Weighted pressure matching with windowed targets for personal sound zones

Vicent Molés-Cases,[1,a] Stephen J. Elliott,[2] Jordan Cheer,[2] Gema Piñero,[1] and Alberto Gonzalez[1]
[1] *Institute of Telecommunications and Multimedia Applications, Universitat Politècnica de València, València 46022, Spain*
[2] *Institute of Sound and Vibration Research, University of Southampton, Southampton SO17 1BJ, United Kingdom*

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.

[https://doi.org(DOI number)]

[XYZ] Pages: 1–12

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[1,2]. 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[3,4], soundfield synthesis[5–7], energy cancellation approaches[8,9], or hybrid approaches[10,11]. Among these, Acoustic Contrast Control (ACC)[12] 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[12]. However, ACC can not synthesize a specific target response in the bright zone. To solve this limitation, the Weighted Pressure Matching (WPM) algorithm has been proposed[13] , 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[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[14], subband-domain[15] and frequency-domain[16]. 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 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)[13] . The performance of ACC and WPM has been compared under reverberant conditions in[17,18]. The target response selected for the bright zone is a key choice for WPM systems, as different targets 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[14–16,19–21]. 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. We therefore propose a variation of the WPM approach, in which a window function

[a] vimoca3@iteam.upv.es

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. 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 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, Section V summarizes the main conclusions.

Notation: Throughout this paper 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, $\mathbb{E}\{\cdot\}$ for expected value, $\mathcal{F}\{\cdot\}$ denotes the Discrete Time Fourier Transform and $\mathbf{I}$ denotes the identity matrix.

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

$$\mathbf{g} = \begin{bmatrix} G_0 & \dots & G_{L-1} \end{bmatrix}^T . \quad (2)$$

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

$$\mathbf{d}_b = \begin{bmatrix} D_{0,b} & \dots & D_{M_b-1,b} \end{bmatrix}^T . \quad (4)$$

Once the model is presented, we describe the Weighting Pressure Matching (WPM) algorithm, which was originally proposed by[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 (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; 3) the energy of the filter coefficients. In (5), λ is a regularization factor that constrains the energy of the filters. Also, κ is weighting factor satisfying $0 \leq \kappa \leq 1$ that is used to balance the solution, e.g., high values of κ 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 straight forward to show that the optimal filter coefficients that minimize (5) are given by

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

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 as the delayed response from one of the loudspeakers to all of the control points in the bright zone[14–16,19–21], i.e.,

$$D_{m,b} = H_{ml_r,b} \, e^{-j2\pi f \tau} , \quad (7)$$

where $l_r \in \{0, \dots, L-1\}$ is the index of the reference loudspeaker, and τ is a modeling delay that assures the causality of the filters. The previous selection aims to synthesize the direct propagation component and all the reverberant components produced by the reference loudspeaker in the bright zone. Now, let us define

$$h_{ml,b}^{(L_w)}(n) = \begin{cases} w_{L_w}(n-\tau_p) \, h_{ml,b}(n) & \text{if } L_w < \infty \\ h_{ml,b}(n) & \text{otherwise} \end{cases} , \quad (8)$$

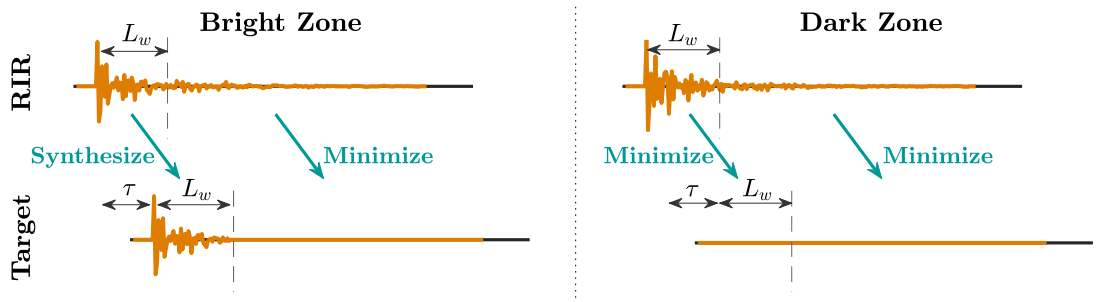


FIG. 1. Schematic to illustrate the effect produced by windowing the target. The upper plots represent the RIR between the reference loudspeaker and one control point in the bright an dark zones, and the lower plots represent the target at these control points using a window of length L_w.

where $h_{ml,q}$ is the RIR between the l-th loudspeaker and the m-th control point in the bright zone, τ_p is the propagation delay corresponding to the direct component of $h_{ml,b}$, and w_{L_w} is a window function defined as $w_{L_w}(n) = 0$ for $|n| \geq L_w$. It is important to highlight that in the previous definition of the window, we assumed that it is centered on $n = 0$. In (8), the window is time-shifted such that its center is located on $n = \tau_p$. Moreover, from the definition of the window 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). The authors in[22] proposed to use windowed RIRs in $\mathbf{H}_b$ and $\mathbf{H}_d$ to compute the filters for the ACC algorithm in order to improve the robustness to perturbations in the environment. In our case, we propose to use windows on the target responses in $\mathbf{d}_b$ in the WPM algorithm to improve the overall performance, so that these become

$$D_{m,b} = H^{(L_w)}_{ml_r,b}\, e^{-j2\pi f\tau}\ , \tag{9}$$

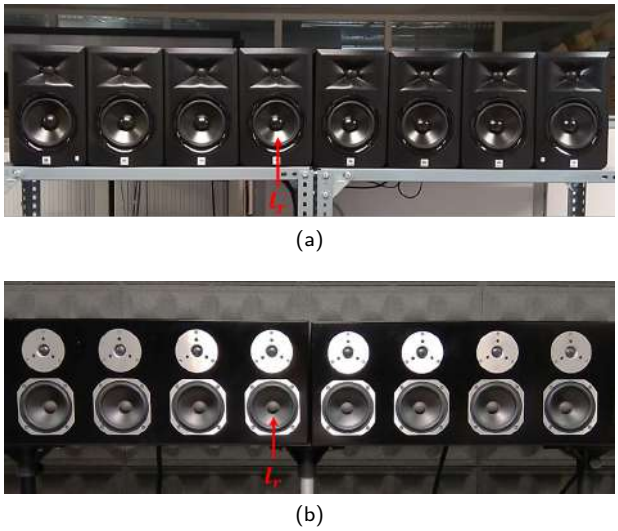
where $H^{(L_w)}_{ml_r,b} = \mathcal{F}\left\{h^{(L_w)}_{ml_r,b}\right\}$. It is important to note that (7) and (9) are equivalent when $L_w = \infty$. However, (9) has the advantage that, by selecting the window length, we can choose which reverberant components we want to synthesize and which to minimize 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 of the RIR 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.

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[23]. 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[24], 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 $\left(\mathbf{H}^H_{d,dif}\mathbf{H}_{d,dif}\right)^{-1}\mathbf{H}^H_{b,dif}\mathbf{H}_{b,dif}$. For the diffuse components, we can assume that $\mathbf{H}^H_{b,dif}\mathbf{H}_{b,dif} \approx \sigma\mathbf{I}$ and $\mathbf{H}^H_{d,dif}\mathbf{H}_{d,dif} \approx \sigma\mathbf{I}$ (where σ is the energy of the diffuse field in the room)[25]. 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 Section 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 win-

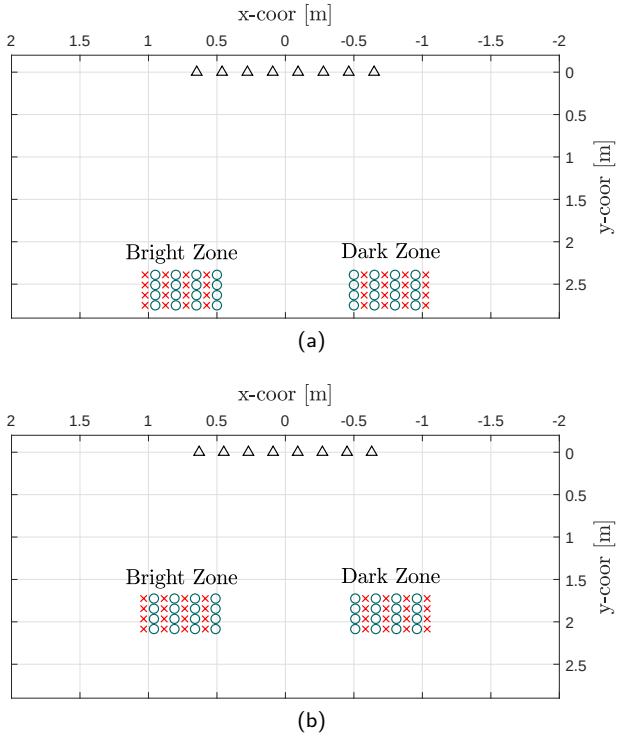
dowing the target response for the bright zone. Finally, we evaluate the robustness to perturbations in the filters computed with a windowed target.



(a)

(b)

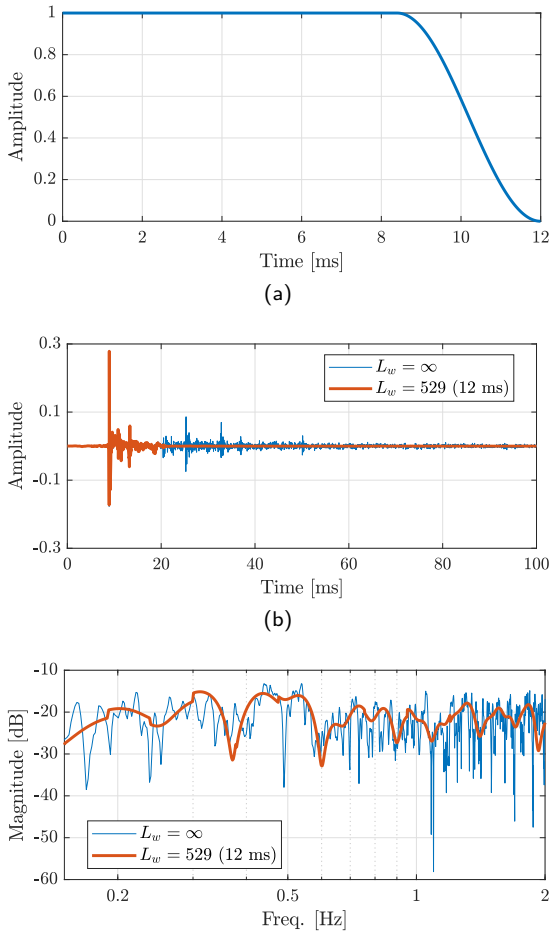
FIG. 2. Array of loudspeakers used in scenario 1 and 2 in a) and b), respectively.



(a)

(b)

FIG. 3. Setups for scenario 1 and 2 in a) and b), respectively. Markers ○ and × denote control and validation points, respectively, and ▵ denotes a loudspeaker. 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.



(a)

(b)

(c)

FIG. 4. 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 microphone of the bright zone for scenario 1 with different window lengths in b) and c), respectively.

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. 2a). The setup is formed by one bright and one dark zone (FIG. 3a). 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.

- *Scenario 2*: Listening room at iTEAM, which is a rectangular room of size 9.07 × 4.45 × 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. 2b) has been used and two zones are considered, which are spatially sampled using two different grids of microphones (see FIG. 3b).

In all cases, the RIRs were measured using the exponentially-swept sine technique[26] with a sampling frequency of 44100 Hz. In the following, we use notation $\overline{\cdot}$ to denote the elements related with 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 (6), where the weighting parameter κ is set to 0.5 if not indicated otherwise. Moreover, the regularization parameter λ is selected such that the array effort (which will be defined next in (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 (9), where we select w_{L_w} as a Tukey window with cosine fraction α=0.3[27] and the reference loudspeaker is $l_r = 3$ (as indicated in FIG. 2a and FIG. 2b). Moreover, when $L_w < \infty$, we apply a 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. 4a, we show an example of the causal part of a Tukey window with L_w=529 (12 ms), and FIG. 4b and FIG. 4c show examples of a target impulse response and a frequency response in the 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

$$\varepsilon_b = \left\| \overline{\mathbf{H}}_b \mathbf{g} - \overline{\mathbf{d}}_b \right\|^2 , \tag{10}$$

where the target is selected using the same criterion used to compute the filters, meaning that the same window w_{L_w} is used to compute and to evaluate the filters. Next, we define the mean energy in the dark zone as

$$E_d = \left\| \overline{\mathbf{H}}_d \mathbf{g} \right\|^2 , \tag{11}$$

and the array effort as

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

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 with the three components of the cost function (5), and whose influence can be adjusted with λ and κ. Now, we define the Acoustic Contrast (AC) as

$$C = \frac{\left\| \overline{\mathbf{H}}_b \mathbf{g} \right\|^2}{\left\| \overline{\mathbf{H}}_d \mathbf{g} \right\|^2} , \tag{13}$$

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 (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

$$K(n) = \frac{1}{N_h} \sum_{\forall m, \forall l, \forall q} \left(\frac{\mathbb{E}\left\{ \left(\overline{h}_{ml,q} - \mu^{(n)}_{ml,q} \right)^4 \right\}}{\left(\sigma^{(n)}_{ml,q} \right)^4} - 3 \right) \tag{14}$$

where $\mu^{(n)}_{ml,q}$ and $\sigma^{(n)}_{ml,q}$ are the mean value and the standard deviation of $\overline{h}_{ml,q}$, respectively, over the interval $n, \ldots, n+L_s-1$, and the expected value for time n is computed over the interval $n, \ldots, n+L_s-1$. In (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[28] show that the Kurtosis is closely related with 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[29] for all frequency-domain plots to improve the readability of the results.

C. Effect of L_w in scenario 1

In this subsection, we present experimental results to show the influence of the window length L_w on the performance for scenario 1. Firstly, we show in FIG. 5 the performance as a function of L_w in terms of: mean energy in the dark zone (in FIG. 5a,5d), MSE in the bright zone (in FIG. 5b,5e), and array effort (in FIG. 5c,5f). In FIG. 5, the metrics in the top and bottom rows are computed with $AE_{max} = 0$ dB and $AE_{max} = 15$ dB, respectively. We can see that the performance for both effort constraints is equal for frequencies above 800 Hz. The motivation is that the matrix that must be inverted in (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

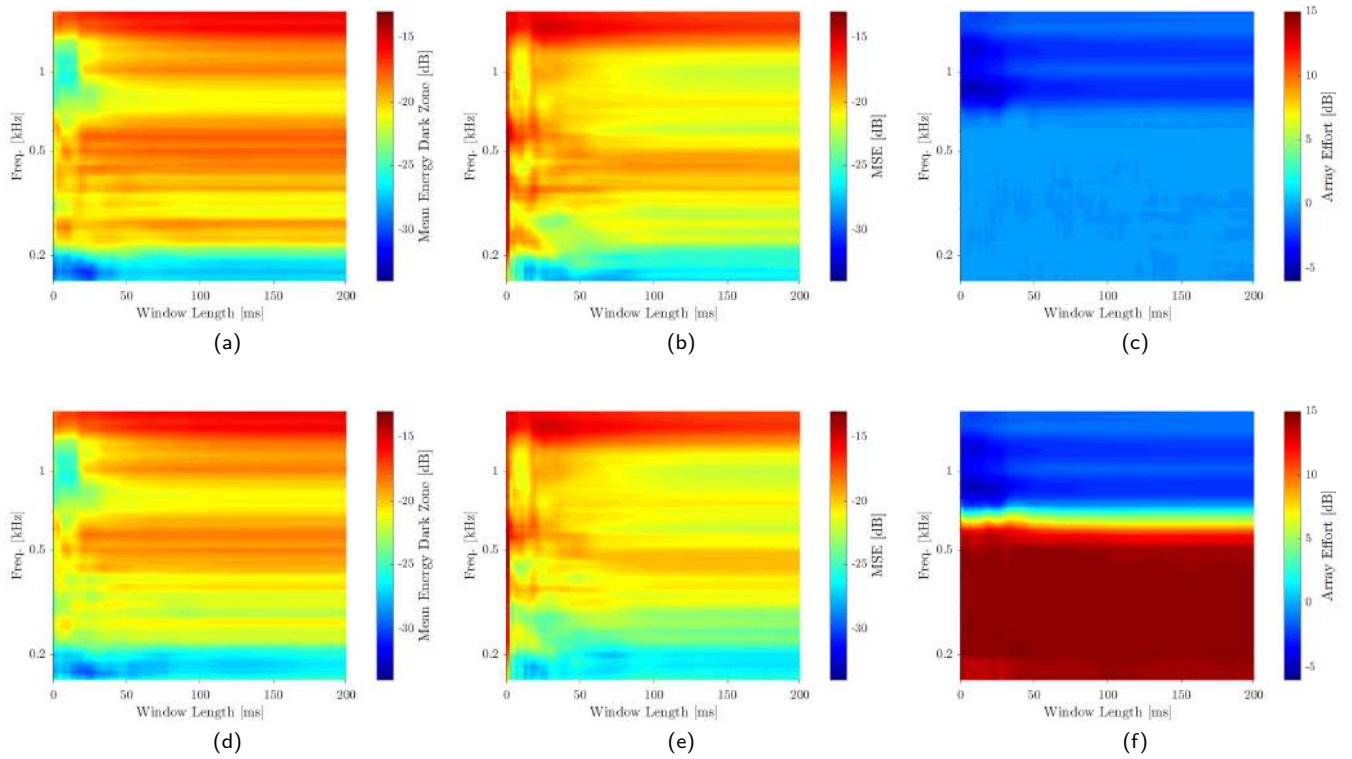


FIG. 5. (color online) Performance as a function of the window length and frequency for scenario 1 in terms of: mean energy in the dark zone (a, d), MSE in the bright zone (b, e), and array effort (c, f). For the top row figures $AE_{max} = 0$ dB, and $AE_{max} = 15$ dB for the bottom figures.

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 FIG. 5b,5e 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 FIG. 5a,5d). This fact indicates that if the window is too short it can degrade the performance. 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. Moreover, for frequencies in the range 150-200 Hz, 400-500 Hz and above 700 Hz these improvements 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, lower energy in the dark zone may be achieved for long windows by tuning the weighting parameter κ, 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 in FIG. 6, where we compare the following configurations:

- Configuration 1: $L_w = 529$ (12 ms) and $\kappa = 0.5$.
- Configuration 2: $L_w = \infty$ and $\kappa = 0.5$.
- Configuration 3: $L_w = \infty$ with frequency dependent κ. For each frequency bin, we search the weighting factor κ (as shown in FIG. 7) that leads to the same MSE as configuration 1.

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 configuration 3 in the comparison. The performance of the three configurations is shown in FIG. 6 in terms of: mean energy in the dark zone (in FIG. 6a), MSE in the bright zone (in FIG. 6b), array effort (in FIG. 6c), and acoustic contrast (in FIG. 6d). 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 configuration 1 can offer lower energy in the dark zone than configuration 2 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 κ). Next,

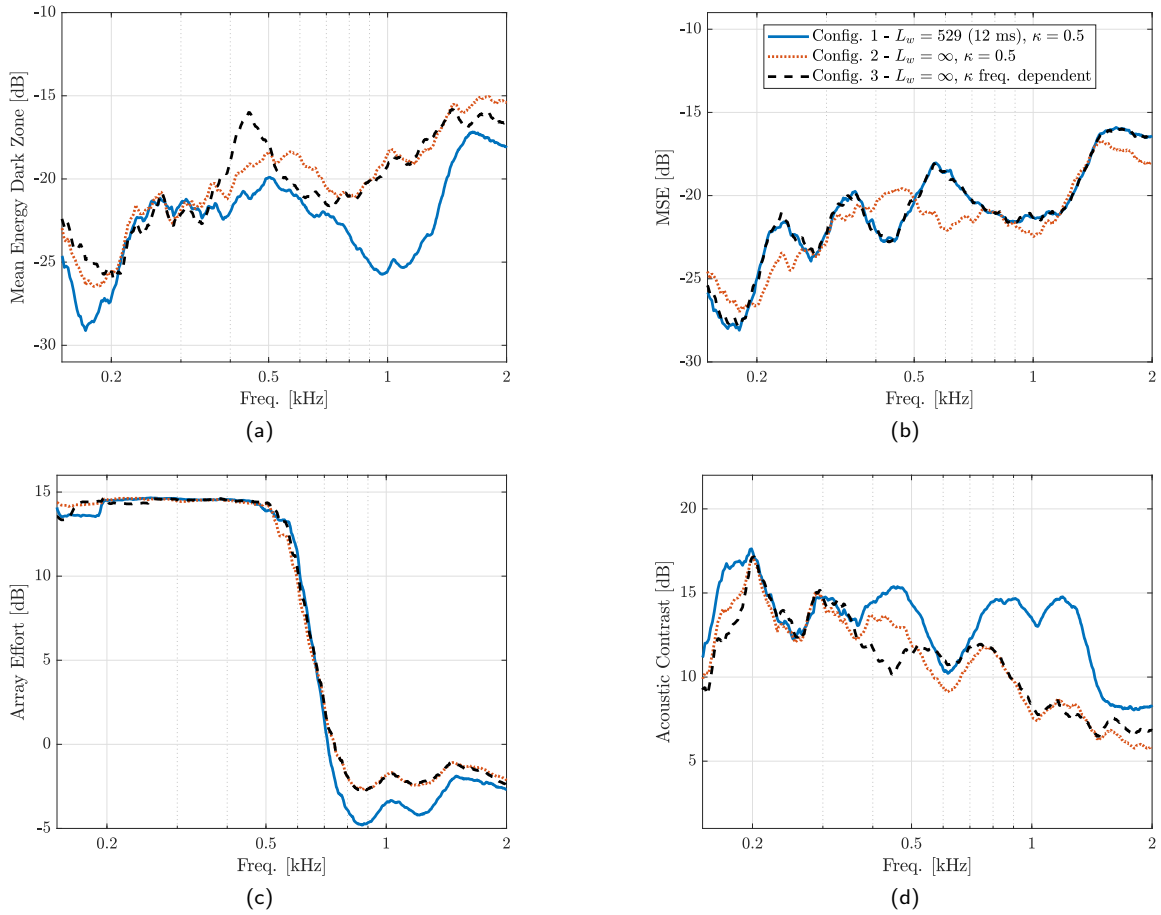


FIG. 6. Performance for three different configurations as a function of frequency for scenario 1 in terms of: (a) mean energy in the dark zone, (b) MSE in the bright zone, (c) array effort, and (d) acoustic contrast. $AE_{max} = 15$ dB is considered.

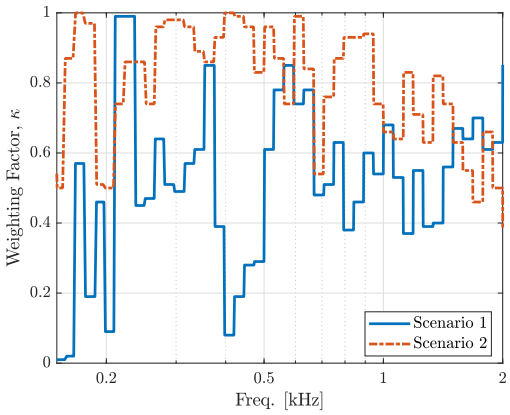


FIG. 7. Weighting factor κ used for configuration 3 in FIG. 6 and FIG. 10 for scenarios 1 and 2, respectively.

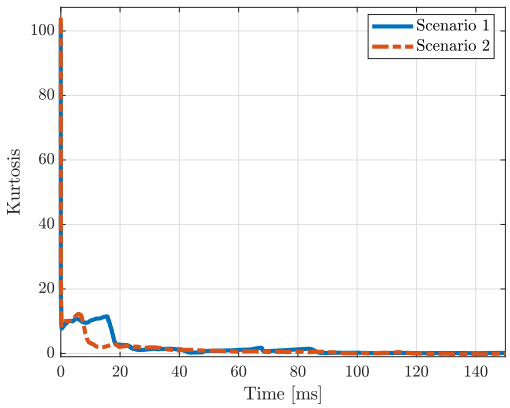


FIG. 8. Mean Kurtosis for scenarios 1 and 2.

if we compare configuration 1 and 3, we can see that configuration 1 leads to lower energy in the dark zone for all the studied frequencies except for the band 200-360 Hz, where configuration 3 presents slightly lower energy. For example, configuration 1 leads to almost 6 dB lower energy in the dark zone than configuration 3 around 1 kHz, and 5 dB lower energy around 180 Hz. The acoustic contrast in FIG. 6d follows the same trends, since for the same MSE level, configuration 1 can offer higher contrast than configuration 3 across almost all of the stud-

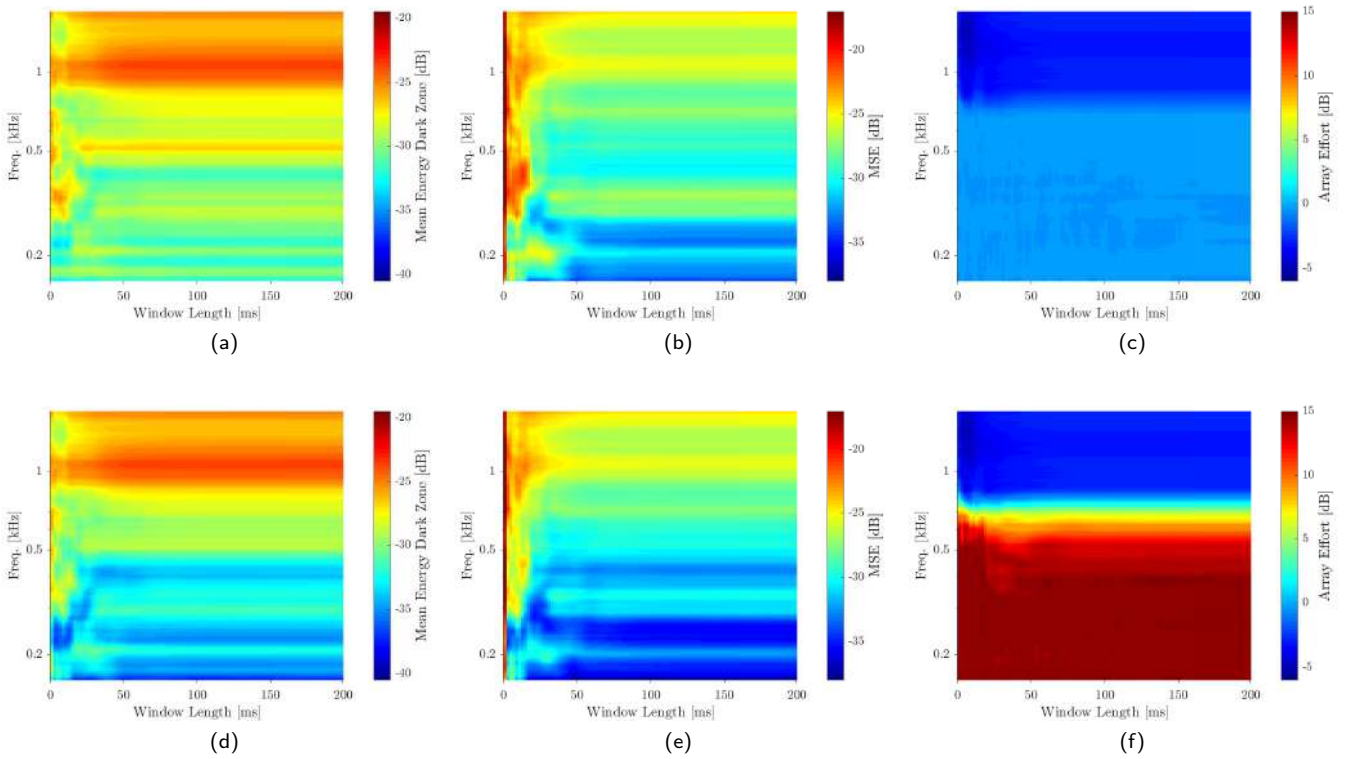


FIG. 9. (color online) Performance as a function of the window length and frequency for scenario 2 in terms of: mean energy in the dark zone (a, d), MSE in the bright zone (b, e), and array effort (c, f). For the top row figures $AE_{max} = 0$ dB, and $AE_{max} = 15$ dB for the bottom figures.

ied frequencies (with improvements of more than 5 dB). Moreover, the array effort required by configuration 1 is lower than for configuration 2 and 3 for frequencies above 700 Hz. The MSE is almost the same for configuration 1 and 3, as expected, and is broadly similar for configuration 2, 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. 8 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 a small value after about 20 ms, indicating that the reverberant components of the RIRs can be assumed diffuse, with a Gaussian PDF, after this time[28]. 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.

D. Effect of L_w in scenario 2

In this subsection, we present experimental results to illustrate the influence of the window length L_w on the performance for scenario 2, with the lower reverberation time. FIG. 9 shows the performance as a function of L_w in terms of: mean energy in the dark zone (in FIG. 9a,9d), MSE in the bright zone (in FIG. 9b,9e), and array effort (in FIG. 9c,9f). In FIG. 9, 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 effort 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 FIG. 9b,9e 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 FIG. 9a,9d). This indicates that 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 FIG. 9b,9e 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.

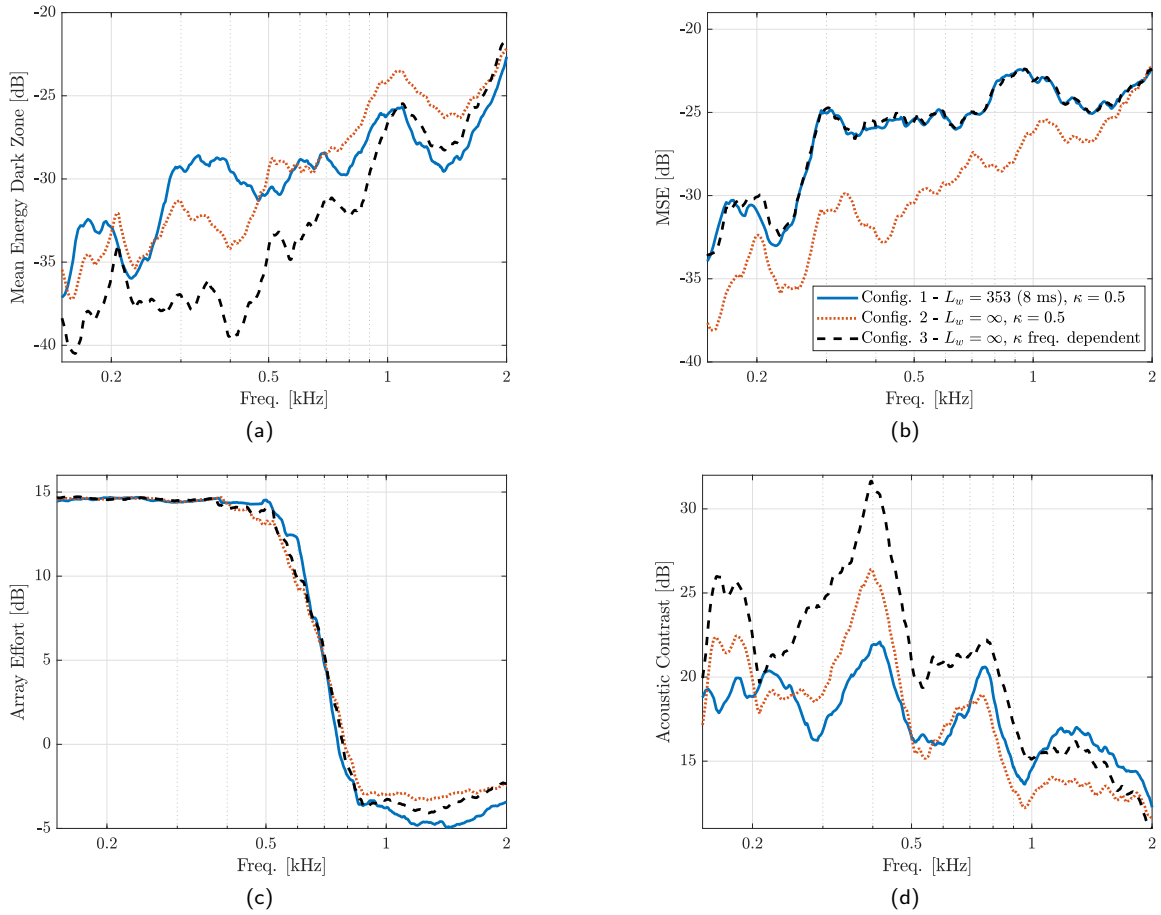


FIG. 10. Performance for three different configurations as a function of frequency for scenario 2 in terms of: (a) mean energy in the dark zone, (b) MSE in the bright zone, (c) array effort, and (d) acoustic contrast. $AE_{max} = 15$ dB is considered.

In order to further study if windowing the target offers performance improvements in this scenario, we present additional results in FIG. 10 where we compare three configurations which are similar to those used above:

- Configuration 1: $L_w = 353$ (8 ms) and $\kappa = 0.5$.
- Configuration 2: $L_w = \infty$ and $\kappa = 0.5$.
- Configuration 3: $L_w = \infty$ with frequency dependent κ. For each frequency bin, we search the weighting factor κ (shown in FIG. 7) that leads to the same MSE as configuration 1.

We again include configuration 3 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. 10 in terms of: mean energy in the dark zone (in FIG. 10a), MSE in the bright zone (in FIG. 10b), array effort (in FIG. 10c), and acoustic contrast (in FIG. 10d). 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 configuration 1 than for configuration 2, but even lower than for configuration 3 above 1 kHz. In this case, however, configuration 2 has a significantly lower MSE than configuration 1 or 3. 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 Section 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. 8 shows the mean Kurtosis of the RIRs for scenario 2, which has high values for times smaller than 10 ms, while it is much smaller 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. 9, again illustrating the use of the Kurtosis in estimating the optimum window length, although only at higher frequencies in this case with a short reverberation time.

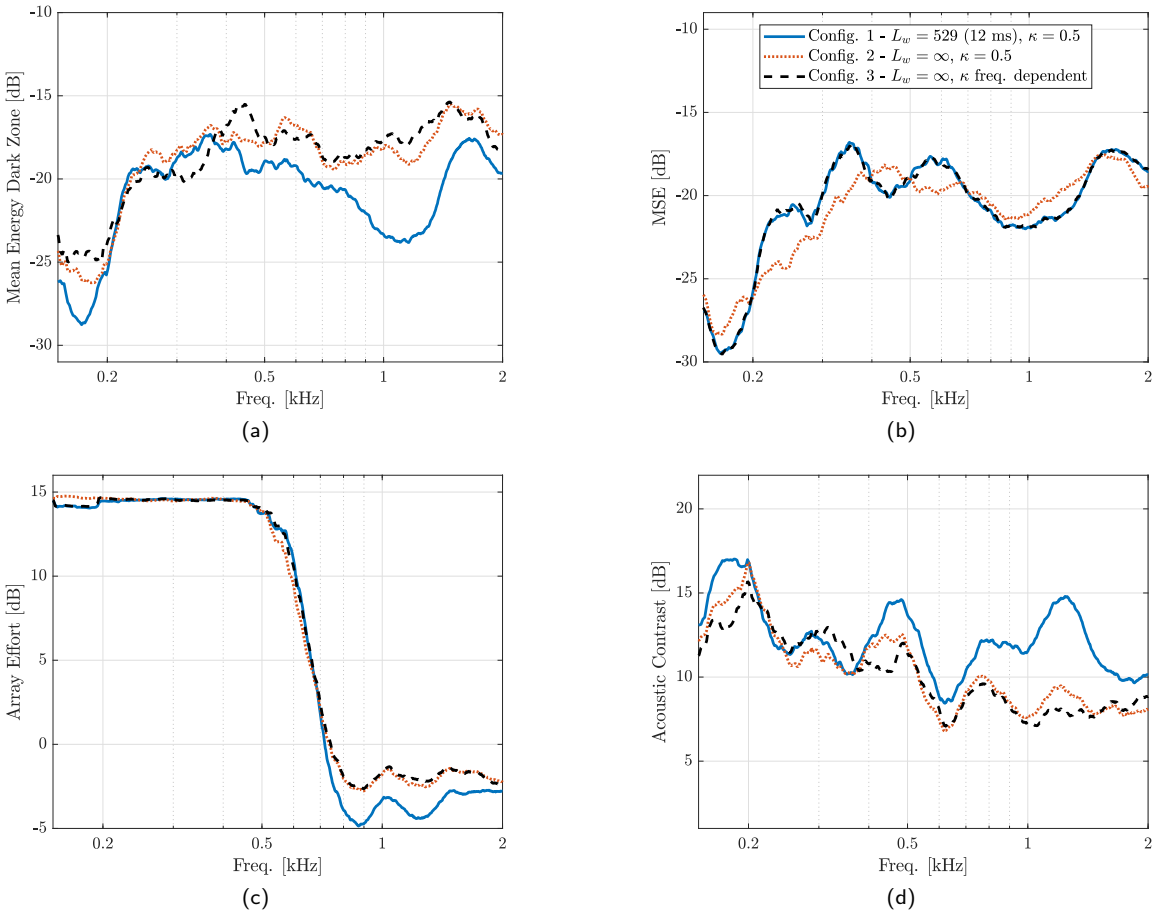


FIG. 11. Performance for three different configurations as a function of frequency for scenario 1 with perturbations in terms of: (a) mean energy in the dark zone, (b) MSE in the bright zone, (c) array effort, and (d) acoustic contrast. $AE_{max} = 15$ dB is considered.

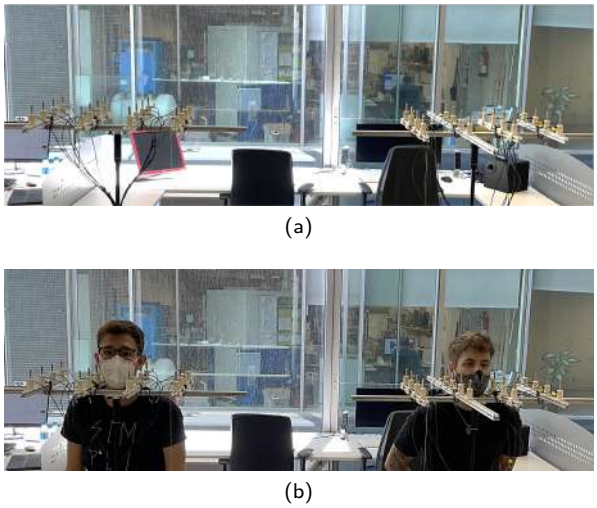


FIG. 12. Setup used to measure the RIRs without perturbations in a), and to measure the RIRs with the perturbations produced by two persons within the zones in b).

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. 11 evaluations results for scenario 1, where the filters are computed with the RIRs measured at the control points without any perturbation (as in FIG. 12a) 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. 12b). The control filters are those calculated without any perturbations, as in Section IV C, but now the performance, as shown in FIG. 11, has been calculated after these perturbations in the RIR. The effect of the perturbations can thus be evaluated by comparing the results in FIG. 11 with those in FIG. 6. 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 configuration 1 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 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 with 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 FPU17/01288. Stephen 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 RTI2018-098085-B-C41 (MCIU/AEI/FEDER, UE).

[1]W. F. Druyvesteyn and R. M. Aarts, "Personal sound," J. Acoust. Soc. Am. **45**(9), 685–701 (1997).

[2]T. Betlehem, W. Zhang, M. A. Poletti, and T. Abhayapala, "Personal sound zones: Delivering interface-free audio to multiple listeners," IEEE Signal Processing Magazine **32**(2), 81–91 (2015).

[3]B. Van Veen and K. Buckley, "Beamforming: a versatile approach to spatial filtering," IEEE ASSP Magazine **5**(2), 4–24 (1988).

[4]E. Mabande and W. Kellermann, "Towards superdirective beamforming with loudspeaker arrays," in *Proc. of the 19th International Congress on Acoustics* (2007).

[5]T. Betlehem and P. D. Teal, "A constrained optimization approach for multi-zone surround sound," in *Proc. of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing* (2011), pp. 437–440.

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

[7]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).

[8]M. Shin, S. Q. Lee, F. M. Fazi, P. A. Nelson, D. Kim, S. Wang, K. Ho Park, and J. Seo, "Maximization of acoustic energy difference between two spaces," J. Acoust. Soc. Am. **128**, 121–131 (2010).

[9]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).

[10]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, and Lang. Process. **25**(8), 1579–1591 (2017).

[11]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 sound field control," in *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing* (2018), pp. 6812–6816.

[12]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).

[13]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).

[14]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, and Lang. Process. **23**(11), 1869–1878 (2015).

[15]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, and Lang. Process. **28**, 2684–2696 (2020).

[16]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).

[17]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).

[18]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).

[19]M. Schneider and E. A. P. Habets, "Iterative dft-domain inverse filter optimization using a weighted least-squares criterion," IEEE/ACM Trans. Audio, Speech, and Lang. Process. **27**(12), 1957–1969 (2019).

[20]L. Vindrola, M. Melon, J.-C. Chamard, B. Gazengel, and G. Plantier, "Personal sound zones: a comparison between frequency and time domain formulations in a transportation context," in *Proc. of the AES 147th Convention* (2019).

[21]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).

[22]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 Audio Engineering Society Convention* (2020).

[23]H. Kuttruff, *Room acoustics*, 5th ed. (Spon Press, London, 2009), p. 52.

[24]S. J. Elliott, J. Cheer, J.-W. Choi, and Y. Kim, "Robustness and regularization of personal audio systems," IEEE/ACM Trans. Audio, Speech, and Lang. Process. **20**(7), 2123–2133 (2012).

[25]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).

[26]A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Proc. of the 108th AES Convention* (2000).

[27]P. Bloomfield, *Fourier Analysis of Time Series: An Introduction*, 2nd ed. (Wiley, New York, 1976), p. 69.

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

[29]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).