

## Weighted pressure matching with windowed targets for personal sound zones

Vicent Molés-Cases,<sup>1,a)</sup> Stephen J. Elliott,<sup>2</sup> Jordan Cheer,<sup>2,b)</sup> Gema Piñero,<sup>1,c)</sup> and Alberto Gonzalez<sup>1,d)</sup>

<sup>1</sup>*Institute of Telecommunications and Multimedia Applications, Universitat Politècnica de València, València 46022, Spain*

<sup>2</sup>*Institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ, United Kingdom*

### ABSTRACT:

Personal sound zones (PSZ) systems use an array of loudspeakers to render independent audio signals to multiple listeners within a room. The performance of a PSZ system, designed using weighted pressure matching, depends on the selected target responses for the bright zone. In reverberant environments, the target responses are generally chosen to be the room impulse responses from one of the loudspeakers to the control points in the selected bright zone. This approach synthesizes the direct propagation component and all the reverberant components in the bright zone, while minimizing the energy in the dark zone. We present a theoretical analysis to show that high energy differences cannot be achieved for the diffuse reverberant components in the bright and dark zones, and so trying to synthesize these components in the bright zone does not lead to the best performance. It is then shown that the performance can be improved by using windowed versions of these measured impulse responses as target signals, in order to control which reverberant components are synthesized in the bright zone and which are not. This observation is supported by experimental measurements in two scenarios with different levels of reverberation.

© 2022 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1121/10.0009275>

(Received 15 September 2021; revised 13 December 2021; accepted 14 December 2021; published online 20 January 2022)

[Editor: Mingsian R. Bai]

Pages: 334–345

### I. INTRODUCTION

Personal sound zones (PSZ) systems use an array of loudspeakers to render different audio signals to different users with minimum leakage between them.<sup>1,2</sup> To achieve this, a set of filters is used to process the audio signals that are fed to the loudspeakers. Different techniques have been proposed to compute the filters, such as beamforming,<sup>3,4</sup> soundfield synthesis,<sup>5–7</sup> energy cancellation approaches,<sup>8,9</sup> or hybrid approaches.<sup>10,11</sup> Among these, acoustic contrast control (ACC)<sup>12</sup> is the algorithm that can achieve highest isolation between the bright and dark zones, where the terms bright and dark zone refer to the regions where we want high and low acoustic energy, respectively.<sup>12</sup> However, ACC cannot synthesize a specific target response in the bright zone. To solve this limitation, the weighted pressure matching (WPM) algorithm has been proposed,<sup>13</sup> which offers the possibility to render a target response in the bright zone while keeping control over the energy in the dark zone. To do so, the authors in Ref. 13 proposed a novel cost function in which a weighting parameter is used to balance the components of the energy in the dark zone and the error with respect to the desired response in the bright zone. The WPM algorithm can be formulated in time-domain,<sup>14</sup>

subband-domain<sup>15</sup> and frequency-domain.<sup>16</sup> In this paper, because of its simplicity, we will use the frequency-domain formulation. Nevertheless, our findings can be generalized to the other formulations of WPM.

The main advantage of WPM over ACC is that it allows us to synthesize a desired target response in the bright zone; however, this is at the cost of higher energy in the dark zone (i.e., higher interference between the users of the system).<sup>13</sup> The performance of ACC and WPM has been compared under reverberant conditions in Refs. 17 and 18. The target response selected for the bright zone is a key choice for WPM systems, as different target responses can lead to different levels of energy in the dark zone. However, this is an aspect that, to the best of our knowledge, has not been previously studied in the literature. The most usual approach in reverberant environments is to select the target as the room impulse response (RIR) produced by one of the speakers of the system in the control points of the bright zone.<sup>14–16,19–21</sup> This approach aims to synthesize all the direct and reverberant components of the RIR in the bright zone while minimizing the energy in the dark zone. The late reverberation components, however, can be assumed to be diffuse for frequencies above the Schroeder frequency. We will show that there is no set of filters that can achieve high energy differences for the diffuse reverberant components in the bright and dark zones. Therefore, trying to synthesize these components in the bright zone while minimizing their energy in the dark zone does not give the best overall performance.

<sup>a)</sup>Electronic mail: vimoca3@iteam.upv.es, ORCID: 0000-0002-3303-953X.

<sup>b)</sup>ORCID: 0000-0002-0552-5506.

<sup>c)</sup>ORCID: 0000-0002-8719-8106.

<sup>d)</sup>ORCID: 0000-0002-6984-3212.

We therefore propose a variation of the WPM approach, in which a window function is applied to the target impulse response for the bright zone. By windowing this response, we can control which reverberant components are synthesized and which are minimized in the bright zone. Windowing has been previously used in the context of PSZ systems in Refs. 18, 22, and 23. In these works, the authors propose to window the RIRs used to compute the filters. The effect of windowing in these cases is similar to regularization, as it makes the filters more robust to inaccuracies in the RIRs. The approach proposed in this paper is conceptually different, as the RIRs used to compute the filters are not windowed, but windowing is instead applied to the targets for the bright zone. We present experimental evaluations in two scenarios that show the performance obtained with different window lengths. The results indicate that windowing the target can lead to performance improvements with respect to the case without windowing. In general, it seems that the optimal window length is frequency and scenario dependent, but the improvements that can be obtained are more significant for mid-high frequencies. Moreover, the results show that the higher the room reverberation, the higher the performance improvements obtained by windowing the target. Finally, we present evaluation results that show that the improvements in the performance obtained with the proposed method are robust to perturbations in the environment.

The paper<sup>1</sup> is structured as follows. Section II studies the WPM algorithm. Section III presents the novel target selection for WPM. Section IV presents experimental results to show the performance of the proposed strategy to select the target under different reverberation levels. Finally, Sec. V summarizes the main conclusions.

## II. WEIGHTED PRESSURE MATCHING

Let us consider a PSZ system that uses an array of  $L$  loudspeakers, and where the bright and dark zones are spatially sampled using  $M_b$  and  $M_d$  control points, respectively. Let us denote  $H_{ml,q}(f)$  as the room frequency response at frequency  $f$  between the  $l$ -th loudspeaker and the  $m$ -th control point in the  $q$ -th zone, where  $q \in \{b, d\}$ , and  $b$  and  $d$  are the indices of the bright and dark zones, respectively. Also, let us define  $G_l(f)$  as the frequency response of the filter used to filter the signals that will be fed to the  $l$ -th speaker. From now on, we will omit index  $f$  for the sake of simplicity. Next, let us define the  $M_q \times L$  matrix containing the room frequency responses between all loudspeakers and all control points in the  $q$ -th zone as

$$\mathbf{H}_q = \begin{bmatrix} H_{00,q} & \dots & H_{0(L-1),q} \\ \vdots & \ddots & \vdots \\ H_{(M_q-1)0,q} & \dots & H_{(M_q-1)(L-1),q} \end{bmatrix}. \quad (1)$$

Similarly, let us define the  $L \times 1$  vector containing the frequency responses of the filters for all loudspeakers in the system as

$$\mathbf{g} = [G_0 \quad \dots \quad G_{L-1}]^T. \quad (2)$$

Then, we can write the  $M_q \times 1$  vector containing the combined frequency responses for the control points in the  $q$ -th zone as

$$\mathbf{x}_q = \mathbf{H}_q \mathbf{g}. \quad (3)$$

Moreover, let us define  $D_{m,b}(f)$  as the target frequency response that we want to synthesize at the  $m$ -th control point of the bright zone (we assume that a null target response is selected for the dark zone). Then, we can write the  $M_b \times 1$  vector containing the target frequency responses for all the control points in the bright zone as

$$\mathbf{d}_b = [D_{0,b} \quad \dots \quad D_{M_b-1,b}]^T. \quad (4)$$

Once the model is presented, we describe the weighting pressure matching (WPM) algorithm, which was originally proposed in Ref. 13. The algorithm aims to find the filter coefficients  $\mathbf{g}$  that minimize the following cost function

$$J(\mathbf{g}) = \frac{\kappa}{M_d} \|\mathbf{H}_d \mathbf{g}\|^2 + \frac{(1-\kappa)}{M_b} \|\mathbf{H}_b \mathbf{g} - \mathbf{d}_b\|^2 + \lambda \|\mathbf{g}\|^2. \quad (5)$$

We can see that Eq. (5) is formed by three terms: (1) the mean energy in the dark zone, (2) the mean square error (MSE) with respect to the target frequency response in the bright zone, and (3) the energy of the filter coefficients. In Eq. (5),  $\lambda$  is a regularization factor that constrains the energy of the filters. Also,  $\kappa$  is the weighting factor satisfying  $0 \leq \kappa \leq 1$  that is used to balance the solution, e.g., high values of  $\kappa$  put more effort into minimizing the mean energy in the dark zone whereas low values put more effort into minimizing the MSE in the bright zone. It is straightforward to show that the optimal filter coefficients that minimize Eq. (5) are given by,<sup>13</sup>

$$\mathbf{g} = \left( \frac{\kappa}{M_d} \mathbf{H}_d^H \mathbf{H}_d + \frac{(1-\kappa)}{M_b} \mathbf{H}_b^H \mathbf{H}_b + \lambda \mathbf{I} \right)^{-1} \frac{(1-\kappa)}{M_b} \mathbf{H}_b^H \mathbf{d}_b. \quad (6)$$

## III. TARGET SELECTION

The target selected for the bright zone heavily influences the performance of WPM; however, this is an aspect that has not been extensively studied in the PSZ related literature. The most common approach in reverberant environments is to select the target impulse response as the delayed response from one of the loudspeakers to all of the control points in the bright zone,<sup>14-16,19-21</sup> i.e.,

$$d_{m,b} = h_{ml_r,b}(n - \tau), \quad (7)$$

where  $l_r \in \{0, \dots, L-1\}$  is the index of the reference loudspeaker, and  $\tau$  is a modelling delay that assures the causality of the filters. The target frequency response  $D_{m,b}$  is obtained

by computing the DTFT of  $d_{m,b}$  in Eq. (7). The previous selection aims to synthesize the direct propagation component and all the reverberant components produced by the reference loudspeaker in the bright zone.

Next, let us define

$$w_{L_w}(n) = 0, \quad \text{for } |n| \geq L_w, \quad (8)$$

which is a window function centered on  $n = 0$ . From the definition of the window in Eq. (8), it is easy to see that its length is  $2L_w - 1$ ; however, from now on, we are going to refer to the length of the window as  $L_w$  (i.e., the number of positive time samples in the window). Now, we alternatively propose to define the target impulse response in the control points of the bright zone as

$$d_{m,b}^{(L_w)}(n) = \begin{cases} w_{L_w}(n - \tau - \tau_p) h_{ml_r,b}(n - \tau) & \text{if } L_w < \infty \\ h_{ml_r,b}(n - \tau) & \text{otherwise,} \end{cases} \quad (9)$$

where  $\tau_p$  is the propagation delay corresponding to the direct component of  $h_{ml_r,b}$ , and in which the window is time-shifted such that its center is located on  $n = \tau_p + \tau$ . Then, we propose to select the target frequency response  $D_{m,b}$  used to form  $\mathbf{d}_b$  as the DTFT of  $d_{m,b}^{(L_w)}$  in Eq. (9). It is important to highlight that, contrary to what is proposed in Refs. 18, 22, and 23, no windowing is applied to the RIRs used to form matrices  $\mathbf{H}_b$  and  $\mathbf{H}_d$ . Also, it is relevant to note that the targets in Eqs. (7) and (9) are equivalent when  $L_w = \infty$ . However, Eq. (9) has the advantage that, by selecting the window length, we can choose which reverberant components we want to synthesize and which to minimize in the bright zone. Then, when selecting a window length that removes certain reverberant components, we are aiming to achieve a de-reverberation of the bright zone. We present in Fig. 1 a schematic to illustrate this effect. In the schematic, we show an example of a RIR between the reference loudspeaker and one control point in the dark and bright zones, and also, the target that we aim to achieve at these control points using a window of length  $L_w$ . From this schematic, it is easy to see that when  $L_w = \infty$ , we want to synthesize all the reverberant components in the bright zone while minimizing the energy of all the components in the dark zone. However, the effect is different when we select finite values for  $L_w$ . For time instants within the window, we seek to

synthesize the direct propagation component and some early reflections in the bright zone, while minimizing the energy in the dark zone. For time instants after the end of the window, we want to minimize the energy of the reverberant components both in the bright and dark zones. The proposed target selection can be directly applied for the time-domain formulation of WPM<sup>14</sup> by using Eq. (9) to form the vector with the target impulse responses in all the control points of the bright zone.

It is not initially clear whether windowing the target response will lead to performance improvements. However, above the Schroeder frequency, the late reverberation components of the RIR become diffuse.<sup>24</sup> These components can then be assumed to be of similar energy but uncorrelated between the different control points. Now, let us define  $\mathbf{H}_{b,dif}$  and  $\mathbf{H}_{d,dif}$  as the matrices containing the diffuse components of the RIR at single frequency  $f$  in the bright and dark zones, respectively. From the ACC algorithm,<sup>25</sup> we know that the maximum acoustic contrast (AC) that can be achieved between the diffuse components of the bright and dark zones is given by the highest eigenvalue of matrix  $(\mathbf{H}_{d,dif}^H \mathbf{H}_{d,dif})^{-1} \mathbf{H}_{b,dif}^H \mathbf{H}_{b,dif}$ . For the diffuse components, we can assume that  $\mathbf{H}_{b,dif}^H \mathbf{H}_{b,dif} \approx \sigma \mathbf{I}$  and  $\mathbf{H}_{d,dif}^H \mathbf{H}_{d,dif} \approx \sigma \mathbf{I}$  (where  $\sigma$  is the energy of the diffuse field in the room).<sup>26</sup> Then, the maximum AC that can be achieved between the diffuse components of the bright and dark zones is approximately 1 (in linear units). This shows that there is no set of filters that is able to provide a significant energy difference between the diffuse components in the bright and dark zones. Thus, selecting a target that tries to synthesize all of the reverberant components, including the diffuse components in the bright zone (i.e., with  $L_w = \infty$ ) is not a good choice. In Sec. IV, we will present evaluation results that show that improvements in the performance can be achieved by windowing the target, i.e., aiming the minimization of the energy of the diffuse components in both zones.

#### IV. RESULTS

In this section, we evaluate the performance of the proposed target selection for different window lengths and for two scenarios with different reverberation times. First, we define the experimental setup and the metrics used for the evaluations. Then, we present evaluation results that show the effect on the performance of windowing the target response for the bright zone. Finally, we evaluate the

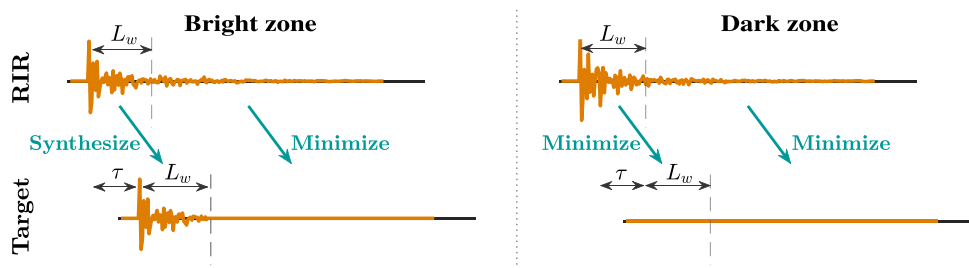


FIG. 1. (Color online) Schematic to illustrate the proposed target selection. The upper plots represent the RIR between the reference loudspeaker and one control point in the bright and dark zones, and the lower plots represent the windowed target at these control points using a window of length  $L_w$ .

robustness to perturbations in the filters computed with a windowed target.

**A. Setup**

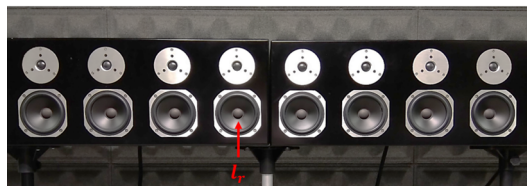
The experimental evaluations have been carried out in two scenarios using rooms with different reverberations levels:

- *Scenario 1:* Office-like room at the Institute of Telecommunications and Multimedia Applications (iTEAM) of size  $7.2 \times 11.72 \times 2.63$  m and reverberation time  $T_{60} = 0.5$  s. The Schroeder frequency for this room is 95 Hz. In this scenario, a linear array of 8 two-way loudspeakers with an inter-element distance of 0.18 m has been used [see Fig. 2(a)]. The setup is formed by one bright and one dark zone [Fig. 3(a)]. In each zone, two different grids of microphones have been used for spatial sampling, such that the RIRs measured with the control grid are used to compute the filters and the RIRs measured with the validation grid are used to evaluate the filters. The highest frequency that can be rendered without spatial aliasing is 1498 Hz.<sup>27</sup>
- *Scenario 2:* Listening room at iTEAM, which is a rectangular room of size  $9.07 \times 4.45 \times 2.65$  m with acoustically treated walls and reverberation time  $T_{60} = 0.18$  s. The Schroeder frequency for this room is 137 Hz. Similarly to scenario 1, a linear array of 8 two-way loudspeakers with an inter-element distance of 0.18 m [Fig. 2(b)] has been used and two zones are considered, which are spatially sampled using two different grids of microphones [see Fig. 3(b)]. The highest frequency that can be rendered without spatial aliasing is 1401 Hz.<sup>27</sup>

In all cases, the RIRs were measured using the exponentially swept sine technique<sup>28</sup> with a sampling frequency of 44 100 Hz. In the following, we use notation  $\mp$  to denote

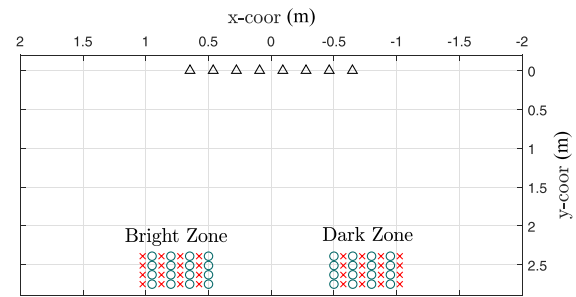


(a)

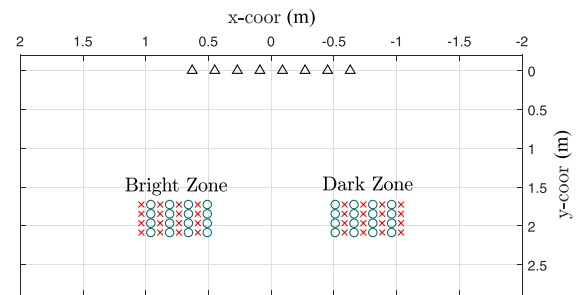


(b)

FIG. 2. (Color online) Array of loudspeakers used in scenarios 1 and 2 in (a) and (b), respectively. The reference loudspeaker used to select the target is indicated using a red arrow in each case.



(a)



(b)

FIG. 3. (Color online) Setups for scenarios 1 and 2 in (a) and (b), respectively. Circle, cross, and triangle markers denote control points, validation points, and loudspeakers, respectively. The walls are located in  $x = \pm 3.2$  m,  $y = \pm 5.86$  m, and  $z = \{0, 2.65\}$  m in (a), and in  $x = \pm 4.53$  m,  $y = \{-0.46, 3.99\}$  m, and  $z = \{0, 2.65\}$  m in (b). The loudspeakers and microphones are located at a height of 1.56 m and 1.51 m in (a) and (b), respectively.

the elements related to the validation grid of microphones. Due to the effects of the spatial aliasing in the arrays of loudspeakers, we limit the study to the frequency range 150–2000 Hz.

The optimal filters have been computed using Eq. (6), where the weighting parameter  $\kappa$  is set to 0.5, if not indicated otherwise. Moreover, the regularization parameter  $\lambda$  is selected such that the array effort [which will be defined next in Eq. (12)] is upper bounded as  $AE \leq AE_{max}$ . The value of  $AE_{max}$  will be indicated in each case. The target for the bright zone has been selected using Eq. (9), where we select  $w_{L_w}$  as a Tukey window with cosine fraction  $\alpha = 0.3$ <sup>29</sup> and the reference loudspeaker is  $l_r = 3$  [as indicated in Figs. 2(a) and 2(b)]. We select a Tukey window because, when applied to the target response, it does not modify the amplitude of the first early reflections of the response and it leads to a smooth transition between the reverberant and null components of the target. Other types of windows have also been investigated but gave similar results. Moreover, when  $L_w < \infty$ , we apply a  $\frac{1}{3}$  octave band equalizer to the target to obtain the same energy in the bright zone as for the case with  $L_w = \infty$ . In Fig. 4(a), we show an example of the causal part of a Tukey window with  $L_w = 529$  (12 ms), and Figs. 4(b) and 4(c) show examples of a target impulse response and a frequency response in the

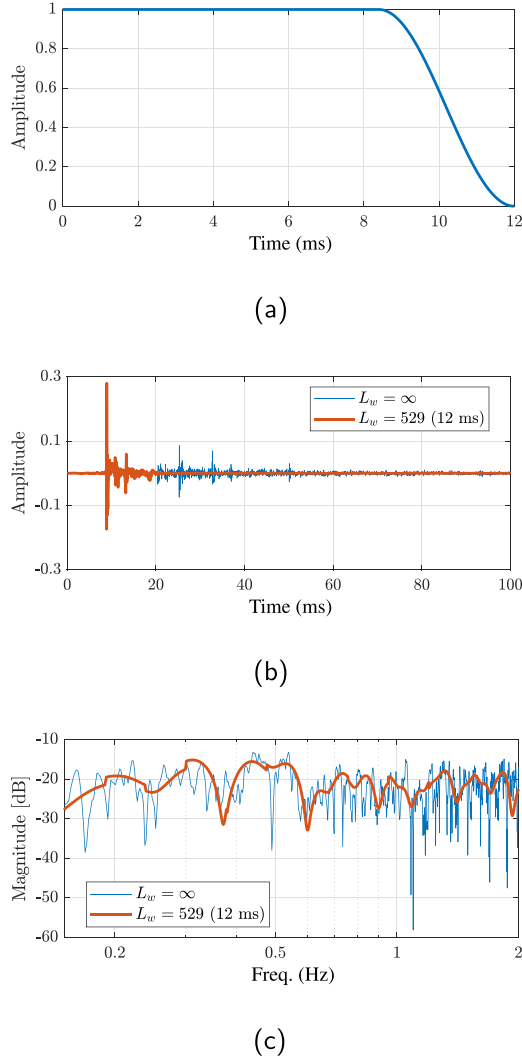


FIG. 4. (Color online) Example of the causal part of a Tukey window with  $L_w = 529$  (12 ms) and cosine fraction  $\alpha = 0.3$  in (a), target impulse response and transfer function in the 0-th control point of the bright zone for scenario 1 with different window lengths in (b) and (c), respectively.

bright zone for scenario 1 with  $L_w = 529$  (12 ms) and  $L_w = \infty$ .

### B. Metrics

Next, we present the metrics used for the evaluations. First, we define the MSE in the bright zone as

$$\varepsilon_b = \|\bar{\mathbf{H}}_b \mathbf{g} - \bar{\mathbf{d}}_b\|^2, \quad (10)$$

where the target is selected using the same criterion used to compute the filters, meaning that the same window  $w_{L_w}$  is applied to obtain the targets  $\mathbf{d}_b$  and  $\bar{\mathbf{d}}_b$  that are used to compute and to evaluate the filters, respectively. Next, we define the mean energy in the dark zone as

$$E_d = \|\bar{\mathbf{H}}_d \mathbf{g}\|^2, \quad (11)$$

and the array effort as

$$AE = \frac{\|\mathbf{g}\|^2}{E_{ref}}, \quad (12)$$

where  $E_{ref}$  is the energy required by the reference loudspeaker to provide the same energy in the bright zone as the array of loudspeakers using the set of filters  $\mathbf{g}$ .

So far, we have defined metrics that are directly related to the three components of the cost function in Eq. (5), and whose influence can be adjusted with  $\lambda$  and  $\kappa$ . Now, we define the acoustic contrast (AC) as

$$C = \frac{\|\bar{\mathbf{H}}_b \mathbf{g}\|^2}{\|\bar{\mathbf{H}}_d \mathbf{g}\|^2}, \quad (13)$$

which is a metric related to the level of acoustic isolation between the bright and dark zones. The AC is not directly present in cost function [Eq. (5)] but is commonly used as a performance indicator for PSZ systems. Finally, we define the mean Kurtosis of the measured RIRs at sample time  $n$  as

$$K(n) = \frac{1}{N_h} \sum_{\forall m, \forall l, \forall q} \left( \frac{\mathbb{E} \left\{ \left( \bar{h}_{ml,q} - \mu_{ml,q}^{(n)} \right)^4 \right\}}{\left( \sigma_{ml,q}^{(n)} \right)^4} - 3 \right), \quad (14)$$

where  $\mu_{ml,q}^{(n)}$  and  $\sigma_{ml,q}^{(n)}$  are the mean value and the standard deviation of  $\bar{h}_{ml,q}$ , respectively, over the interval  $n, \dots, n + L_s - 1$ , and the expected value for time  $n$  is computed over the interval  $n, \dots, n + L_s - 1$ . In Eq. (14),  $N_h$  denotes the total number of RIRs for which the mean Kurtosis is computed. The Kurtosis is a measure of the tailedness of a sample distribution. For a Gaussian probability density function (PDF), the Kurtosis value is 0. The authors in Ref. 30 show that the Kurtosis is closely related to the diffuseness of late reverberations. In particular, they show that the early part of the RIR, containing the direct propagation component and strong deterministic reflections, is unlikely to have a Gaussian distribution, so it presents high Kurtosis levels. However, the late diffuse components of the RIR present Kurtosis values close to 0. The motivation is that the reflection density is high for the diffuse part of the RIR, which makes it more likely that the RIR has a Gaussian distribution. For the computation of the Kurtosis, we used a length segment of  $L_s = 882$  (20 ms) and all the RIRs have been aligned, such that their direct propagation component is located at sample index  $n = 0$ . From now on, we use a third-octave averaging<sup>31</sup> for all frequency-domain plots to improve the readability of the results.

### C. Performance for scenario 1

In this subsection, we present experimental results to evaluate the performance of the proposed target selection for scenario 1.

First, we compare the proposed target selection with the approach employed in Refs. 18, 22, and 23 in which not only the target used to form  $\mathbf{d}_b$  is windowed but also the

RIRs used to form  $\mathbf{H}_q$ . We show in Fig. 5 the performance of the proposed approach as a function of  $L_w$  in terms of: mean energy in the dark zone [in Figs. 5(a) and 5(d)], MSE in the bright zone [in Figs. 5(b) and 5(e)], and array effort [in Figs. 5(c) and 5(f)]. Similarly, the performance of the approach used in Refs. 18, 22, and 23 is shown in Fig. 6. As expected, the performance in both cases is identical for very long windows, e.g., 200 ms. We can see in Fig. 6 that windows shorter than 100 ms highly degrade the performance for the case with windowed RIRs and targets. This is because relevant reverberation components of the RIRs are not taken into account in the optimization. On the contrary, the proposed approach takes into account all the reverberant components of the RIR in the optimization but tries to minimize the energy of some of these components. The results in Fig. 5 show that short windows can offer better performance in this case. These results indicate that the performance of the system can be improved with the proposed approach by selecting short windows, while this is not possible with the approach used in Refs. 18, 22, and 23. From now on, we focus on further evaluating the performance of the proposed method.

Next, we discuss the effect of the window length in the performance of the proposed target selection. In Fig. 5, the metrics in the top and bottom rows are computed with  $AE_{max} = 0$  dB and  $AE_{max} = 15$  dB, respectively, but the performance for both effort constraints is equal for frequencies above 800 Hz. The motivation is that the matrix that must be inverted in Eq. (6) is well conditioned above 800 Hz, and

then, even with a regularization parameter of  $\lambda = 0$  does not lead to a high array effort. For frequencies below 800 Hz, the performance is different for the two effort constraints, however, the effect of the window length in the performance is similar in both cases. Then, the following analysis is valid for the two studied constraints. We can see in Figs. 5(b) and 5(e) that very short windows (of 1 ms or less) lead to an MSE that is at least 10 dB worse than for longer windows, while the energy generated in the dark zone is similar [in Figs. 5(a) and 5(d)]. This degradation appears because very short windows aim to suppress the first early reflections, and also, to equalize the response of the reference loudspeaker in the bright zone, which is too restrictive and leads to important errors in the bright zone. Also, we can see that, in general, short windows (of about 12 ms) present a significantly lower energy in the dark zone than longer windows. The motivation is that these windows remove the diffuse components from the target in the bright zone, and then, the designed filters are capable of minimizing the diffuse field, both in the bright and dark zones. Then, the improvements obtained by windowing the target are not obtained because we force the cancellation of some high energy early reflections, but because we target the cancellation of the diffuse reverberant tail. In particular, for frequencies in the range of 150–200 Hz, 400–500 Hz, and above 700 Hz, the improvements in the energy in the dark zone are not at the cost of substantially higher MSE. For frequencies 200–400 Hz and 500–700 Hz, short windows lead to lower energy in the dark zone at the cost of worse MSE. For these frequency bands,

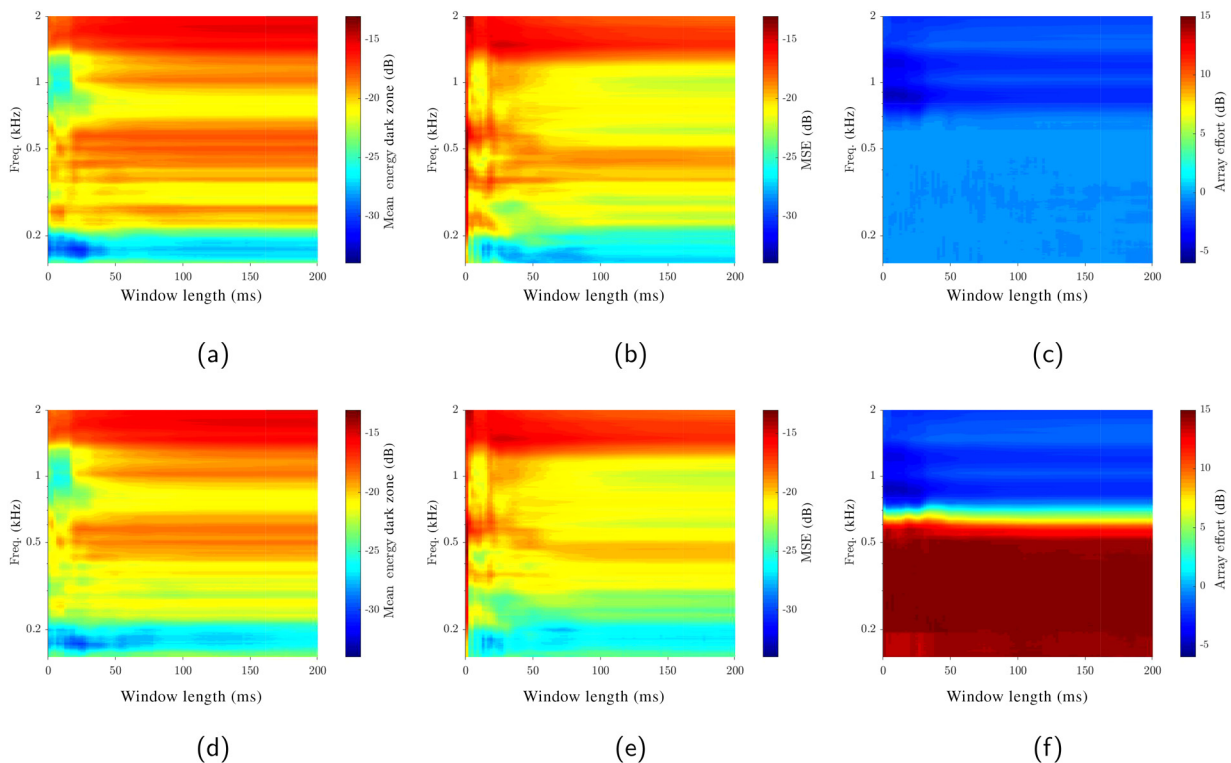


FIG. 5. (Color online) Performance of the proposed target selection in Sec. III as a function of the window length and frequency for scenario 1 in terms of: mean energy in the dark zone (a) and (d), MSE in the bright zone (b) and (e), and array effort (c) and (f). For the top row figures  $AE_{max} = 0$  dB, and  $AE_{max} = 15$  dB for the bottom figures.

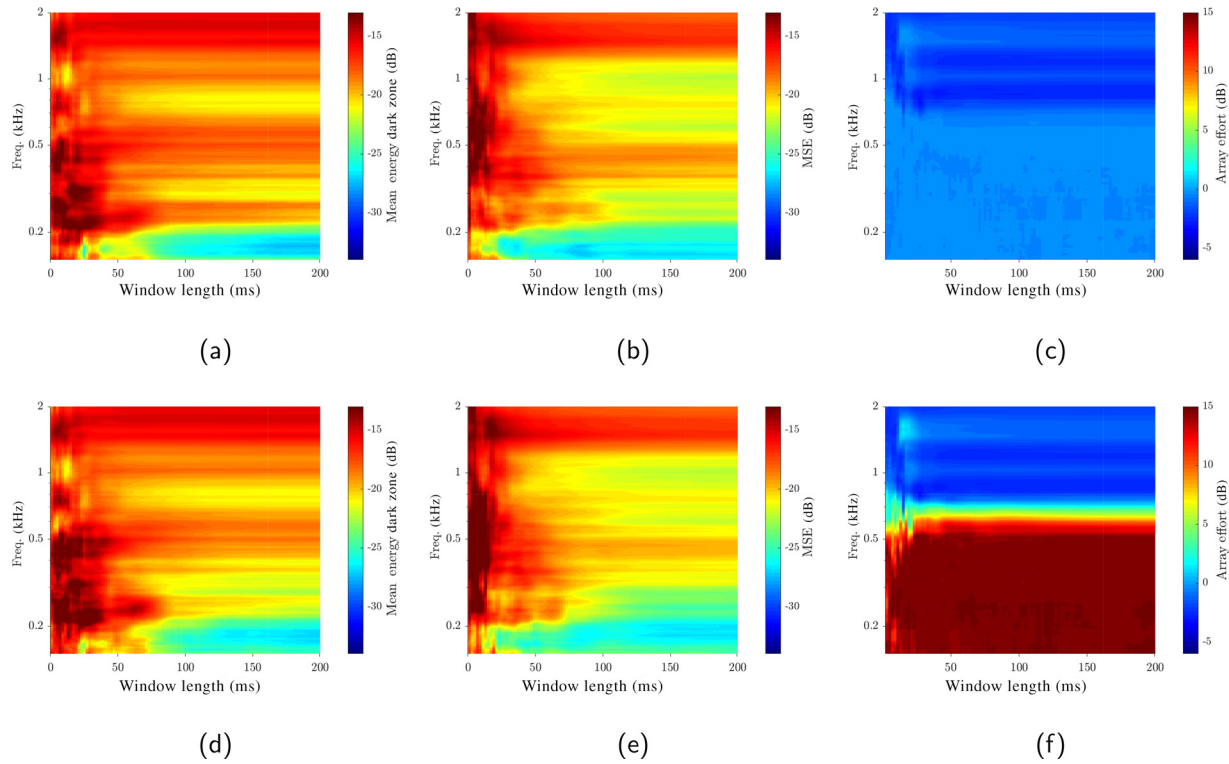


FIG. 6. (Color online) Performance of the approach used in Refs. 18, 22, and 23 as a function of the window length and frequency for scenario 1 in terms of: mean energy in the dark zone (a) and (d), MSE in the bright zone (b) and (e), and array effort (c) and (f). In this case, both the RIR and the target are windowed. For the top row figures  $AE_{max} = 0$  dB, and  $AE_{max} = 15$  dB for the bottom figures.

lower energy in the dark zone may be achieved for long windows by tuning the weighting parameter  $\kappa$ ; therefore, at this point, it is not yet clear if real improvements in the performance are achieved with short windows in all the studied frequency range.

Now, we present additional results where we compare the following configurations:

- *Win. Target (WT)*:  $L_w = 529$  (12 ms) and  $\kappa = 0.5$ .
- *Full Target (FT)*:  $L_w = \infty$  and  $\kappa = 0.5$ .
- *Full Target with Frequency Dependent  $\kappa$  (FT-FD)*:  $L_w = \infty$  and  $\kappa$  selected as shown in Fig. 7 to achieve the same MSE as WT.

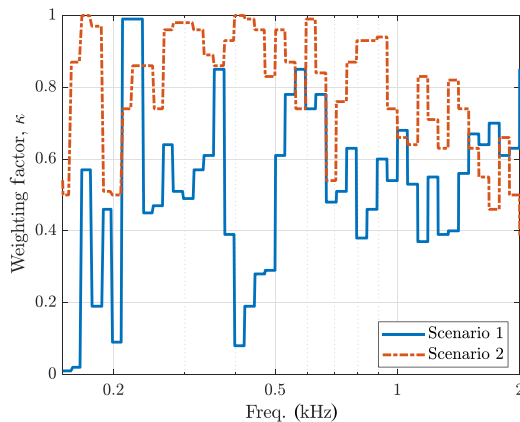


FIG. 7. (Color online) Weighting factor  $\kappa$  for FT-FD in scenarios 1 and 2.

It is important to note that in order to fairly determine which window length is able to provide lower energy in the dark zone, we must compare their performance for the case in which their MSE is equal. This means that reductions in the energy in the dark zone for one of the configurations compared to another are then not at the cost of higher MSE in the bright zone. This is the motivation to include FT-FD in the comparison. The performance of the three configurations is shown in Fig. 8 in terms of: mean energy in the dark zone [in Fig. 8(a)], MSE in the bright zone [in Fig. 8(b)], array effort [in Fig. 8(c)], and acoustic contrast [in Fig. 8(d)]. We only present results for  $AE_{max} = 15$  dB, as we can see in Fig. 5 that the effect of the window length on the performance is similar for different effort constraints. We see that WT can offer lower energy in the dark zone than FT and also lower or equal MSE in certain frequency bands, e.g., 150–200 Hz, 400–500 Hz, 700–900 Hz, and 1.1–1.5 kHz. This means that a longer window cannot offer better performance than  $L_w = 529$  for these frequencies, neither in terms of energy in the dark zone nor in terms of MSE (even if we tune  $\kappa$ ). Next, if we compare WT and FT-FD, we can see that WT leads to lower energy in the dark zone for all the studied frequencies except for the band 200–360 Hz, where FT-FD presents slightly lower energy. For example, WT leads to almost 6 dB lower energy in the dark zone than FT-FD around 1 kHz, and 5 dB lower energy around 180 Hz. Obtaining lower energy in the dark zone does not necessarily mean that the acoustic contrast is higher, since the energy reduction could be produced because the energy in the bright zone is

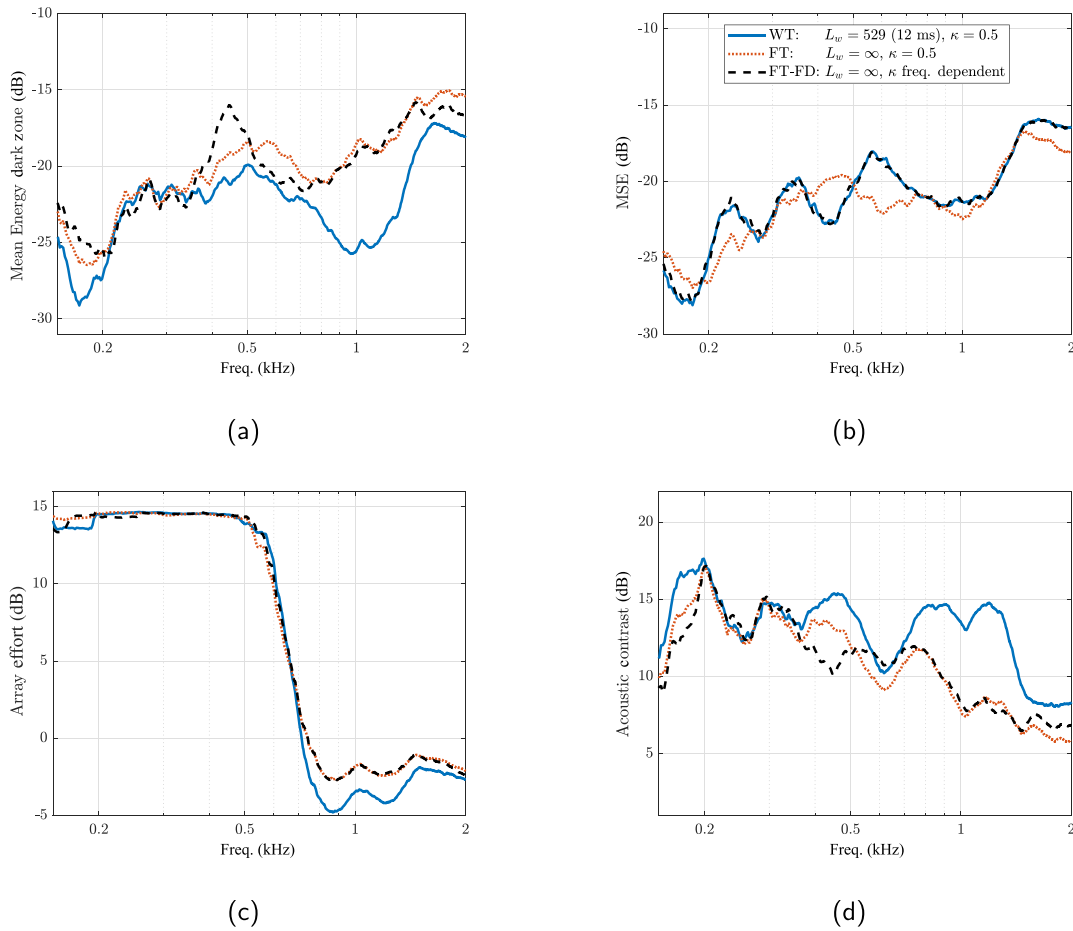


FIG. 8. (Color online) Performance for three different configurations as a function of frequency for scenario 1 in terms of: (a) mean energy in the dark zone, (b) MSE in the bright zone, (c) array effort, and (d) acoustic contrast.  $AE_{max} = 15$  dB is considered.

also lower. This aspect is especially important for the comparison of targets windowed with different window lengths, as these targets present different energy levels for the bright zone at individual frequencies (although their mean energy is equal across  $\frac{1}{3}$  octave bands). Nonetheless, the results in Fig. 8(d) indicate that, in general, the acoustic contrast follows an inverse trend compared with those for the energy in the dark zone, since for the same MSE level, WT can offer higher contrast than FT-FD across almost all of the studied frequencies (with improvements of more than 5 dB). Moreover, the array effort required by WT is lower than for FT and FT-FD for frequencies above 700 Hz. The MSE is almost the same for WT and FT-FD, as expected, and is broadly similar for FT, although slightly higher or lower for different frequencies. From these results, we can conclude that, for scenario 1, windowing the target response with a short window of length  $L_w = 529$  (12 ms) leads to significantly better performance than the case without windowing for most of the studied frequencies. This indicates that targeting the minimization of the energy of the diffuse reverberant components in the bright and dark zones can lead to great improvements in the performance for this scenario.

Finally, Fig. 9 shows the mean Kurtosis of the RIRs for scenario 1. We can observe that the Kurtosis has high values

for the early part of the RIR, where the direct component and the early reflections are located, and drops to values below 3 after about 20 ms, indicating that the reverberant components of the RIRs can be assumed diffuse, with a Gaussian PDF, after this time.<sup>30</sup> It is for window lengths greater than about 20 ms that the performance starts to deteriorate in Fig. 5. Then, the Kurtosis provides a useful metric for the selection of the window length.

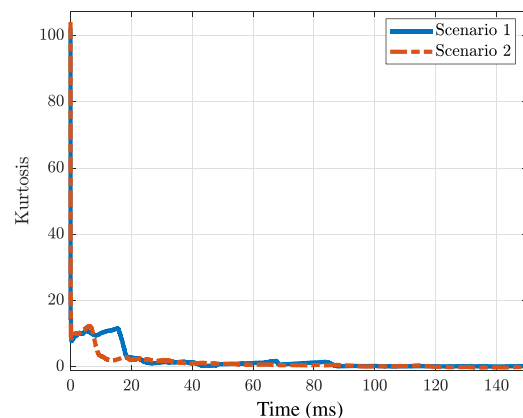


FIG. 9. (Color online) Mean Kurtosis for scenarios 1 and 2.



### D. Performance for scenario 2

Next, we study the performance of the proposed target selection for scenario 2, with the lower reverberation time. Figure 10 shows the performance as a function of  $L_w$  in terms of: mean energy in the dark zone [in Figs. 10(a) and 10(d)], MSE in the bright zone [in Figs. 10(b) and 10(e)], and array effort [in Figs. 10(c) and 10(f)]. In Fig. 10, the metrics in the top row are again computed with  $AE_{max} = 0$  dB, and  $AE_{max} = 15$  dB for the bottom row. As for scenario 1, the performance is similar for both constraints at frequencies above 800 Hz, as are the effects of the window length for frequencies below 800 Hz in both cases. We can see in Figs. 10(b) and 10(e) that very short windows (of 1 ms or less) lead to a much higher MSE than longer windows, while the energy generated in the dark zone is similar [in Figs. 10(a) and 10(d)]. Similarly to scenario 1, this is produced because very short windows aim to suppress the first early reflections and to equalize the response of the reference loudspeaker in the bright zone, which leads to high MSE. Then, selecting a window that is too short can significantly degrade the performance. The mean energy in the dark zone is lower for shorter windows (of about 8 ms) than longer ones, particularly for frequencies above 1 kHz. However, we can also see in Figs. 10(b) and 10(e) that, in general, the MSE is higher for short windows. In this case, the lower energy generated in the dark zone when using short windows comes at the cost of higher MSE in the bright zone.

To further study if windowing the target offers performance improvements in this scenario, we present additional

results in Fig. 11 where we compare three configurations which are similar to those used above:

- *Win. Target (WT)*:  $L_w = 353$  (8 ms) and  $\kappa = 0.5$ .
- *Full Target (FT)*:  $L_w = \infty$  and  $\kappa = 0.5$ .
- *Full Target with Frequency Dependent  $\kappa$  (FT-FD)*:  $L_w = \infty$  and  $\kappa$  selected as shown in Fig. 7 to achieve the same MSE as WT.

We again include FT-FD to fairly determine which window length is able to provide lower energy in the dark zone, while giving the same MSE. The performance of these configurations is shown in Fig. 11 in terms of: mean energy in the dark zone [in Fig. 11(a)], MSE in the bright zone [in Fig. 11(b)], array effort [in Fig. 11(c)], and acoustic contrast [in Fig. 11(d)]. As for scenario 1, we only present results for  $AE_{max} = 15$  dB to avoid redundancy. Above about 700 Hz, the energy in the dark zone is lower for WT than for FT, but even lower than for FT-FD above 1 kHz. In this case, however, FT has a significantly lower MSE than WT or FT-FD. From these results, we can conclude that windowing the target response for the bright zone with a short window of 8 ms can lead to better performance in terms of acoustic contrast than the case without windowing for frequencies above 1 kHz in this scenario. These results and the results in Sec. IV C show that the optimal window length is frequency and scenario dependent. Comparing the results for the two scenarios, we can see that the higher the reverberation in the room, the greater the improvement obtained by windowing the target response in the bright zone.

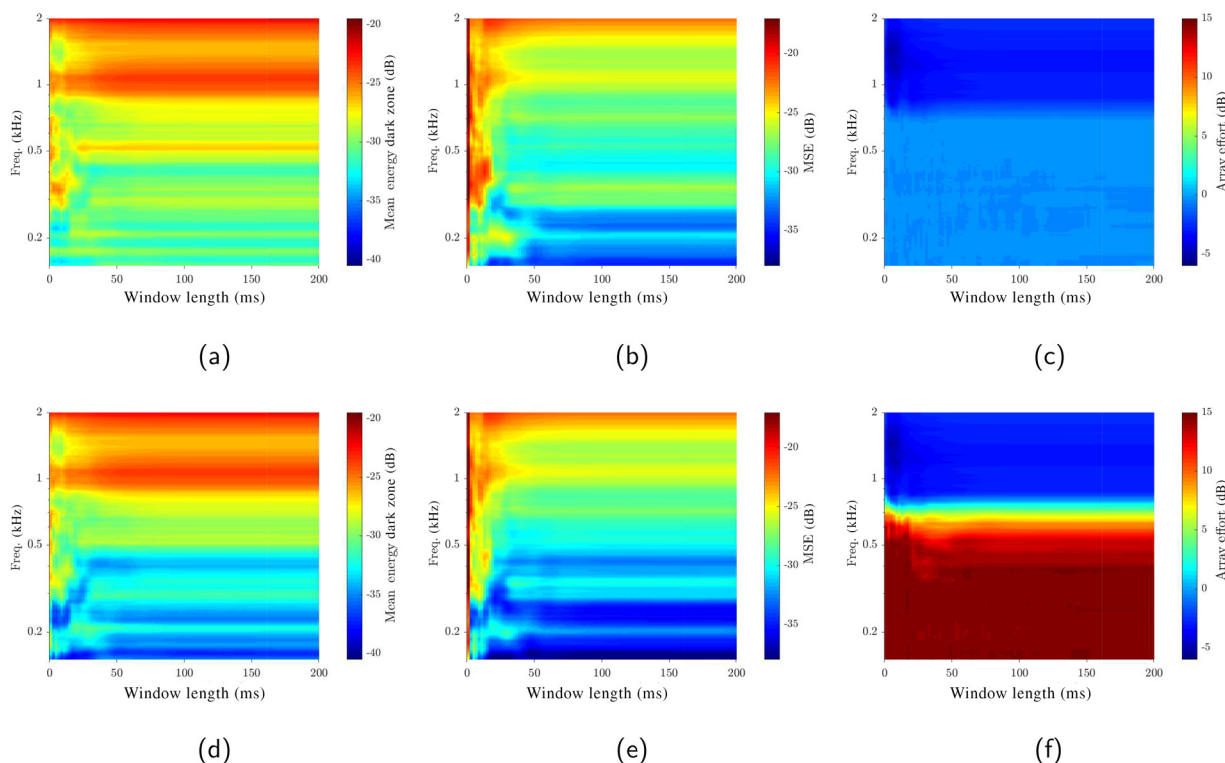


FIG. 10. (Color online) Performance of the proposed target selection in Sec. III as a function of the window length and frequency for scenario 2 in terms of: mean energy in the dark zone (a) and (d), MSE in the bright zone (b) and (e), and array effort (c) and (f). For the top row figures,  $AE_{max} = 0$  dB, and  $AE_{max} = 15$  dB for the bottom figures.

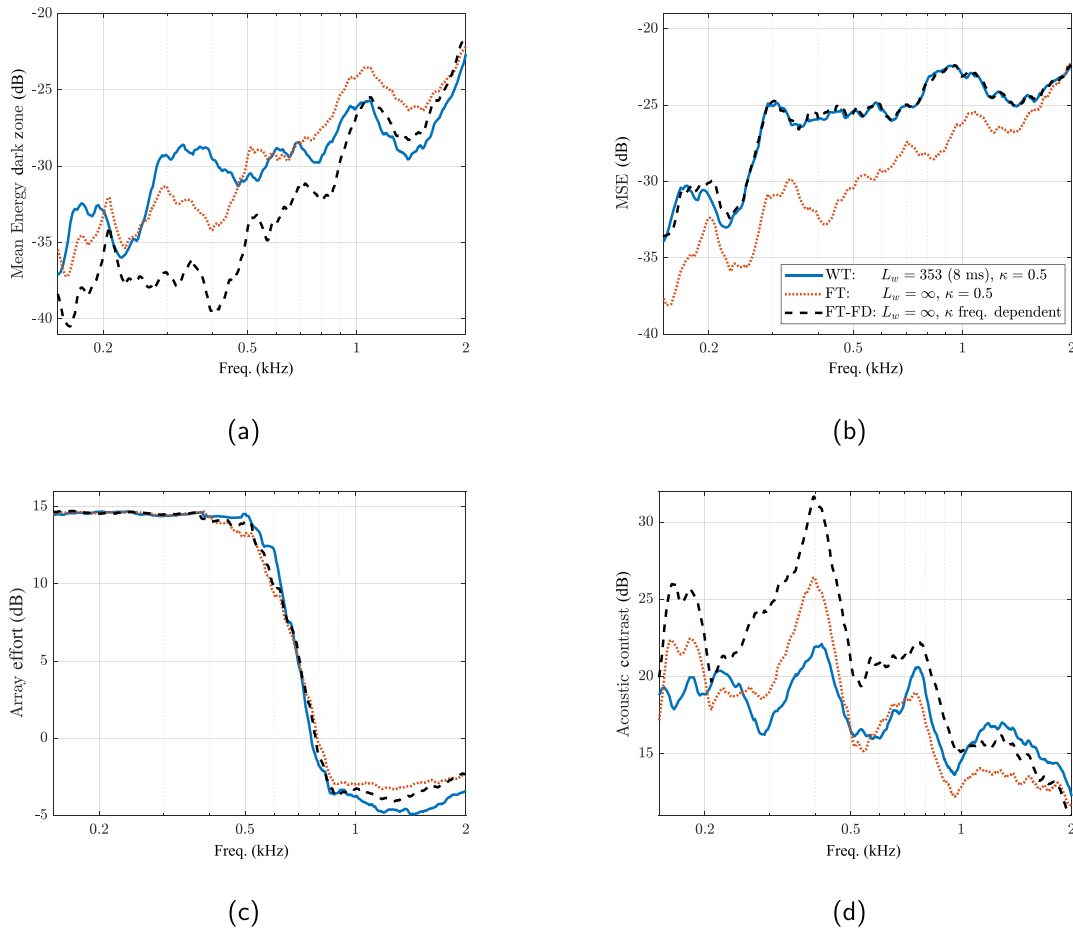


FIG. 11. (Color online) Performance for three different configurations as a function of frequency for scenario 2 in terms of: (a) mean energy in the dark zone, (b) MSE in the bright zone, (c) array effort, and (d) acoustic contrast.  $AE_{max} = 15$  dB is considered.

Figure 9 shows the mean Kurtosis of the RIRs for scenario 2, which has high values for times smaller than 10 ms, while it drops below 3 for times greater than about 10 ms. This indicates that the reverberant components of the RIRs are diffuse after about 10 ms, which is the window length after which the performance starts to degrade in Fig. 10, again illustrating the use of the Kurtosis in estimating the window length, although only at higher frequencies, in this case, with a short reverberation time.

### E. Robustness against perturbations

We now evaluate whether the improvements obtained by windowing the target are robust to perturbations in the environment. To this end, we present in Fig. 12 evaluations results for scenario 1, where the filters are computed with the RIRs measured at the control points without any perturbation [as in Fig. 13(a)] and evaluated using the RIRs measured at the control points when perturbations in the RIR, due to two people located within the zones, are present [as in Fig. 13(b)]. The control filters are those calculated without any perturbations, as in Sec. IV C, but now the performance, as shown in Fig. 12, has been calculated after these perturbations in the RIR. For comparison purposes, we also include in Fig. 12 the performance of the filters when

evaluated without perturbations (as in Sec. IV C). We can see that the perturbations have generally increased the energy in the dark zone and the MSE in the bright zone with respect to the case without perturbations. The mean energy in the dark zone is still significantly smaller for the windowed target in WT than it is without the window, and the MSE is also again broadly similar in the two cases. We can thus conclude that the performance improvements obtained by selecting a short window for scenario 1 are robust to perturbations in the environment.

### V. CONCLUSIONS

In this paper, we proposed a novel approach to select the target response in the bright zone for the weighting pressure matching (WPM) algorithm in personal sound zones systems. In previous works, the target for the bright zone has generally been selected to be the room impulse response (RIR) from one loudspeaker to all the control points in the bright zone. The aim is thus to synthesize the direct propagation component and all the reverberant components in the bright zone, while minimizing the energy of all components in the dark zone. The late reverberant components, however, are diffuse above the Schroeder frequency and it is shown that there is no set of filters that can lead to high energy

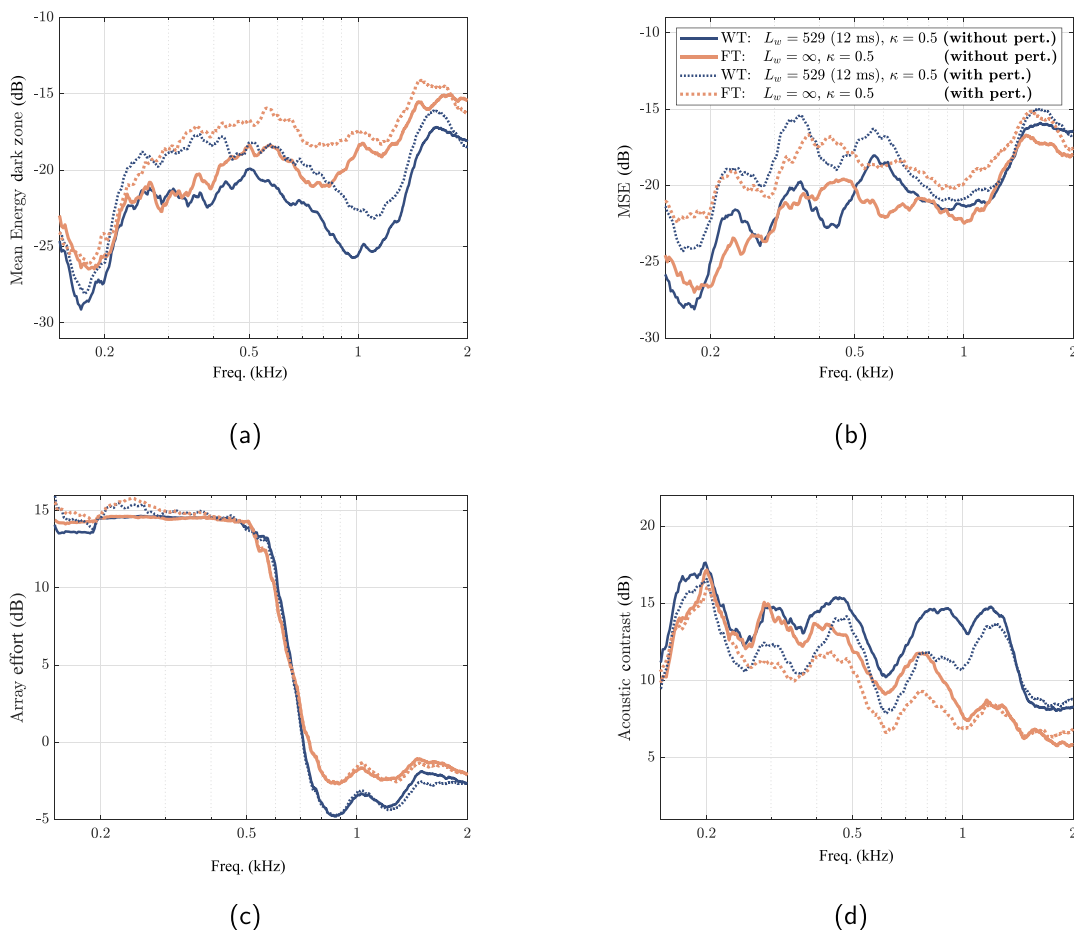
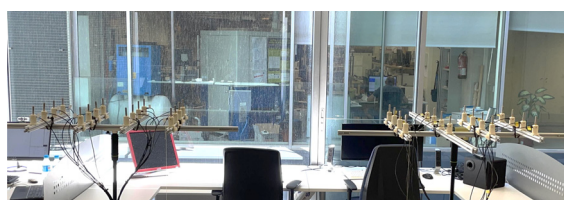


FIG. 12. (Color online) Performance for WT and FT as a function of frequency for scenario 1 with perturbations in terms of: (a) mean energy in the dark zone, (b) MSE in the bright zone, (c) array effort, and (d) acoustic contrast.  $AE_{max} = 15$  dB is considered. The performance of the filters evaluated without perturbations is also shown for comparison purposes.



(a)



(b)

FIG. 13. (Color online) Setup used to measure the RIRs without perturbations in (a), and to measure the RIRs with the perturbations produced by two persons within the zones in (b).

differences between the diffuse reverberant components in the bright and dark zones, so trying to synthesize all of the reverberation components in the bright zone while minimizing their energy in the dark zone is not a good strategy. Alternatively, we proposed to window the RIRs from one loudspeaker to all the control points in the bright zone and use these responses as target for the bright zone. This approach allows us to select which reverberant components we want to synthesize in the bright zone by choosing the window length. Experimental evaluation results in two rooms with different levels of reverberation show the effect of windowing the target response. The results showed that windowing the target response can lead to lower energy in the dark zone and higher acoustic contrast than the case without windowing, with similar mean square error (MSE) in the bright zone in the case of the more reverberant room and with similar array effort. Specifically, improvements of up to 6 dB in the acoustic contrast are observed for a room with  $T_{60} = 500$  ms when a window length of 12 ms is used in the target impulse responses. The window length that offers best performance is, in general, both frequency and

scenario dependent. The Kurtosis of the RIRs, which is related to their level of diffuseness, is also shown to give a good indication of the best window length. Also, we observed that greater improvements with respect to the case without windowing are obtained for mid-high frequencies. The improvements obtained by windowing the target are greater for scenarios with high reverberation level and are also robust to perturbations in the room impulse responses.

## ACKNOWLEDGMENTS

Vicent Molés-Cases was supported by the Spanish Ministry of Education through Grant No. FPU17/01288. Stephen J. Elliott was supported by the EPSRC DigiTwin project (EP/R006768/1). Jordan Cheer was supported by the Intelligent Structures for Low Noise Environments (ISLNE) EPSRC Prosperity Partnership (EP/S03661X/1). This research was supported by the Spanish Ministry of Science, Innovation and Universities through Grant No. RTI2018-098085-B-C41 (MCIU/AEI/FEDER, UE). Open access was funded by Universitat Politècnica de València through Grant No. PAID-12-21.

<sup>1</sup>Throughout this paper, matrices and vectors are represented by upper- and lowercase boldface letters, respectively,  $(\cdot)^T$  stands for transpose,  $(\cdot)^H$  stands for conjugate transpose,  $\|\cdot\|$  for vector 2-norm,  $\mathbb{E}\{\cdot\}$  for expected value,  $\mathcal{F}\{\cdot\}$  denotes the discrete time Fourier transform (DTFT), and  $\mathbf{I}$  denotes the identity matrix.

<sup>1</sup>W. F. Druyvesteyn and R. M. Aarts, "Personal sound," *J. Acoust. Soc. Am.* **96**, 3281 (1994).

<sup>2</sup>T. Betlehem, W. Zhang, M. A. Poletti, and T. Abhayapala, "Personal sound zones: Delivering interface-free audio to multiple listeners," *IEEE Signal Process. Mag.* **32**(2), 81–91 (2015).

<sup>3</sup>B. Van Veen and K. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.* **5**(2), 4–24 (1988).

<sup>4</sup>E. Mabande and W. Kellermann, "Towards superdirective beamforming with loudspeaker arrays," in *Proceedings of the 19th International Congress on Acoustics*, Madrid, Spain, September 2017 (2007).

<sup>5</sup>T. Betlehem and P. D. Teal, "A constrained optimization approach for multi-zone surround sound," in *Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing*, Prague, Czech Republic, May 2011 (2011), pp. 437–440.

<sup>6</sup>Y. J. Wu and T. Abhayapala, "Spatial multizone soundfield reproduction: Theory and design," *IEEE Trans. Audio Speech Lang. Process.* **19**(6), 1711–1720 (2011).

<sup>7</sup>P. Coleman, P. J. B. Jackson, M. Olik, and J. A. Pedersen, "Personal audio with a planar bright zone," *J. Acoust. Soc. Am.* **136**, 1725–1735 (2014).

<sup>8</sup>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.* **128**, 121–131 (2010).

<sup>9</sup>Y. Cai, M. Wu, L. Liu, and J. Yang, "Time-domain acoustic contrast control design with response differential constraint in personal audio systems," *J. Acoust. Soc. Am.* **135**, 252–257 (2014).

<sup>10</sup>F. Olivieri, F. M. Fazi, S. Fontana, D. Menzies, and P. A. Nelson, "Generation of private sound with a circular loudspeaker array and the

weighted pressure matching method," *IEEE/ACM Trans. Audio, Speech Lang. Process.* **25**(8), 1579–1591 (2017).

<sup>11</sup>A. Canclini, D. Marković, M. Schneider, F. Antonacci, E. A. P. Habets, A. Walther, and A. Sarti, "A weighted least squares beam shaping technique for soundfield control," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, Calgary, Alberta, Canada, April 2018 (2018), pp. 6812–6816.

<sup>12</sup>J.-W. Choi and Y.-H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," *J. Acoust. Soc. Am.* **111**(4), 1695–1700 (2002).

<sup>13</sup>J.-H. Chang and F. Jacobsen, "Sound field control with a circular double-layer array of loudspeakers," *J. Acoust. Soc. Am.* **131**, 4518–4525 (2012).

<sup>14</sup>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.* **23**(11), 1869–1878 (2015).

<sup>15</sup>V. Molés-Cases, G. Piñero, M. De Diego, and A. Gonzalez, "Personal sound zones by subband filtering and time domain optimization," *IEEE/ACM Trans. Audio, Speech Lang. Process.* **28**, 2684–2696 (2020).

<sup>16</sup>J. Cheer and S. Elliott, "Design and implementation of a personal audio system in a car cabin," in *Proc. of the Meetings on Acoustics* (2013).

<sup>17</sup>M. Møller and M. Olsen, "Sound Zones: On Performance Prediction of Contrast Control Methods," in *Proc. of the AES International Conference on Sound Field Control* (2016).

<sup>18</sup>M. Olik, J. Francombe, P. Coleman, P. J. Jackson, M. Olsen, M. Møller, R. Mason, and S. Bech, "A comparative performance study of sound zoning methods in a reflective environment," in *Proc. of the AES International Conference* (2013).

<sup>19</sup>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. Proc.* **27**(12), 1957–1969 (2019).

<sup>20</sup>L. Vindrola, M. Melon, J.-C. Chamard, B. Gazengel, and G. Plantier, "Pressure matching with forced filters for personal sound zones application," in *Proc. of the AES 147th Convention* (2019).

<sup>21</sup>L. Vindrola, M. Melon, J.-C. Chamard, and B. Gazengel, "Pressure matching with forced filters for personal sound zones application," *J. Audio Eng. Soc.* **68**(11), 832–842 (2020).

<sup>22</sup>Q. Zhu, P. Coleman, M. Wu, and J. Yang, "Robust acoustic contrast control with reduced in-situ measurement by acoustic modeling," *J. Audio Eng. Soc.* **65**(6), 460–473 (2017).

<sup>23</sup>M. Ebri, N. Strozzi, F. M. Fazi, A. Farina, and L. Cattani, "Individual Listening Zone with Frequency-Dependent Trim of Measured Impulse Responses," in *Proc. Of the 149th AES Convention* (2020).

<sup>24</sup>H. Kuttruff, *Room Acoustics*, 5th ed. (Spon Press, London, 2009), p. 52.

<sup>25</sup>S. J. Elliott, J. Cheer, J.-W. Choi, and Y. Kim, "Robustness and regularization of personal audio systems," *IEEE/ACM Trans. Audio Speech Lang. Proc.* **20**(7), 2123–2133 (2012).

<sup>26</sup>M. F. Simón-Gálvez, S. J. Elliott, and J. Cheer, "The effect of reverberation on personal audio devices," *J. Acoust. Soc. Am.* **135**, 2654–2663 (2014).

<sup>27</sup>S. Spors and R. Rabenstein, "Spatial aliasing artifacts produced by linear and circular loudspeaker arrays used for wave field synthesis," in *Proc. of the 120th AES Convention*, Paris, France, May 2006 (2006).

<sup>28</sup>A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Proc. of the 108th AES Convention*, Paris, France, February 2000 (2000).

<sup>29</sup>P. Bloomfield, *Fourier analysis of time series: An introduction*, 2nd ed. (Wiley, New York, 1976), p. 69.

<sup>30</sup>C.-H. Jeong, "Kurtosis of room impulse responses as a diffuseness measure for reverberation chambers," *J. Acoust. Soc. Am.* **139**(5), 2833–2841 (2016).

<sup>31</sup>P. D. Hatziantoniou and J. N. Mourjopoulos, "Generalized fractional-octave smoothing of audio and acoustic responses," *J. Audio Eng. Soc.* **48**(4), 259–280 (2000).