of them. She is a member of the Editorial Board of the Digital Signal Processing Journal. From 2013 to 2016 she served in the Spanish Chapter of the IEEE Women in Engineering. She has been a visiting professor at the University of Illinois at Urbana-Champaign (2011) and the Imperial College London (2016). She is a Senior Member of the IEEE, member of the EURASIP, and Founding Member of the Spanish Association for Research and Teaching Universitas.

Maria de Diego (Senior Member, IEEE) received the M.Sc. degree in telecommunications engineering and the Ph.D degree from the Universitat Politècnica de València, Valencia, Spain, in 1994 and 2003, respectively. She is currently working as an Associate Professor in digital signal processing and communications. She is author of more than 30 journal papers and around 50 conference papers. She is a Senior Member of the IEEE and a Member of the International Institute of Acoustics and Vibration (IIAV), having been elected to the board of directors during the period 2015 to 2019. Moreover, she currently serves as Associate editor in the IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE. Her research interests are focused on active noise and equalization systems, distributed sound processing and audio applications over wireless acoustic networks, non-linear adaptive filtering, and finally sound quality and psychoacoustics.

Alberto Gonzalez (Senior Member, IEEE) received the graduate degree (Hons.) in telecommunication engineering from the Universitat Politecnica Catalunya, Barcelona, Spain, and the Ph.D. degree (magna cum laude) from the Universitat Politècnica de València, Valencia, Spain, in 1997.

During 1995, he worked as a Visiting Researcher in the Institute of Sound and Vibration Research, University of Southampton, Southampton, U.K., and is currently heading a research group in audio and communications digital signal processing. He has published more than 100 papers in international technical journals and renowned conferences in the fields of signal processing and applied acoustics and serves as the Dean of the Telecommunication Engineering School since June 2012. His research interests include optimization of computation methods for detection and decoding in digital communications and distributed sound signal processing. He belonged to the EURASIP Special Area Team on Acoustic, Speech and Music Signal Processing.