OPEN ACCESS Freely available online

CC BY

J. Cadavid, M. Bo Møller, T. Van Waterschoot, S. Bech, and J. Østergaard, "Effects of Reduced Information in the Performance of Low-Frequency Sound Zones," *J. Audio Eng. Soc.*, vol. 73, no. 7/8, pp. 429–445 (2025 Jul./Aug.). https://doi.org/10.17743/jaes.2022.0207.

Effects of Reduced Information in the Performance of Low-Frequency Sound Zones

JOSÉ CADAVID,[1,*] *AES Student Member*, **MARTIN BO MØLLER,**[2]
(jmct@es.aau.dk) (mim@bang-olufsen.dk)

TOON VAN WATERSCHOOT,[3] **SØREN BECH,**[1,2] *AES Fellow* **AND JAN ØSTERGAARD**[1]
(toon.vanwaterschoot@esat.kuleuven.be) (sbe@es.aau.dk) (jo@es.aau.dk)

[1]*Department of Electronic Systems, Aalborg University, Aalborg, Denmark*
[2]*Bang & Olufsen, Struer, Denmark*
[3]*Department of Electrical Engineering, KU Leuven, Leuven, Belgium*

Sound zone techniques enable separate audio playback at distinct listening areas in rooms. This requires using loudspeaker arrays, multiple control microphones per zone, and carefully designed control filters. The filters are based on room impulse responses between all loudspeaker-microphone pairs that, in some cases, can be obtained with very short acquisition times. Low frequencies, featuring longer wavelengths, allow for fewer spatial sampling points inside the sound zones, simplifying the setup and decreasing data processing requirements. This study evaluates the effects of reducing room impulse responses acquisition times and sampling points in low-frequency sound zone rendering. Different combinations of both strategies and multiple spatial sampling arrangements were explored. Their performance was evaluated in two acoustic conditions under objective metrics, and the area of influence of the system around the sound zones was also studied. Both strategies were found effective in reducing the amount of information included while performing as good as cases comprising all information available. For instance, under short reverberation time, four microphones and 150-ms acquisition time performed comparably to 15 microphones and 16-s acquisition time. Using more sampling points reduced the effects of their arrangement in the performance. Both strategies resulted in faster control filter calculations, suggesting improved efficiency.

0 INTRODUCTION

Multiple situations require reproducing different sounds in designated areas of a room, called bright zones, and generating low sound pressure levels in other areas, known as dark zones. This selective audio reproduction is made through dedicated control filters applied to an array of loudspeakers. Such filters are based on precise information about the transfer functions between the loudspeakers and spatial sampling points in these sound zones, usually obtained in the form of room impulse responses (RIRs) [1]. To achieve a high signal-to-noise ratio and a wide frequency range, measurement signals lasting several seconds are generally used [2]. Additionally, obtaining enough data about large areas also demands a higher number of sampling points, which result in a more precise description of the sound field [3].

The need for these RIRs poses one of the biggest limitations of current sound zone methods, translated in their incapability to adapt to changes. More precisely, successful sound zones rendering involves very precise room acoustics information, and thus, any change in the environment and/or the system will affect the resulting performance [4, 5]. Therefore, if changes are to be accounted for, new measurements would have to be performed and new control filters computed. This makes it necessary to implement fast measurement and calculation processes.

For fast RIR acquisition, Cadavid et al. [6] proposed using very short measurement signals that, compared to traditionally longer measurements, still achieve high sound zones performance. The second requirement, fast processing, has been approached both from the computational and the information points of view. The former focuses on the efficiency of the algorithms (e.g., by means of frequency domain or subbands implementations) or by using reduced-rank matrices for simplified, successful solutions [7–9]. The latter consists in reducing the information used to per-

*To whom correspondence should be addressed, e-mail: jmct@es.aau.dk.

form the computations by selectively excluding parts of the RIRs [10–12] or by including fewer sampling points. As an example, Elliott et al. [13] showed that a single microphone allows controlling the sound field over a region with a one-tenth of the wavelength radius. This single-point concept has been applied in active noise control to cancel low-frequency noise inside a car cabin [14].

These ideas have also been applied for sound zones, where fewer control points were used for personal head-tracked systems [15] or to extend their region of influence by sound field interpolation [16]. However, little research has been done on spatial sampling and/or reduced information specifically for low-frequency sound zones. The focus on this frequency range arises from the different strategies required to render sound zones over the frequency spectrum: whereas the size of a linear array of sources becomes unpractical in the low frequencies, middle and high frequencies are more conveniently reproduced, respectively, by linear arrays and highly directive sources [17, 18]. Hence, low-frequency sound fields can be better controlled through an array of sources distributed over the room. Further research in this range of frequencies under realistic acoustic conditions aims to improve the reproduction of full-band audio signals, which would positively impact several applications in domestic, automotive, and health care environments.

The effects of using fewer sampling points and shorter RIR acquisition times in the performance of low-frequency sound zones was partially explored by the authors in a previous work. Presented at the 156th Convention of the Audio Engineering Society [19], the paper also introduces a data set of RIRs for low-frequency sound field control. These RIRs were acquired for the range of 30 Hz to 600 Hz, with different acquisition times, in two rooms with contrasting acoustic conditions. Based on the aforementioned publication and using the related data set, this work revises and extends the original analyses, providing a more comprehensive understanding of the aforementioned strategies for expedited sound zones rendering. More specifically, it aims to explore the influence of these strategies when jointly applied in the sound zones context, taking into account the effects of the frequency range of interest.

In line with the original work, the performance has been evaluated under objective metrics describing primarily the acoustic performance, splitting the RIRs of the data set in two subsets: $\mathrm{R}^{\mathbf{d}}$, exclusively for designing the control filters; and $\mathrm{R}^{\mathbf{e}}$, exclusively for evaluating their performance. This is to avoid using the same information to design and evaluate the sound zones, which, as observed by Møller et al., could yield misleading results [3]. Additionally, following the observation by Elliot et al. [13], the region the system can inherently control given the range of frequencies is also evaluated. Finally, previous experiments on the impact of the strategies in the computational requirements have also been revised.

The methods implemented in this work to design and evaluate sound zones are described in SEC. 1. Details about the use of the data set and about tests and simulations are included in SEC. 2. Results are presented in SEC. 3 and discussed in SEC. 4. Overall conclusions and ideas for further work are summarized in SEC. 5.

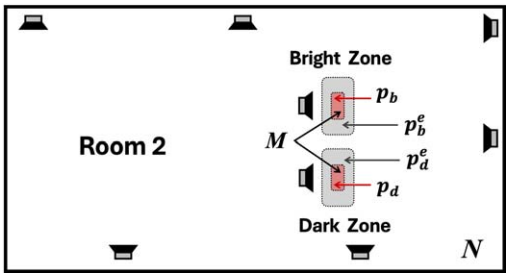


Fig. 1. Diagram of system setup to create two sound zones as deployed in room 2 of the data set in [19] (see SEC. 2.1). Sampled areas are as shown in Fig. 2, where the bright and dark zones are marked in red. Their pressures are denoted, respectively, $\mathbf{p}_b$ and $\mathbf{p}_d$, and the corresponding evaluation pressures are $\mathbf{p}_b^{\mathbf{e}}$ and $\mathbf{p}_d^{\mathbf{e}}$.

1 WEIGHTED LEAST SQUARES METHOD FOR SOUND ZONES RENDERING

To combine the advantages of different approaches for control filter design, Chang and Jacobsen [20] proposed the so-called weighted least squares (WLS) method. By setting the value of the weighting factor $0 \leq \beta \leq 1$, the method allows to trade-off between two objectives: reducing the pressure in the dark zone $\mathbf{p}_d$ through destructive interference, and matching the pressure in the bright zone $\mathbf{p}_b$ and a predefined reference pressure $\mathbf{p}_r$ by minimizing the difference between them.

A system as the one depicted in Fig. 1 allows creating two sound zones by controlling the sound generated by the N loudspeakers. To obtain information about the response of the room at the zones, each one is spatially sampled by a set of M microphones, resulting in $2NM$ RIRs, where

$$\mathbf{h}^{\mathbf{d}}{}_{mn} = \left[h^d_{mn}(0) \;\; h^d_{mn}(1) \;\; \cdots \;\; h^d_{mn}(J-1)\right]^{\mathrm{T}} \tag{1}$$

is the $J \times 1$ RIR of the microphone-loudspeaker pair mn. These RIRs compose the set $\mathrm{R}^{\mathbf{d}}$, used in this work to design both the reference pressure $\mathbf{p}_r$, and the control filters

$$\mathbf{w} = \left[\mathbf{w}_1^{\mathrm{T}} \;\; \mathbf{w}_2^{\mathrm{T}} \;\; \cdots \;\; \mathbf{w}_N^{\mathrm{T}}\right]^{\mathrm{T}} \tag{2}$$

with the n-th $I \times 1$ individual filter

$$\mathbf{w}_n = \left[w_n(0) \;\; w_n(1) \;\; \cdots \;\; w_n(I-1)\right]^{\mathrm{T}}. \tag{3}$$

When applying the N control filters to the loudspeakers, the interaction of the reproduced signals is expected to render the intended sound zones. For this, an optimization problem can be formulated with the cost function

$$\begin{aligned}\mathrm{J}_{\mathrm{WLS}}(\mathbf{w}) = {} & (1-\beta)[(\mathbf{H}_b\mathbf{w} - \mathbf{p}_r)^{\mathrm{T}}(\mathbf{H}_b\mathbf{w} - \mathbf{p}_r)] \\ & + \beta(\mathbf{H}_d\mathbf{w})^{\mathrm{T}}\mathbf{H}_d\mathbf{w},\end{aligned} \tag{4}$$

as defined by Gálvez et al. [21] for the WLS method in the time domain, with a Dirac impulse input signal. Note on the one hand that the first term is the square of the error resulting from matching $\mathbf{p}_r$ and $\mathbf{p}_b$ weighted by $(1-\beta)$. The second term is the square of the error in the dark zone

weighted by β, quantifying the difference between the pressure in the dark zone $\mathbf{p}_d$ and an implicit reference pressure equal to zero. For $\beta = 0$, the system will only attempt to match $\mathbf{p}_r$ and $\mathbf{p}_b$, resulting in an existing method known as pressure matching [22]. And, for $\beta = 1$, the only goal will be to reduce $\mathbf{p}_d$, resembling traditional active noise control. On the other hand, for the chosen input signal, the pressures $\mathbf{p}_b$, $\mathbf{p}_d$, and $\mathbf{p}_r$ have been expressed as vectors of size $M(J + I - 1) \times 1$. They contain the respective pressure signals of the M microphones on each zone: $\mathbf{p}_{b_m}$, $\mathbf{p}_{d_m}$, and $\mathbf{p}_{r_m}$.

Pressures $\mathbf{p}_b$ and $\mathbf{p}_d$ correspond to the convolution between the control filters and the RIRs in the bright and dark zones, respectively. These RIRs are arranged in $\mathbf{H}_b$ and $\mathbf{H}_d$, where for the corresponding RIRs,

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{11} & \cdots & \mathbf{H}_{1N} \\ \vdots & \mathbf{H}_{mn} & \vdots \\ \mathbf{H}_{M1} & \cdots & \mathbf{H}_{MN} \end{bmatrix} \tag{5}$$

are an $M(J + I - 1) \times LI$ block matrix composed by the $(J + I - 1) \times I$ convolutional matrices

$$\mathbf{H}_{mn} = \begin{bmatrix} h^d_{mn}(0) & 0 & 0 \\ \vdots & \ddots & 0 \\ h^d_{mn}(J-1) & \ddots & h^d_{mn}(0) \\ 0 & \ddots & \vdots \\ 0 & 0 & h^d_{mn}(J-1) \end{bmatrix}. \tag{6}$$

Additionally, the reference pressure $\mathbf{p}_r$ uses the RIRs of a reference source and, instead of the control filters, the input signal is filtered by a delayed impulse to introduce a modeling delay τ_{mod} [21].

To improve the performance of the system, three additional constraints were added to $\mathrm{J_{WLS}}$ in Eq. (4). The first one performs the well-known Tikhonov regularization, ensuring realistic control filter coefficients and reducing the sensitivity of the system to external variations and calculation errors [1]. The second additional constraint adjusts the working bandwidth to the frequency range of interest, described by the coefficients of a finite impulse response filter and included in the block matrix $\mathbf{R}_B$ [21]. The last constraint allows reducing possible artifacts arising from pre- and postringing in the control filters by shaping their envelope with a piecewise function included in the diagonal matrix $\mathbf{R}_E$ [23]. These penalties are, respectively, weighted by λ, γ, and ε in the resulting cost function:

$$\begin{aligned} \mathrm{J_{RWLS}}(\mathbf{w}) = {} & (1-\beta)[(\mathbf{H}_b\mathbf{w} - \mathbf{p}_r)^{\mathrm{T}}(\mathbf{H}_b\mathbf{w} - \mathbf{p}_r)] \\ & + \beta(\mathbf{H}_d\mathbf{w})^{\mathrm{T}}\mathbf{H}_d\mathbf{w} \\ & + \lambda\mathbf{w}^{\mathrm{T}}\mathbf{I}_{NI}\mathbf{w} + \gamma\mathbf{w}^{\mathrm{T}}\mathbf{R}_B\mathbf{w} + \epsilon\mathbf{w}^{\mathrm{T}}\mathbf{R}_E\mathbf{w}\ , \end{aligned} \tag{7}$$

with $\mathbf{I}_{NI}$ the identity matrix with size $NI \times NI$. The cost function $\mathrm{J_{RWLS}}$, when minimized with respect to $\mathbf{w}$, leads to the optimal control filters. For this, the derivative of $\mathrm{J_{RWLS}}$ is equated to zero, resulting in

$$\mathbf{w} = \Phi^{-1}(1-\beta)\mathbf{H}_b^{\mathrm{T}}\mathbf{p}_r\ , \tag{8}$$

where $\Phi = [(1-\beta)\mathbf{R}_b + \beta\mathbf{R}_d + \lambda\mathbf{I}_{NI} + \gamma\,\mathbf{R}_B + \varepsilon\mathbf{R}_E]$.

The two main goals defined in the WLS method also suggest natural ways to evaluate the performance of a sound zone rendering system. Reducing the pressure in the dark zone is usually assessed through the ratio of the energies in the zones, here referred to as the acoustic contrast ratio (ACR). Matching $\mathbf{p}_b$ and $\mathbf{p}_r$ in the bright zone is usually evaluated through the error between their energies, also understood as the reproduction error and here calculated as the normalized mean squared error (NMSE) [24]. Given an input signal containing K samples, these are averaged over time and space as

$$\mathrm{ACR_t} = 10\log_{10}\left[\frac{\sum_{k=1}^{K}\sum_{m=1}^{M}|\mathbf{p}^{\mathbf{e}}_{b_m}[k]|^2}{\sum_{k=1}^{K}\sum_{m=1}^{M}|\mathbf{p}^{\mathbf{e}}_{d_m}[k]|^2}\right], \tag{9}$$

$$\mathrm{NMSE_t} = 10\log_{10}\left[\frac{\sum_{k=1}^{K}\sum_{m=1}^{M}|\mathbf{p}^{\mathbf{e}}_{r_m}[k]-\mathbf{p}^{\mathbf{e}}_{b_m}[k]|^2}{\sum_{k=1}^{K}\sum_{m=1}^{M}|\mathbf{p}^{\mathbf{e}}_{r_m}[k]|^2}\right], \tag{10}$$

where $\mathbf{p}^{\mathbf{e}}_{b_m}$ and $\mathbf{p}^{\mathbf{e}}_{d_m}$ are the evaluation pressures at microphone m in the bright and dark zones, respectively. Instead of $\mathrm{R}^{\mathbf{d}}$, these are calculated with the $\mathrm{R}^{\mathbf{e}}$ set of RIRs, by convolving them with the input signal filtered by the control filters. Similarly, $\mathbf{p}^{\mathbf{e}}_{r_m}$ is the reference evaluation pressure at microphone m in the bright zone, corresponding to the pressure generated by the reference source on its own, without any control. In the frequency domain, the respective evaluation pressures are denoted $\bar{\mathbf{p}}^{\mathbf{e}}_b$, $\bar{\mathbf{p}}^{\mathbf{e}}_d$, and $\bar{\mathbf{p}}^{\mathbf{e}}_r$. Hence, both metrics can be evaluated for a frequency f, referred to, respectively, as $\mathrm{ACR_f}$ and $\mathrm{NMSE_f}$, and calculated as

$$\mathrm{ACR_f}[f] = 10\log_{10}\left[\frac{\sum_{m=1}^{M}|\bar{\mathbf{p}}^{\mathbf{e}}_{b_m}[f]|^2}{\sum_{m=1}^{M}|\bar{\mathbf{p}}^{\mathbf{e}}_{d_m}[f]|^2}\right], \tag{11}$$

$$\mathrm{NMSE_f}[f] = 10\log_{10}\left[\frac{\sum_{m=1}^{M}|\bar{\mathbf{p}}^{\mathbf{e}}_{r_m}[f]-\bar{\mathbf{p}}^{\mathbf{e}}_{b_m}[f]|^2}{\frac{1}{F'}\sum_{f'=0}^{F'}\sum_{m=1}^{M}|\bar{\mathbf{p}}^{\mathbf{e}}_{r_m}[f']|^2}\right]. \tag{12}$$

Note that $\bar{\mathbf{p}}^{\mathbf{e}}_{r_m}[f']$ is averaged over F' frequencies, spanning over the same frequencies that are evaluated. That is, $F' = F$ in this work.

Eqs. (11) and (12) allow for calculation of $\mathrm{ACR_{f\,3}}$ and $\mathrm{NMSE_{f\,3}}$, one-third-octave band averages of $\mathrm{ACR_f}$ and $\mathrm{NMSE_f}$. In that case, $\bar{\mathbf{p}}^{\mathbf{e}}_{b_m}$, $\bar{\mathbf{p}}^{\mathbf{e}}_{d_m}$ and $\bar{\mathbf{p}}^{\mathbf{e}}_{r_m}$ are averaged for $f_l \le f \le f_u$. Given a one-third-octave band with center frequency f_c, $f_l = 2^{(-1/6)} \cdot f_c$, and $f_u = 2^{(1/6)} \cdot f_c$.

2 METHODOLOGY

In general, the experiments carried out consisted in evaluating the performance of two low-frequency sound zones: a bright and a dark zone. These were rendered in two different rooms with contrasting acoustic properties by using eight

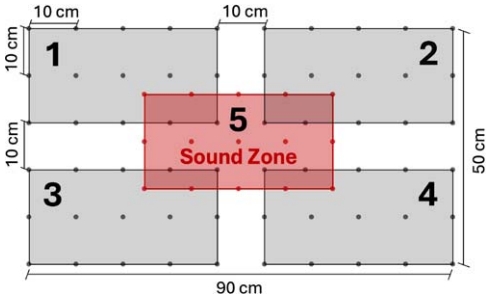


Fig. 2. Measurement positions in both areas of the two rooms included in the data set [19]. The rectangles depict the 15-microphone array in five positions inside both areas that are 30 cm apart from each other. In this study, position 5 corresponds to the sound zones and forms sets $\mathbf{R^d}$ and $\mathbf{R^e_5}$. Positions 1 to 4 sample the areas around the sound zones and form the set $\mathbf{R^e_{1-4}}$.

subwoofers placed on the floor. In both rooms, as shown in Fig. 1 for room 2, six subwoofers were distributed unevenly next to the walls and two next to the sound zones. Both zones are located almost in the middle of each room 1.2 m apart from each other and have a size of 20 × 40 cm. In addition to the zones, a 90 × 50 cm area around each of them was spatially sampled to evaluate the sound field inside and outside of the sound zones. The idea behind such a setup is evaluating low-frequency sound zones under acoustic conditions similar to those in domestic spaces, assuming fixed listener positions. The area of the sound zones would enfold a human head and allow for slight horizontal displacements.

2.1 Low-Frequency Room Impulse Responses

The convention paper on which this work is based [19] also introduced a data set of RIRs for sound field control in the low frequencies with an array of loudspeakers distributed in the room around the sound zones. Because such a data set is also used in this work, it is worth reminding the reader of the following key aspects about the setup and measurements performed for the data set:

- It includes RIRs measured in two rooms with different acoustic conditions. Room 1 has dimensions 8.12 × 7 × 3 m and, featuring several absorptive surfaces, has a reverberation time of $T_{20} = 0.29$ s. Conversely, room 2 has dimensions 7.8 × 4.14 × 3 m and comprises less absorption; it presents a highly reflective environment with a $T_{20} = 0.55$ s. These reverberation times were averaged over one-third-octave bands from 50 Hz to 630 Hz [6].
- The system deployed was composed of 10-in. custom-made subwoofers placed around and next to the sound zones, as shown in Fig. 1. The RIRs were acquired with an array of 15 microphones, distributed in a 3 × 5 grid with 10-cm separation between nodes.
- In each room, two different areas were each sampled in five different positions, as shown in Fig. 2. Position 5 includes measurements at a 1-m and 1.05-m height and corresponds to the sound zones in this study. Positions 1 to 4 include measurements around position 5 taken at a 1.05-m height only. Both areas are 1.2 m apart.
- The RIRs were acquired with the synchronized swept sine (SSS) [25] technique which, in addition to exponentially varying frequency sweeps as measurement signal, uses an analytical expression for the deconvolution filter.
- The SSS measurement signals range from 30 Hz to 600 Hz acquired at $f_s = 48$ kHz and, given the frequency range, downsampled to $f_s = 1.2$ kHz. To include the reverberant tail after the signal, a silence was included in the acquisition time. Henceforth, acquisition times are expressed in bold, indicating the length of the SSS signal and of the following silence, with the respective units. For instance, **200ms + 100ms** denotes a 200 ms SSS signal followed by 100 ms of silence.
- The SSS signals were faded with a raised-cosine window applied at the extremes to ensure smooth displacements of the loudspeakers at the beginning and the end of the playback. Table 1 details the different acquisition times and their respective window lengths.
- Additional details about the rooms, deployments and measurements, can be found in the data set [19].

Table 1. Duration of SSS signals, silences, and fades evaluated.

SSS	15 s	1 s	200	100	50	36
Silence	1 s	0.5 s	100	50	25	18
Fades	20	20	20	20	2	2

Note: Values are in milliseconds unless otherwise specified.

In this work, the different RIRs of the data set were used the same way as in the original convention paper: areas 1 and 2 correspond to bright and dark zones, respectively. Among these, three different sets were used for design and evaluation:

- $\mathbf{R^d}$ is composed by RIRs in position 5, acquired at a 1-m height, and used exclusively for the design of the control filters. Hence, it includes all acquisition times in Table 1, allowing to assess the effect of this factor. These RIRs correspond to $\mathbf{h^d}_{mn}$ in Eq. (1).
- $\mathbf{R^e_5}$ comprises RIRs in position 5 acquired at a 1.05-m height. It is used for evaluation at the sound zones to avoid designing and evaluating with the same RIRs.
- $\mathbf{R^e_{1-4}}$ is composed by RIRs in positions 1 to 4, at a 1.05-m height, and used for evaluation around the sound zones.

As shown in Fig. 2, the $\mathbf{R^e_{1-4}}$ set samples a 90 × 50-cm area around each sound zone and sets the $\mathbf{R^d}$ and $\mathbf{R^e_5}$ sample a 20 × 40-cm area corresponding to the sound zones. Areas in set $\mathbf{R^e_{1-4}}$ are 0.3 m apart, and both zones are 1.2 m apart. Sets $\mathbf{R^e_5}$ and $\mathbf{R^e_{1-4}}$ compose $\mathbf{R^e}$, used to calculate the evaluation pressures in Eqs. (9) to (12). Thus,

to evaluate with all information available, these two sets include only RIRs acquired at **15s + 1s**.

2.2 Spatial Sampling

The effect of the spatial sampling in the performance was evaluated following the methodology of the original article. Thus, the control filters were designed using different numbers of RIRs in the set $\mathbf{R^d}$, representing the use of more or fewer microphones. More specifically, different cases were tested for M sampling points used in a particular case, with $M = 1, \ldots, 15$. For a given M, $\mathbf{c} = [m_1, \ldots, m_M]$ is a $1 \times M$ vector specifying a certain combination of M sampling points. For instance, $M = 1$ offers 15 different combinations, each composed only by one of the 15 microphones in the array. Similarly, $M = 15$ results only in one combination comprising all 15 microphones. Hence, selecting M microphones out of the 15 microphones available results in C_M^{15} combinations $\mathbf{c}$, with

$$C_M^{15} = \binom{15}{M} = \frac{15!}{M!\,(15-M)!}\,. \tag{13}$$

Using the combinatronics notation in [26], this can be expressed as $\binom{15}{M}$, where the set S is composed by all possible combinations $\mathbf{c}$ comprising M microphones. The microphones specified by each $\mathbf{c}$ were selected in both zones to design the control filters for a single evaluation, and the resulting performance was evaluated with Eqs. (9) to (12). Additionally, the average of each metric for all possible combinations was calculated.[1] For the respective individual metrics, these are referred to as $\overline{\mathrm{ACR}}_\mathrm{t}$, $\overline{\mathrm{NMSE}}_\mathrm{t}$, $\overline{\mathrm{ACR}}_\mathrm{f}$, $\overline{\mathrm{NMSE}}_\mathrm{f}$, $\overline{\mathrm{ACR}}_{\mathrm{f}\,\mathbf{3}}$, and $\overline{\mathrm{NMSE}}_{\mathrm{f}\,\mathbf{3}}$. In these analyses, all possible combinations for each number of spatial sampling points M were used, not only a random selection. This approach allowed for drawing wider, general conclusions, independent of specific microphone arrangements.[2]

2.3 Sound Zones Simulation

Using the respective measured RIRs, two sound zones were simulated in each room. Each of the N control filters was $I = 100$ taps long, and the weighting factor $\beta = 0.97$ was chosen to maximize the contrast between both zones. The additional weights were $\lambda = 10^{-3}$, $\gamma = 10^{-5}$, and $\varepsilon = 10^{-7}$. The modeling delay was $\tau_{\mathrm{mod}} = 20$ ms. The input signal was defined as a Dirac impulse to directly evaluate the response of the system [21]. The desired reference pressure $\mathbf{p}_r$ was the one generated by the loudspeaker the closest to the bright zone, and for each room, the same $\mathbf{p}_r$ was used for all the cases and combinations evaluated.

[1]Averages were applied over the linear quantities [i.e., the ratios of pressures inside the logarithms in Eqs. (9) to (12)].

[2]Because the combinations were drawn from a grid of points, the terms "microphone combinations" and "microphone arrangements" are henceforth used interchangeably.

3 RESULTS

This section presents the results obtained with the described methodology in terms of the objective metrics defined in SEC. 1: the ACR and the NMSE. These quantify, respectively, the acoustic separation between zones, and the reproduction accuracy in the bright zone. Analyses in the time domain were originally presented in [19] and are here revised and extended to, among others, analyses in the frequency domain. In all cases, the evaluated frequency range spans from 30 Hz to 600 Hz.

3.1 Sampling Points and Acquisition Time

Results in this section focus on the effects of RIR acquisition time and number of microphones in the control filters design. The amount of sampling points is emulated by the number of RIRs of the set $\mathbf{R^d}$ included in Eq. (8).

3.1.1 Analyses in the Time Domain

Figs. 3 and 4 compare the $\overline{\mathrm{ACR}}_\mathrm{t}$ and $\overline{\mathrm{NMSE}}_\mathrm{t}$, respectively, for both acoustic conditions considered. Evaluated with the $\mathbf{R_5^e}$ RIRs, each value is the average of all possible combinations of the respective number of microphones for each of the six acquisition times evaluated. The individual performances, ACR_t and NMSE_t, were evaluated according to Eqs. (9) and (10), and the range from lowest to highest is presented as vertical intervals for each microphone count. For ease of visualization, all these ranges are also shown with shaded areas. Finally, from Eqs. (9) and (10), higher ACR_t and low NMSE_t values denote better performances. For example, $\mathrm{ACR}_\mathrm{t} = 20$ dB means that the energy in the bright zone is almost 160 times greater than in the dark zone. And an $\mathrm{NMSE}_\mathrm{t} = -3$ dB means that the normalized difference between the energies of $\mathbf{p}_b$ and $\mathbf{p}_r$ equals 0.5, indicating an error of 50% in the bright zone.

In room 1, the longest three acquisition times perform the same in terms of $\overline{\mathrm{ACR}}_\mathrm{t}$ and range of values, reaching up to 22.7 dB $\overline{\mathrm{ACR}}_\mathrm{t}$ values. The **100ms + 50ms** time shows the same trend and for fewer than three microphones shows almost the same values as the longer times. However, the maximum ACR_t is 21.5 dB. With respect to these four longer times, **50ms + 25ms** and **36ms + 18ms** degrade the performance by more than 5 dB and 10 dB, respectively. For all times, fewer sampling points resulted in larger differences between the lowest and the highest ACR_t values, partly due to the number of possible arrangements $\mathbf{c}$ for each M. Additionally, the $\overline{\mathrm{ACR}}_\mathrm{t}$ is closer to the highest ACR_t for almost all cases, suggesting that most of the microphone combinations result in high ACR. Such performances can be initially attributed to the choice of the weighting factor $\beta = 0.97$, which allocates greater resources to increasing the contrast between the zones.

The previous observations are also valid for room 2, only with two exceptions: the maximum contrast is lower, reaching up to 18 dB, and such a value is obtained only with the two longest acquisition times. In other words, contrast is degraded in room 1 with **100ms + 50ms** and shorter times, whereas it is degraded in room 2 with **200ms + 100ms** and shorter times. These two differences

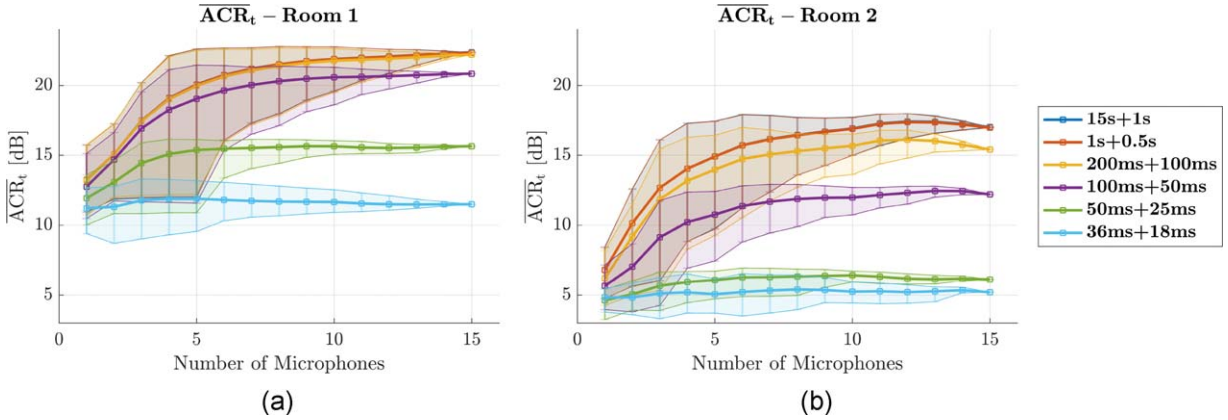


Fig. 3. ACR_t averaged over all combinations of each number of microphones, with RIRs from six acquisition times in room 1 (a) and room 2 (b). The vertical ranges denote the lowest and highest ACR_t among the combinations of a given number of microphones and for ease of visualization have been encircled by the shaded areas. Note that, in both rooms, **15s + 1s** and **1s + 0.5s** overlap, and in room 1, they also overlap with **200ms + 100ms**.

between rooms, mainly due to their distinct acoustic conditions, will be henceforth consistently noted.

The NMSE curves observed in Fig. 4 share important similarities with those in Fig. 3. For instance, room 1 presents lower values than room 2, translated in better performance; and with some exceptions, fewer microphones present worse $\overline{\text{NMSE}}_t$ and a wider range of values than using multiple microphones. However, opposite of the $\overline{\text{ACR}}_t$, the differences between acquisition times and over a number of spatial sampling points are relatively small. This similarity across cases is also greatly influenced by the choice of the weight β, the value of which translates in suboptimal reproduction error in the bright zone. This can also explain the symmetry in the ranges of NMSE_t values with respect to the $\overline{\text{NMSE}}_t$ averages. Finally, same as observed for the $\overline{\text{ACR}}_t$, a very low or high number of microphones result in less $\overline{\text{NMSE}}_t$ variations across microphone arrangements due to the amount of existing possibilities for each M.

Figs. 3 and 4 also reveal interesting relations between sampling points and acquisition times. For instance, in room 1, $\overline{\text{ACR}}_t$ values measured with five, seven, and twelve points at **100ms + 50ms** closely approximate those obtained with four, five, and six points at **200ms + 100ms**. Similarly, in room 2, $\overline{\text{ACR}}_t$ values with five, seven, and twelve points at **200ms + 100ms** approximate those measured with four, five, and six points at **1s + 0.5s**.

3.1.2 Analyses in the Frequency Domain

This section analyzes the behavior of the different metrics in the frequency domain. As a starting point, it was observed that in both rooms, the three longest acquisition times present high similarity over almost all frequencies from 50 Hz to at least 450 Hz.

Relative differences of the averaged metrics were calculated by direct subtraction and, as shown in Fig. 5, these were smaller than ± 1 dB over most of that range for **200ms + 100ms** in room 1 and for **1s + 0.5s** in room 2. These differences, shown here only for 15 microphones, gradually decrease for fewer microphones. Additionally, cases with fewer than five microphones perform similarly above 450 Hz as well. Appearing counterintuitive, this indicates that higher numbers of microphones make the differences between acquisition times more pronounced. Based on this, further analyses will take **200ms + 100ms** as reference ac-

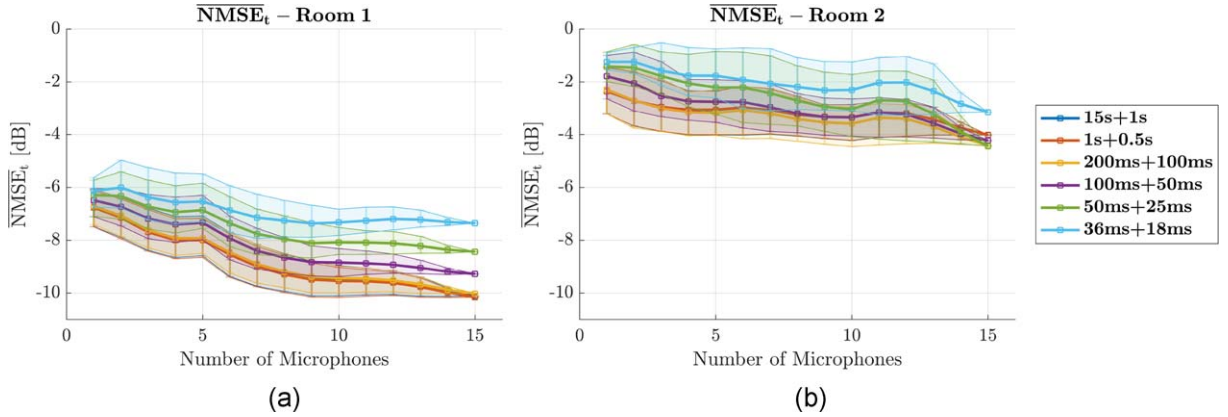


Fig. 4. NMSE_t averaged over all combinations of each number of microphones, with RIRs from six acquisition times in room 1 (a) and room 2 (b). The vertical ranges denote the lowest and highest NMSE_t among the combinations of a given number of microphones and for ease of visualization have been encircled by the shaded areas. Note that, in both rooms, **15s + 1s** and **1s + 0.5s** overlap, and in room 1, they also overlap with **200ms + 100ms**.

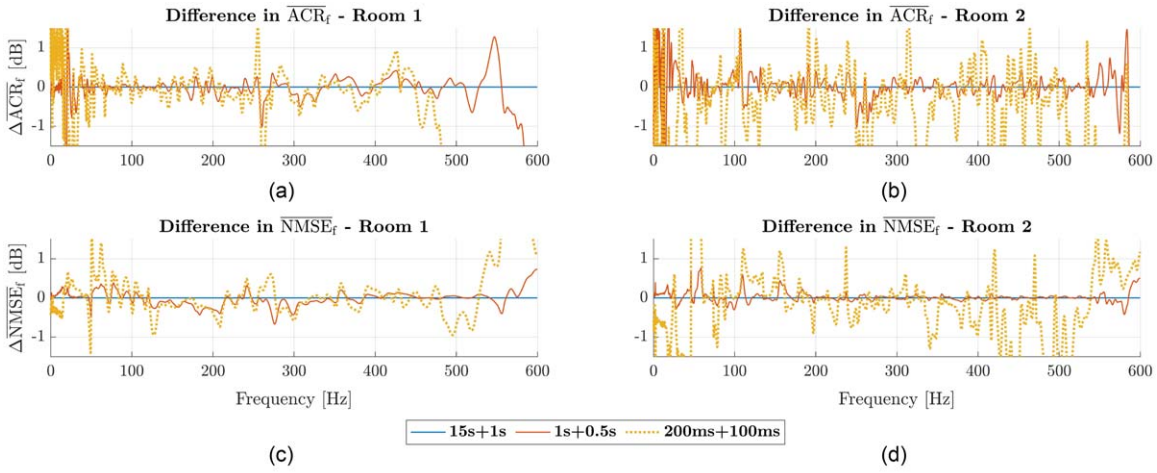


Fig. 5. Relative differences in $\overline{\mathrm{ACR}}_\mathrm{f}$ in room 1 (a) and room 2 (b) and for $\overline{\mathrm{NMSE}}_\mathrm{f}$ in room 1 (c) and room 2 (d) for **1s + 0.5s** and **200ms + 100ms** with respect to **15s + 1s**. Presenting the most pronounced differences, only the case with 15 microphones is shown.

quisition times for room 1 and **1s + 0.5s** for room 2. This choice is reasonable because frequencies above 450 Hz and even below are usually played by an additional loudspeaker system [17, 18]. Additionally, the 450-Hz limit observed for **200ms + 100ms** was found to be defined by the SSS fadeout, acting as a low-pass filter in the frequency domain [6].

Fig. 6 details all the ACR_f and NMSE_f curves in room 1 for one, six, and 12 microphones for **200ms + 100ms**. The curves with the lowest and highest average (i.e., ACR_t and $\mathrm{NMSE}_\mathrm{t)}$ are highlighted. These are compared with the performance of 15 microphones, comprising all information available. Because one microphone provides less information, the substantial performance degradation with respect to 15 microphones was expected. Additionally, the tight range between curves reveals similar performance for all single-microphone evaluations, indicating that such degradation occurs irrespective of the position of the microphone inside the zone.

When six microphones are used, the range of metric values can be greater than 15 dB for certain frequencies, meaning there is high variation between combinations. However, some of the combinations perform as well as 15 microphones, as observed for ACR_f and NMSE_f, respectively, below 430 Hz and 300 Hz. Therefore, using fewer microphones forces a more careful selection of their combinations and potentially introduces a higher dependence on the location of the sound zones in the room.

For the 12 microphones case, it is observed that more points reduce the gap between the best and worse performing arrangement, making all closer to the reference. This improvement in performance with more spatial sampling points, also observed in Figs. 3 and 4, is due to the increased redundancy in the information acquired, resulting in smaller differences between the different arrangements.

It is worth noting, also in Fig. 6, that the maximum ACR_f performance in room 1 lies above 20 dB between (at least) 50 Hz and 400 Hz. Above 400 Hz, the ACR_f decays below 10 dB before reaching 500 Hz due to the aforementioned SSS fadeout. Conversely, the NMSE_f does not present such a roll-off in the higher range, and values around -10 dB are achieved over most frequencies. Because the reference pressure $\mathbf{p}_r^\mathbf{e}$ is defined with RIRs measured with windowed SSS, the error presents no roll-off above 400 Hz, as opposed to the ACR_f.

The effects of an increased number of microphones were also observed in room 2, and for brevity, only the case with six microphones is shown in Fig. 7. Similarly, the ACR_f achieved with the reference of 15 microphones is closely followed up to about 400 Hz, as is the NMSE_f over almost all frequencies. However, the different acoustic conditions greatly degrade both metrics, as detailed previously. Due to this, the roll-off observed for the ACR_f in room 1 is not noticeable in room 2, and the NMSE_f instead of changing around a fixed value, worsens over frequency. This behavior was also observed for acquisition times longer and shorter than **1s + 0.5s**.

These analyses in the time and frequency domain have shown equivalences in the performance metrics for different amounts of information included in the design process through acquisition times and number of microphones. Moreover, highly reflective environments reduce the maximum achievable performance levels and increase the minimum acquisition time and number of microphones required to achieve these peak values. Such equivalences and analyses have been summarized in SEC. 4.1.

3.2 Area of Control

3.2.1 Broadband Performance

This section studies the influence of the system outside of the sound zones through spatial analyses of the sound fields. For this, the RIRs in set $\mathbf{R}^\mathbf{e}_{\mathbf{1-4}}$ were used to evaluate the performance at those points inside and around the sound zones with the metrics described in SEC. 1. To better visualize the results, maps of the spatial distribution of both metrics were plotted applying MATLAB's bilinear interpolation. Because no constraint for spatial control was included in the cost function in Eq. (4), the maps reflect the control the system can inherently exert.

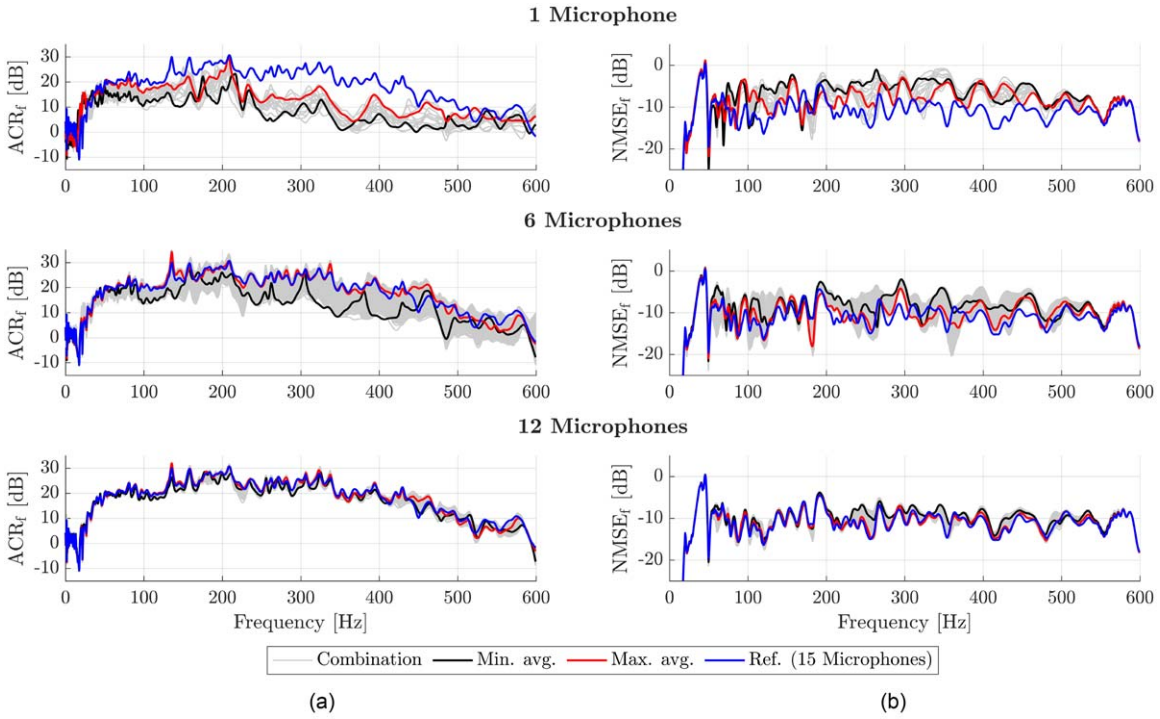


Fig. 6. ACR (a) and NMSE (b) evaluated in the frequency domain for room 1 and **200ms + 100ms** acquisition time. All combinations of one, six, and 12 microphones are shown and compared to the 15 microphones case.

The spatial distributions of the $\overline{\text{ACR}}_\text{t}$ and $\overline{\text{NMSE}}_\text{t}$ in room 1 are shown in Fig. 8. These compare the performance for 15, eight, four, and one spatial sampling points with the four shortest acquisition times (i.e., **200ms + 100ms** to **36ms + 18ms)**. As stated in SEC. 2.2, the $\overline{\text{ACR}}_\text{t}$ and $\overline{\text{NMSE}}_\text{t}$ metrics correspond to the average of all possible microphone combinations. In the figure for each metric, the worse performance at the sound zones corresponds to the case with a single RIR acquired with **36ms + 18ms**, located at the bottom-right. The other maps show the effects of increasing the acquisition time and the number of spatial sampling points: the former is observed from right to left, and the latter from bottom to top. The best-performing case is then 15 RIRs acquired with **200ms + 100ms**, observed at the top-left of the maps of each metric. Interestingly, such a performance appears closely matched by even four microphones and **100ms + 50ms** acquisition time.

Likewise, the influence of the system around the sound zones is also affected by the spatial sampling and the RIR acquisition time. Most of the cases reveal that the controlled areas extend on average up to 10 cm outside of the sound zones. Comparable results were reported in the range of 200 Hz to 500 Hz by Qiao et al. [15]. For sound zones rendering, such an effect is a double-edged sword: whereas it allows freedom of movement within zones, it can increase interference between zones too close together.

Consequently, the $\overline{\text{ACR}}_\text{t}$ and $\overline{\text{NMSE}}_\text{t}$ maps in room 2 showed the highest degree of control inside the sound zones and, to a lesser extent for the $\overline{\text{ACR}}_\text{t}$, outside of the zones as well. Additionally, similar $\overline{\text{ACR}}_\text{t}$ values are achieved for **1s + 0.5s** and **200ms + 100ms** when more than four sampling points are included. Conversely, $\overline{\text{NMSE}}_\text{t}$ maps show poor influence of the system in the surrounding areas, and some cases with more microphones and/or longer acquisition times perform better than other cases with less infor-

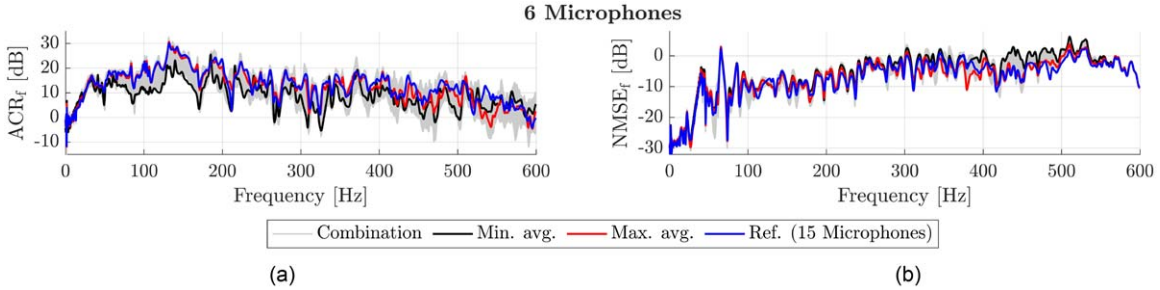


Fig. 7. ACR (a) and NMSE (b) evaluated in the frequency domain for room 2 and **1s + 0.5s** acquisition time. All combinations of six microphones are shown and compared to the 15 microphones case.

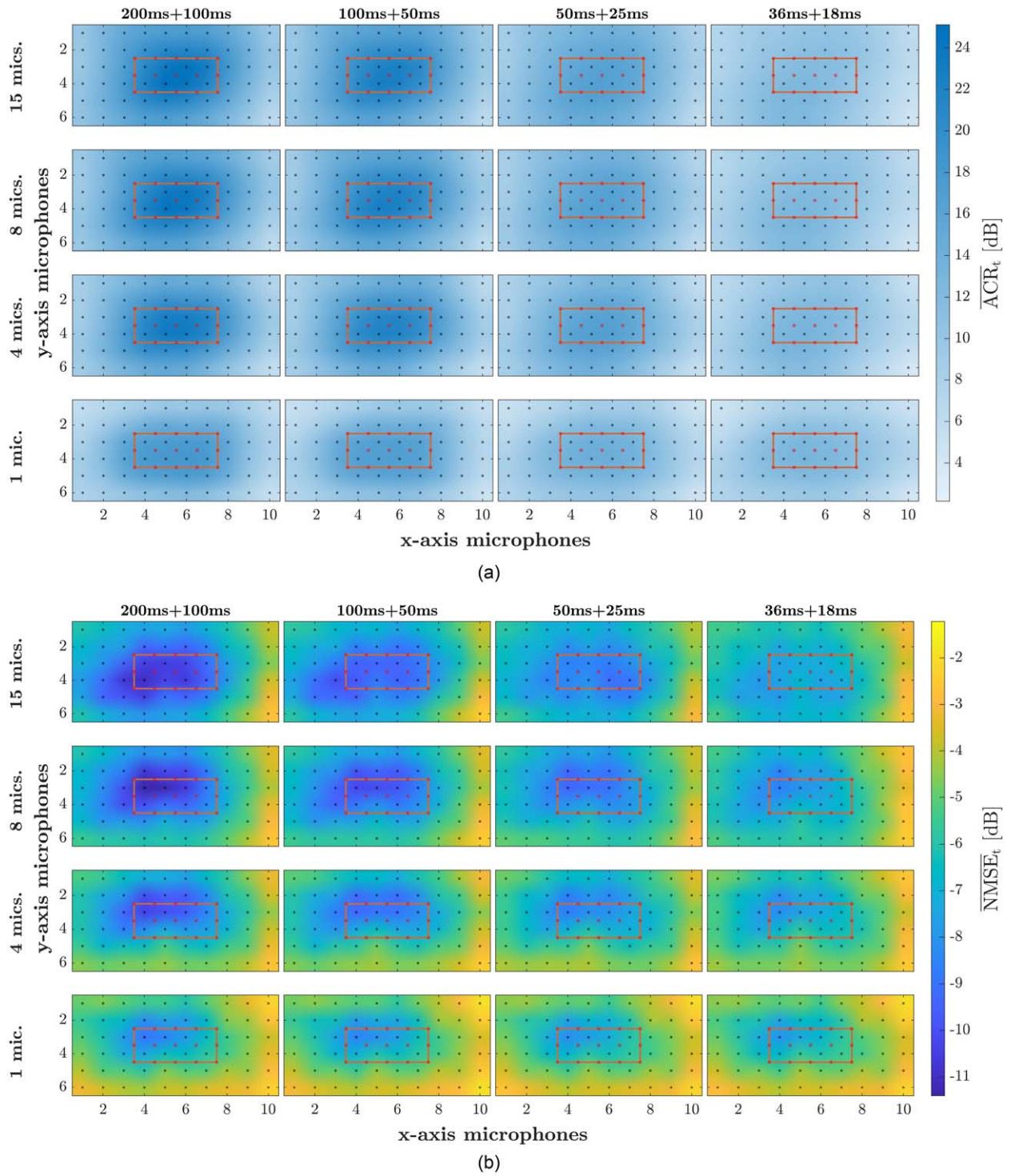


Fig. 8. Spatial distribution in room 1 of the $\overline{\mathrm{ACR}}_t$ (a) and $\overline{\mathrm{NMSE}}_t$ (b) inside (red rectangle) and around the bright zone. The **200ms + 100ms**, **100ms + 50ms**, **50ms + 25ms**, and **36ms + 18ms** acquisition times are sorted over columns, and cases with 15, eight, four, and one microphone are sorted over rows. For each metric, the lowest and highest values of all evaluations and cases define the range of colors, as indicated in the color bars. On each map, *x* and *y* axes correspond to microphone count, where black dots represent the $\mathbf{R}^{\mathbf{e}}_{\mathbf{1-4}}$ grid, and red dots represent the $\mathbf{R}^{\mathbf{e}}_{\mathbf{5}}$ grid.

mation. As observed in SEC. 3.1, this is highly influenced by the acoustic conditions and β.

Finally, note that $\overline{\mathrm{ACR}}_t$ and $\overline{\mathrm{NMSE}}_t$ values in Figs. 3 and 4, evaluated with the RIRs in the $\mathbf{R}^{\mathbf{e}}_{\mathbf{5}}$ set, are related to those in Figs. 8 and 9 at the sound zones (red rectangles). For instance, the difference in the highest achievable metric values between rooms is clearly recognizable in the range of values represented in the color bars.

3.2.2 Performance at Frequency Bands

To gain better insights in the frequency domain, the previous analyses were performed for six one-third-octave bands

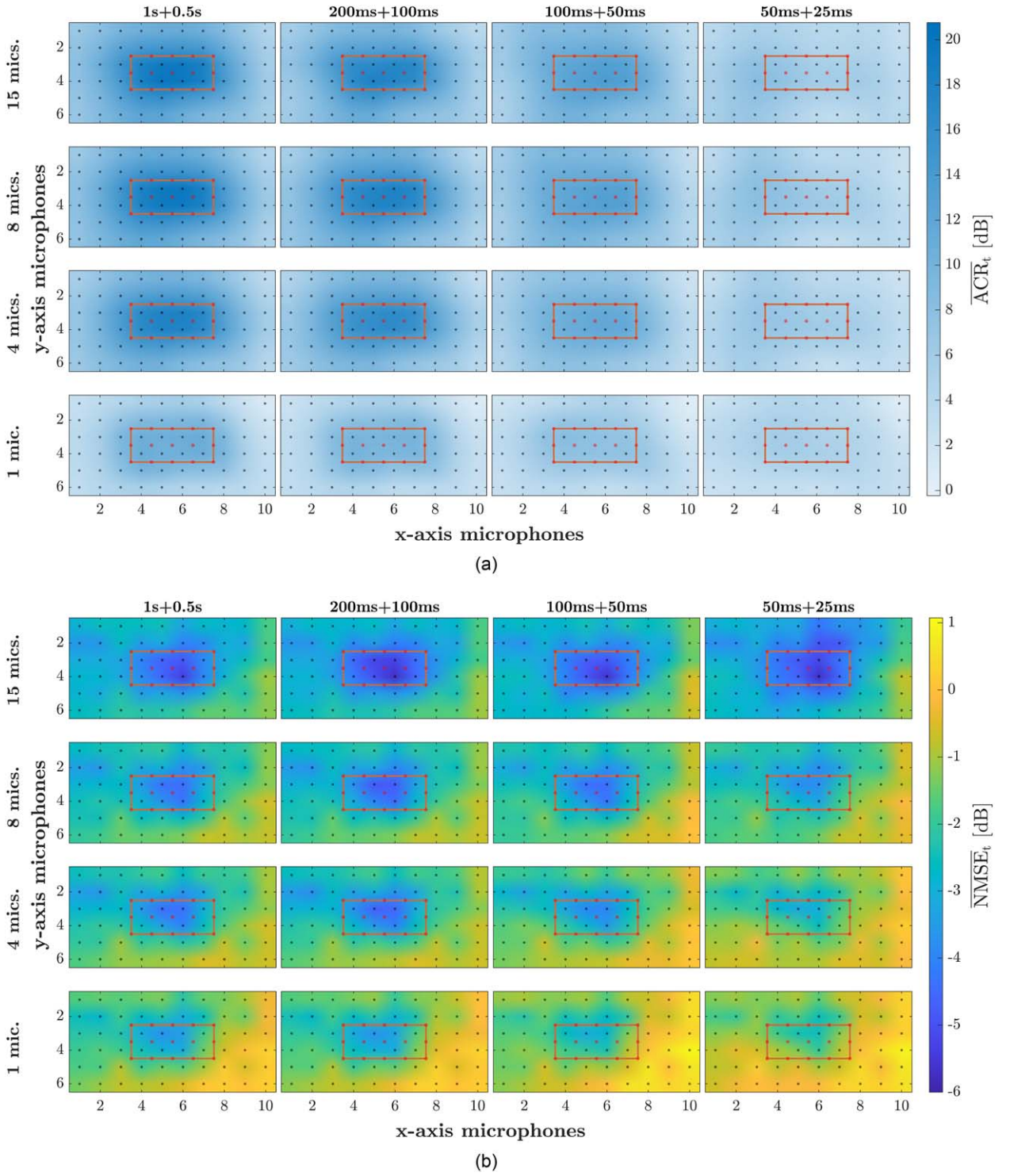


Fig. 9. Spatial distribution in room 2 of the $\overline{\mathrm{ACR}}_t$ (a) and $\overline{\mathrm{NMSE}}_t$ (b) inside (red rectangle) and around the bright zone. The **1s + 0.5s**, **200ms + 100ms**, **100ms + 50ms**, and **50ms + 25ms** acquisition times are sorted over columns, and cases with 15, eight, four, and one microphones are sorted over rows. For each metric, the lowest and highest values of all evaluations and cases define the range of colors, as indicated in the color bars. On each map, *x* and *y* axes correspond to microphone count, where black dots represent the $\mathbf{R^e_{1-4}}$ grid, and red dots represent the $\mathbf{R^e_5}$ grid.

centered at 63, 125, 200, 250, 315 and 450 Hz. Fig. 10 shows a map of both $\overline{\mathrm{ACR}}_{\mathbf{f3}}$ and $\overline{\mathrm{NMSE}}_{\mathbf{f3}}$ for room 1 and room 2, respectively. Accordingly, these analyses were performed for **200s + 100ms** time in room 1 and **1s + 0.5s** in room 2. For brevity, only cases with one and 15 microphones were included. The limits of the respective color bars are the same as in Figs. 8 and 9, and the sound zones are also marked with red rectangles. Because the values observed in the maps inside these red rectangles relate to values in Figs. 6 and 7, specific phenomena in these figures can be analyzed in space with the maps. Keep in mind that averages in the maps include microphones outside of the sound

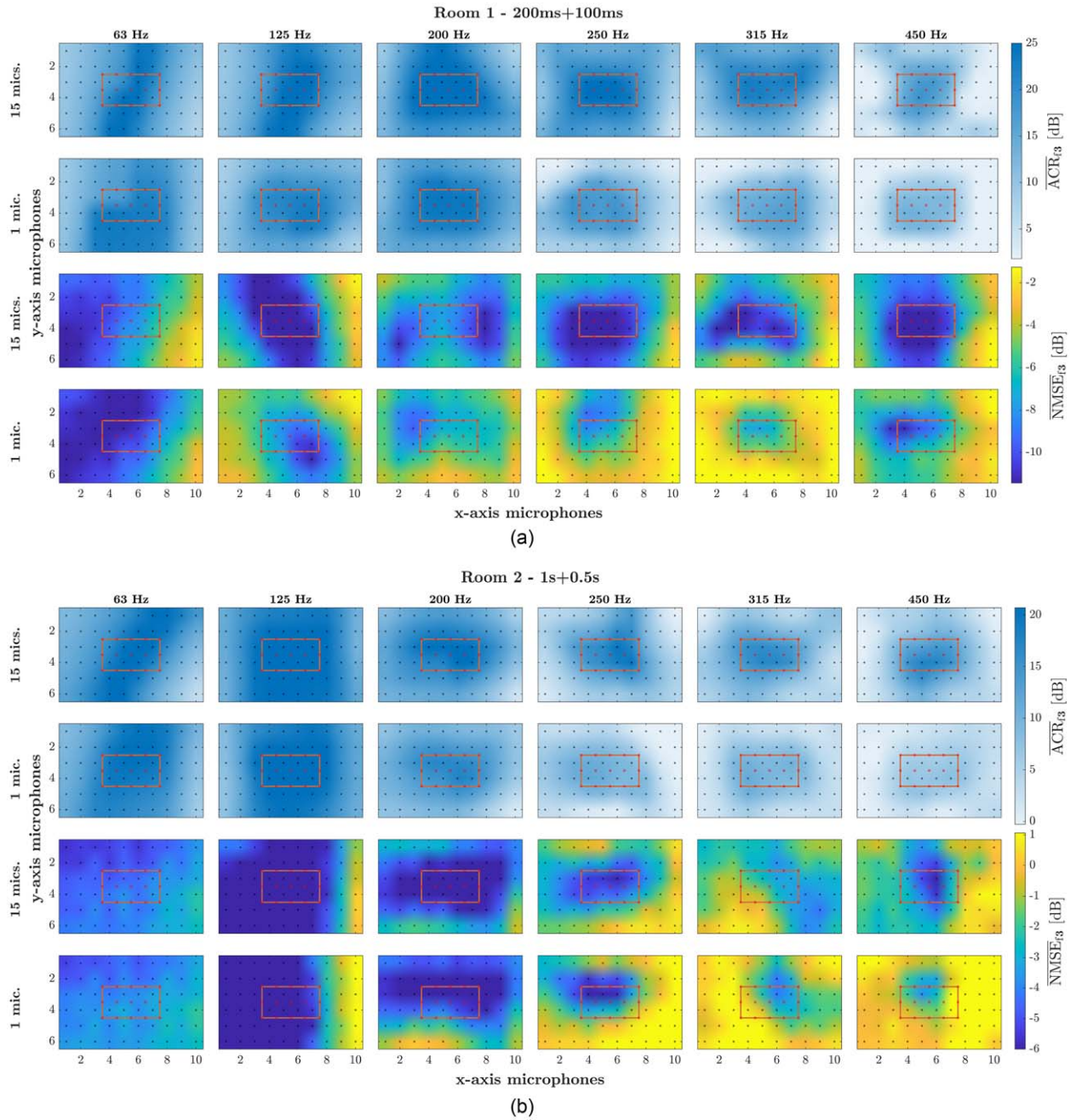


Fig. 10. Spatial distribution of the $\overline{\mathrm{ACR}}_{\mathbf{f3}}$ and $\overline{\mathrm{NMSE}}_{\mathbf{f3}}$ at the bright zone (red rectangle) and around it, evaluated for **200s + 100ms** in room 1 (a) and **1s + 0.5s** in room 2 (b). Values are averaged over one-third-octave frequency bands centered in 63, 125, 200, 250, 315 and 450 Hz. For both rooms, the first row on each metric corresponds to cases with 15 microphones and the second row to one microphone. Conventions and color bar limits are the same as Fig. 8.

zones, therefore presenting lower overall values than those in Figs. 6 and 7.

A quick glimpse of Fig. 10 reveals, as expected, a higher homogeneity in the sound fields for lower frequencies. This is in addition to the degradation of performance already observed for more reverberant environments and/or reduced number of microphones. However, and despite the lower attainable levels, even a single microphone allows extending the region of control outside of the sound zones. This is in agreement with findings in [13, 14]. Hence, using more points to spatially sample the sound zones allows controlling a wider area, especially under less reflective acoustic conditions. Moreover, relative to each room, 15 microphones allow for obtaining high performance values inside and and outside of the sound zones, as observed for 63 Hz and 125 Hz.

It has been already observed that, even though $\beta = 0.97$ favors ACR over NMSE performance, $\mathrm{ACR_f}$ values around 20 dB and $\mathrm{NMSE_f}$ around -10 dB can be achieved. This, however, is not the case for all frequencies. For example, at 200 Hz in room 1 and 63 Hz in room 2, $\mathrm{ACR_f}$ and $\mathrm{NMSE_f}$ values are high, and maps show that lower error is achieved outside rather than inside the bright zone. By observing Figs. 6 and 7, both cases can be understood as

Table 2. Average calculation times of control filters and respective widest confidence intervals.

	Acquisition Times			
Number of Microphones	**15s + 1s**	**200ms + 100ms**	**100ms + 50ms**	**50ms + 25ms**
15	5.26 s	184.20	147.14	114.70
8	2.84 s	117.12	100.31	81.67
4	1.41 s	83.09	68.07	61.28
2	750.21	55.78	58.63	56.28
1	323.06	50.74	59.88	58.03
Maximum CI	± 21	± 17	± 11	± 4

Note: Values are milliseconds unless otherwise specified.

particular occurrences in the NMSE sense. Fig. 6 shows that for six, 12 and 15 microphones (the reference) in room 1, frequencies around 200 Hz is one of the few ranges where the NMSE_f is especially high for all numbers and arrangements of microphones.

Similarly, Fig. 7 shows that in room 2, NMSE_f above 0 dB is achieved at 63 Hz, corresponding to the highest peak observed below 100 Hz. Because the Schroeder frequency [27] of room 1 is 84 Hz, this event is unlikely due to a specific room mode, contrary to what the respective NMSE_{f3} map may suggest. Conversely, the Schroeder frequency in room 2 equals 150 Hz, and thus, this specific occurrence at 63 Hz could be due to modes in room 2.

3.3 Effects on Computations

So far, the effect of reduced information in the acoustical performance has been assessed. This section moves on to the computational complexity, evaluated in terms of the calculation time required to calculate the control filters [7]. For the tests, such calculations were performed with RIRs of room 1 for **200s + 100ms**, **100ms + 50ms**, **50ms + 25ms**, and **15s + 1s**, taken as the reference. The times measured included building the convolutional matrices $\mathbf{H}_b$ and $\mathbf{H}_d$, generating the reference pressure $\mathbf{p}_r$, and computing Eq. (8). For the latter, the `mldivide` command (\ operator) was used for the matrix inversion. Tests were performed in MATLAB R2022a, 64 bit, using a 13-in. MacBookPro (2020) with a 2.3 GHz Quad-Core Intel Core i7 processor. The time was measured with the commands `tic` and `toc` with $f_s = 1.2$ kHz and each filter set to $I = 100$ taps long. The averages of 100 repetitions and widest confidence interval (CI) under verified normality are included in Table 2.

The tests reveal that faster calculations are achieved with fewer microphones and shorter RIRs, associated to shorter acquisition times. For example, with respect to the **15s + 1s** reference, the calculation times for **200s + 100ms** with 15, eight, four, two, and one microphones are reduced, respectively, 28, 24, 17, 13 and six times. And on average, each reduction of the number of spatial sampling points in Table 2 halves the calculation time for **15s + 1s** and reduces it around 1.5 times for the other acquisition times. Finally, calculating Eq. (8) and building $\mathbf{H}_b$ and $\mathbf{H}_d$ are the processes where most of the time is invested. Therefore, the explored strategies can be combined with more efficient implementations to obtain even faster processing times [8, 18].

4 DISCUSSION

4.1 Influence of the System

Results presented in SEC. 3 revealed key insights of the effects of the two strategies evaluated. To summarize:

- Different combinations of acquisition times and numbers of spatial sampling points allow achieving similar performances. Hence, shorter times and fewer sampling points can be used to reach similar results as cases using more information. This also reveals that longer RIR acquisition times can compensate for fewer number of microphones and vice versa.
- Adverse acoustic conditions, as highly reflective spaces, can limit the best achievable metric values in a given room. Even though the two strategies explored were found incapable of counteracting this effect, increasing the control filter length I could result in improved performance under these conditions.
- Using more spatial sampling points reduces the influence of their specific arrangement and the variability in the resulting performances among possible combinations. This also extends the regions controlled outside of the sound zones.
- The lower the frequency, the less information may be required to obtain high performance values and the wider the area of influence.
- Less information results in faster calculations.

The previous findings were highly influenced by three key aspects. First, the response of the loudspeakers by design exhibit a low-frequency roll-off from around 45 Hz. The effects can be clearly observed in Figs. 6 and 7. Because the power emitted gradually decreases with decreasing frequency, the system is incapable of generating high contrast between zones. Conversely, because both the reference and generated pressures $\mathbf{p}_r$ and $\mathbf{p}_b$ include the same roll-off, the NMSE is very low. Similarly, frequencies below 30 Hz laying outside of the intended range to control present poor ACR and low NMSE.

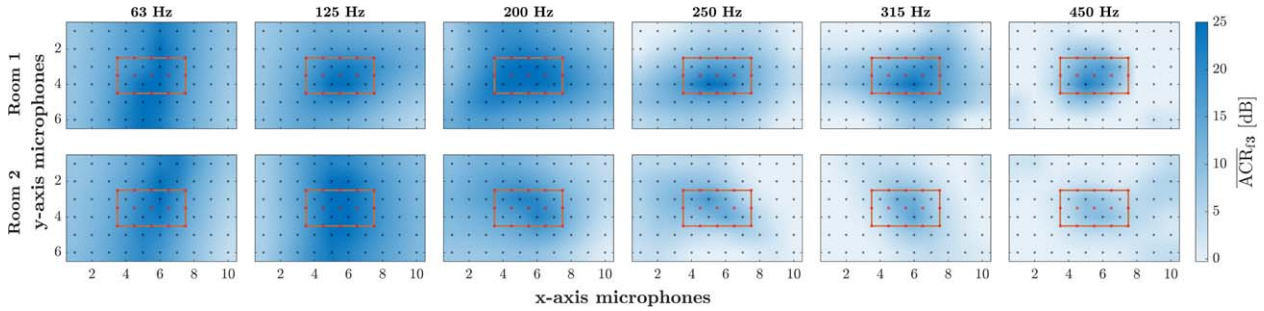


Fig. 11. Comparison of $\overline{\mathrm{ACR}}_{\mathbf{f3}}$ for the center microphone, evaluated at the bright zone (red rectangle) and around it in room 1 (top row) and room 2 (bottom row). The acquisition times are, respectively, **200ms + 100ms** and **1s + 0.5s**. Values are averaged over one-third-octave frequency bands centered in 63, 125, 200, 250, 315, and 450 Hz. Conventions are the same as those in Fig. 8.

The second aspect are the fadeouts applied to the SSS signals for RIR acquisition. These were previously found to influence the response of the system above certain frequencies for the different acquisition times [6]. According to Eq. (35) in [25], the 20-ms fadeout window starts acting at 557 Hz for the **1s + 0.5s** acquisition time, 414 Hz for **200ms + 100ms**, and 287 Hz for **100ms + 50ms**. For **50ms + 25ms**, the 2-ms window applied acts above 518 Hz.

The third aspect is the weight β, which greatly influenced the variability in the NMSE values achieved in both rooms. From the mathematical derivation, it is expected that lower values of β result in improved NMSE and worse ACR performance. However, as observed in this study, the effect of a given β is highly scenario dependant. In other words, a particular value of β can result in different NMSE or ACR changes, and therefore, general conclusions across rooms are difficult to draw. This has also been observed by several authors for different deployments in an anechoic chamber [21], a semianechoic chamber [28], and other typical rooms [11].

Unfortunately, little research has been done on the subjective evaluation of sound zones indirectly linked to the choice of β when rendered with the WLS method. Moreover, to the best of the authors' knowledge, no research linking directly ACR and NMSE with subjective preferences has been made, and minimum desired values for these metrics are still undetermined. Regarding the level of separation, most efforts have been focused on assessing the audio-on-audio interference inside a given zone [29]. Regarding the sound quality, Baykaner et al. assessed it in connection with the planarity of the sound field in the bright zone [30].

Finally, the setup in this study was intended for domestic spaces, where fixed listening zones can be created for activities such as home cinemas, working spots, or even silent areas for studying or sleeping. Nevertheless, the insights derived through it can be translated to other applications, such as automotive, healthcare, gaming, or other entertainments. In all cases, the accurate and efficient reproduction of higher-frequency ranges would require different strategies, as mentioned in the Introduction [17, 18]. Moreover, applications requiring wide zones, such as reinforcement systems for restaurants, exhibitions, or conference rooms, would require deploying large systems, making this approach cumbersome and expensive. In such cases, recent technologies comprising compact arrays of loudspeakers would be a more suitable solution [31].

4.2 A Single-Microphone Case

The methodology described in SEC. 2 comprises averaging over all possible microphone combinations to extract wide, general conclusions, avoiding basing the analyses on very specific cases. This, however, prevented other insights from being observed, such as the actual spatial performance of a single microphone. Due to the averaging over 15 individual positions, maps in Figs. 8 to 10 show worse values than those each microphone can actually achieve.

The same analyses over on-third-octave bands were then performed for a single microphone. Fig. 11 compares the spatial distribution of the $\overline{\mathrm{ACR}}_{\mathbf{f3}}$ evaluated for the microphone at the center of the array, both in room 1 with **200s + 100ms** and in room 2 with **1s + 0.5s**. Both maps show that, even for frequencies as high as 450 Hz, the region of influence extends about 10 cm around the microphone. More specifically, $\mathrm{ACR}_{\mathbf{f3}}$ values above 15 dB and 8 dB can be obtained over such an area in room 1 and room 2, respectively. This is in good agreement with the 10-dB reduction over one-tenth of the wavelength around a single point reported in active control by Elliot et al. [13]. Moreover, due to the dry acoustics in room 1, some bands exhibit greater reductions within this area and/or reductions of more than 10 dB beyond one-tenth of the wavelength. Given the similarity with maps in Fig. 10, NMSE_{f3} maps were not included.

4.3 Intermicrophone Distance

Because the microphone array performs a discrete sampling of the sound field in space, the distance between sampling points defines the maximum frequency for which two samples can be obtained per wavelength. Such a spatial discretization translates in a maximum frequency above which sound field control can no longer be exerted [32].

For instance, Zhao et al. reported strong degradation in the contrast due to spatial aliasing [33]. In their study, the frequency range of interest spanned up to 4 kHz, and the 8-cm microphone separation created aliasing effects above

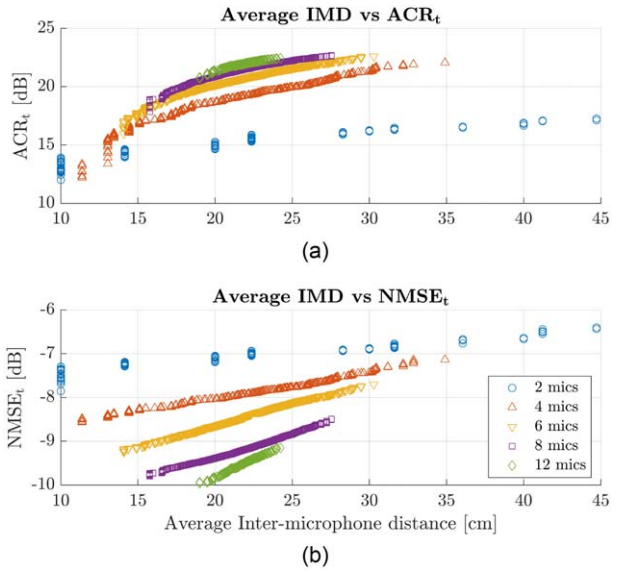


Fig. 12. Average IMD versus ACR_t (a) and $NMSE_t$ (b), calculated over all pairs of microphones in all combinations of two, four, six, eight, and 12 microphones. Evaluated with **100ms + 50ms** RIRs in room 1; the same trends were observed for all acquisition times in both rooms.

2 kHz. Similar effects were then expected in this work. The different microphone combinations resulted in intermicrophone distances (IMDs) between 10 cm and 44.7 cm, with maximum controllable frequencies of 1.7 kHz and 384 Hz, respectively. Hence, additional analyses of specific arrangements featuring diverse IMDs were conducted and, against expectations, no influence of spatial aliasing in the objective performance metrics was found. This discrepancy with observations by Zhao et al. can be attributed to the range of frequencies studied and to the size of the sampled areas with respect to the wavelengths of the corresponding frequency ranges. Thus, spatial aliasing effects are expected to be observed under comparable analyses.

Finally, the IMD was used to investigate the microphone arrangements leading to better performance (i.e., high ACR and low NMSE). The average IMD between all pairs of microphones was calculated for each combination of two, four, six, eight, and 12 microphones and plotted against their respective ACR_t and $NMSE_t$ values. Calculated with **100ms + 50ms** acquisition time of RIRs of room 1, these are depicted in Fig. 12. For all number of microphones, the greater the average IMD, the higher the ACR and, more interestingly, the higher the NMSE as well. This is related to the size of the sampled area, which is proportional to the average IMD. A small sampled area, being easier to control, implies a more successful pressure matching. However, a high contrast will only be attained in that reduced region. Conversely, a large, sparsely sampled area will lead to a more homogenous contrast over a wider region at the expense of controlling the pressure accurately only in the vicinity of the control points. Observed in both rooms and for all acquisition times, this behavior constitutes a key insight for the actual implementation of sound zones.

5 CONCLUSION

Based on and extending previous work presented at the 156th Convention of the Audio Engineering Society [19], this article explored the effects of reduced information in the design of the control filters for low-frequency sound zones. Such a reduction in information was accomplished through two strategies: shorter RIR acquisition times [6] and fewer spatial sampling points. The performance of different combinations of both strategies was assessed in terms of the differences in energy between zones, the reproduction error in the bright zone, and the calculation time of the control filters. The overall area controlled by the system, inside and around the rendered sound zones, was also evaluated.

Different tests and analyses were carried out with a data set of RIRs for low-frequency sound field control measured in two rooms with different acoustic properties over different acquisition times. The findings showed that the same or a similar performance can be obtained when using all or a reduced amount of the information available, the former case resulting in faster control filters calculation. In other words, faster RIR measurements and fewer sampling points still lead to the highest achievable performance in a given room. However, such maximum performances are greatly defined by the acoustic conditions and the system, and the degradations introduced can not be counterbalanced by the explored strategies. Additionally, the higher the number of sampling points, the lower the variability between the performances of their possible combinations; and even when a single microphone is used, the lower the frequency, the wider the area controlled. Moreover, analyses on the IMDs revealed the existence of a trade-off between acoustic contrast and reproduction error arising from the spatial sampling in this frequency range.

Further work could combine the explored strategies with computationally more efficient implementations [8], explore methods to improve the control outside of the sound zones [34], and study minimal spatial sampling configurations for lower redundancy and increased robustness to sound zone locations [35].

6 ACKNOWLEDGMENT

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Grant Agreement No. 956369 and has been supported by KU Leuven internal fund C14/21/075.

7 REFERENCES

[1] T. Betlehem, W. Zhang, M. A. Poletti, and T. D. Abhayapala, "Personal Sound Zones: Delivering Interface-Free Audio to Multiple Listeners," *IEEE Signal Process Mag.*, vol. 32, no. 2, pp. 81–91 (2015 Mar.). https://doi.org/10.1109/msp.2014.2360707.

[2] G.-B. Stan, J.-J. Embrechts, and D. Archambeau, "Comparison of Different Impulse Response Measurement

Techniques," *J. Audio Eng. Soc.*, vol. 50, no. 4, pp. 249–262 (2002 Apr.).

[3] M. B. Møller and M. Olsen, "Sound Zones: On Performance Prediction of Contrast Control Methods," in *Proceedings of the AES International Conference on Sound Field Control* (2016 Jul.), paper 18308.

[4] M. Olsen and M. B. Møller, "Sound Zones: On the Effect of Ambient Temperature Variations in Feed-Forward Systems," presented at the *142th Convention of the Audio Engineering Society* (2017 May), paper 9806.

[5] L. Habib and M. Melon, "Effect of Loudspeaker Non-Linearities on the Performance of Personal Sound Zones," in *Proceedings of the 10th Convention of the European Acoustics Association Forum Acusticum* (Torino, Italy) (2023 Sep.). https://doi.org/10.61782/fa.2023.0193.

[6] J. Cadavid, M. B. Møller, C. S. Pedersen, S. Bech, T. van Waterschoot, and J. Østergaard, "Performance of Low-Frequency Sound Zones With Very Fast Room Impulse Response Measurements," *J. Acoust. Soc. Am.*, vol. 155, no. 1, pp. 757–768 (2024 Jan.). https://doi.org/10.1121/10.0024519.

[7] 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," presented at the *147th Convention of the Audio Engineering Society* (2019 Oct.), paper 10216.

[8] 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.*, vol. 28, pp. 2684–2696 (2020 Sep.). https://doi.org/10.1109/taslp.2020.3023628.

[9] L. Shi, T. Lee, L. Zhang, J. K. Nielsen, and M. G. Christensen, "A Fast Reduced-Rank Sound Zone Control Algorithm Using the Conjugate Gradient Method," in *Proceedings of the ICASSP IEEE International Conference on Acoustics, Speech and Signal Process.*, pp. 436–440 (Barcelona, Spain) (2020 May). https://doi.org/10.1109/ICASSP40776.2020.9054461.

[10] M. Ebri, N. Strozzi, F. M. Fazi, A. Farina, and L. Cattani, "Individual Listening Zone With Frequency-Dependent Trim of Measured Impulse Responses," presented at the *149th Convention of the Audio Engineering Society* (2020 Oct.). paper 10409.

[11] V. Molés-Cases, S. J. Elliott, J. Cheer, G. Piñero, and A. Gonzalez, "Weighted Pressure Matching With Windowed Targets for Personal Sound Zones," *J. Acoust. Soc. Am.*, vol. 151, no. 1, pp. 334–345 (2022 Jan.). https://doi.org/10.1121/10.0009275.

[12] J. Cadavid, M. B. Møller, S. Bech, T. van Waterschoot, and J. Østergaard, "Performance of Low Frequency Sound Zones Based on Truncated Room Impulse Responses," in *Proceedings of the 17th International Audio Mostly Conference*, pp. 239–245 (St. Pölten, Austria) (2022 Sep.). https://doi.org/10.1145/3561212.3561248.

[13] S. J. Elliott, P. Joseph, A. Bullmore, and P. A. Nelson, "Active Cancellation at a Point in a Pure Tone Diffuse Sound Field," *J. Sound Vib.*, vol. 120, no. 1, pp. 183–189 (1988 Jan.). https://doi.org/10.1016/0022-460X(88)90343-4.

[14] H. Chen, P. Samarasinghe, and T. D. Abhayapala, "In-Car Noise Field Analysis and Multi-Zone Noise Cancellation Quality Estimation," in *Proceedings of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pp. 773–778 (Hong Kong, China) (2015 Dec.) https://doi.org/10.1109/APSIPA.2015.7415377.

[15] Y. Qiao and E. Choueiri, "Optimal Spatial Sampling of Plant Transfer Functions for Head-Tracked Personal Sound Zones," presented at the *154th Convention of the Audio Engineering Society* (2023 May), paper 10659.

[16] J. Brunnström, S. Koyama, and M. Moonen, "Variable Span Trade-Off Filter for Sound Zone Control With Kernel Interpolation Weighting," in *Proceedings of IEEE International Conference onAcoustics, Speech Signal Processing (ICASSP)*, pp. 1071–1075 (Singapore) (2022 May). https://doi.org/10.1109/icassp43922.2022.9746550.

[17] W. Druyvesteyn and J. Garas, "Personal Sound," *J. Audio Eng. Soc.*, vol. 45, no. 9, pp. 685–701 (1997 Sep.).

[18] M. B. Møller, J. Martinez, and J. Østergaard, "Reduced Complexity for Sound Zones With Subband Block Adaptive Filters and a Loudspeaker Line Array," *J. Acoust. Soc. Am.*, vol. 155, no. 4, pp. 2314–2326 (2024 Apr.). https://doi.org/10.1121/10.0025464.

[19] J. Cadavid, M. B. Møller, T. van Waterschoot, S. Bech, and J. Østergaard, "Spatial Sampling Versus Acquisition Time of Room Impulse Responses for Low-Frequency Sound Zones," presented at the *156th Convention of the Audio Engineering Society* (2024 Jun.), paper 10706.

[20] J.-H. Chang and F. Jacobsen, "Sound Field Control With a Circular Double-Layer Array of Loudspeakers," *J. Acoust. Soc. Am.*, vol. 131, no. 6, pp. 4518–4525 (2012 Jun.). https://doi.org/10.1121/1.4714349.

[21] M. F. S. Gálvez, S. J. Elliott, and J. Cheer, "Time Domain Optimization of Filters Used in a Loudspeaker Array for Personal Audio," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 23, no. 11, pp. 1869–1878 (2015 Nov.). https://doi.org/10.1109/taslp.2015.2456428.

[22] M. Poletti, "An Investigation of 2-D Multizone Surround Sound Systems," presented at the *125th Convention of the Audio Engineering Society* (2008 Oct.), paper 7551.

[23] M. Møller and M. Olsen, "Sound Zones: On Envelope Shaping of FIR Filters," in *Proceedings of the 24th International Congress on Sound and Vibration (ICSV 24)*, pp. 613–620 (London, UK) (2017 Jul.). https://doi.org/10.25144/24283.

[24] M. B. Møller and J. Østergaard, "A Moving Horizon Framework for Sound Zones," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 28, pp. 256–265 (2020 Nov.). https://doi.org/10.1109/TASLP.2019.2951995.

[25] A. Novak, P. Lotton, and L. Simon, "Synchronized Swept-Sine: Theory, Application, and Implementation," *J. Audio Eng. Soc.*, vol. 63, no. 10, pp. 786–798 (2015 Nov.). https://doi.org/10.17743/jaes.2015.0071.

[26] R. P. Stanley, *Enumerative Combinatorics Volume 1* (Cambridge University Press, Cambridge, UK, 2011), 2nd ed.

[27] H. Kuttruff, *Room Acoustics* (Spon, London, UK, 2000), 4th ed.

[28] S. Zhao and I. Burnett, "Adaptive Personal Sound Zones Systems With Online Plant Modelling," in *Proceedings of the 24th International Congress on Acoustics* (Gyeongju, Korea) (2022 Oct.).

[29] J. Francombe, *Perceptual Evaluation of Audio-on-Audio Interference in a Personal Sound Zone System*, Ph.D. thesis, University of Surrey, Guildford, UK (2014 Aug.)

[30] K. Baykaner, P. Coleman, R. Mason, P. J. Jackson, J. Francombe, M. Olik, and S. Bech, "The Relationship Between Target Quality and Interference in Sound Zone," *J. Audio Eng. Soc.*, vol. 63, no. 1/2, pp. 78–89 (2015 Jan./Feb.). https://doi.org/10.17743/jaes.2015.0007.

[31] E. Start, "Loudspeaker Matrix Arrays: Challenging the Way We Create and Control Sound," in *Proceedings of the AES Acoustics & Sound Reinforcement Conference* (2024 Jan.), paper 23.

[32] J. Benesty, M. M. Sondhi, and Y. Huang, eds., *Springer Handbook of Speech Processing* (Springer, Heidelberg, Germany, 2008), 1st ed.

[33] S. Zhao, Q. Zhu, E. Cheng, and I. S. Burnett, "A Room Impulse Response Database for Multi-zone Sound Field Reproduction (L)," *J. Acoust. Soc. Am.*, vol. 152, no. 4, pp. 2505–2512 (2022 Oct.). https://doi.org/10.1121/10.0014958.

[34] M. Poletti, T. Abhayapala, and P. Samarasinghe, "Interior and Exterior Sound Field Control Using Two Dimensional Higher-Order Variable-Directivity Sources," *J. Acoust. Soc. Am.*, vol. 131, no. 5, pp. 3814–3823 (2012 May). https://doi.org/10.1121/1.3699192.

[35] S. A. Verburg, F. Elvander, T. van Waterschoot, and E. Fernandez-Grande, "Optimal Sensor Placement for the Spatial Reconstruction of Sound Fields," *EURASIP J. Audio Speech Music Process.*, vol. 2024, no. 1, p. 41 (2024 Aug.). https://doi.org/10.1186/s13636-024-00364-4.

THE AUTHORS


José Miguel Cadavid Tobón


Martin Bo Møller


Toon van Waterschoot


Søren Bech


Jan Østergaard

José Miguel Cadavid Tobón graduated as sound engineer (2014) from San Buenaventura University in Medellín, Colombia. He received his M.Sc. (2019) in electroacoustics from Le Mans Université in France. He is currently a Ph.D. fellow at Aalborg University, Denmark, as part of the Marie Skłodowska-Curie Training Network, "SOUNDS". His Ph.D. project focuses on signal processing for audio and acoustics applied to sound zones rendering. His professional and research experience also include, among others, room acoustics and sound systems design.

•

Martin Bo Møller received an M.Sc. degree in engineering acoustics from the Technical University of Denmark in 2011. Since then, he has been employed by Bang & Olufsen A/S (B&O), and in 2020 he completed a part-time Ph.D. degree, sponsored by B&O, in the Department of Electronic Systems, Aalborg University, Denmark. His research interests include sound field control, loudspeaker array processing, and loudspeaker and room interaction.

•

Toon van Waterschoot received M.Sc. (2001) and Ph.D. (2009) degrees in electrical engineering, both from KU Leuven, Belgium, where he is currently professor and Campus Chair of the Group T Leuven and De Nayer Sint-Katelijne-Waver Campuses. He has been a Consolidator Grantee of the European Research Council (ERC) and Head of the STADIUS Center for Dynamical Systems, Signal Processing, and Data Analytics, and has previously also held teaching and research positions at Delft University of Technology in the Netherlands and the University of Lugano in Switzerland. His research interests are in signal processing, machine learning, and numerical optimization, applied to acoustic signal enhancement, acoustic modeling, audio analysis, and audio reproduction.

•

Søren Bech is Director of Research at Bang & Olufsen and professor of audio perception at Aalborg University's Section AI and Sound in the Technical Faculty of Information Technology and Design. He received a Master of Science and a Ph.D. from the Technical University of Denmark. He has authored over 52 peer-reviewed papers, more than 100 conference contributions, and led 19 funded international research projects. A Fellow of the Acoustical Society of America and the Audio Engineering Society (AES), he has served as governor and vice president of the AES and currently acts as associate technical editor for the AES journal. He has also been vice-chair of the International Telecommunication Union working group 10/3. His research interests include psychoacoustics and, in particular, human perception of reproduced sound in small and medium-sized rooms. Other interests include experimental procedures and statistical analysis of data from sensory analysis of audio and video quality.

•

Jan Østergaard received a master of science in electrical engineering from Aalborg University, Denmark, and a Ph.D. from Delft University of Technology, the Netherlands. He worked as a research and development engineer at ETI, A/S in Aalborg and ETI, Incorporated in Virginia, USA. He also served as a postdoctoral researcher at the University of Newcastle in Australia. Currently, he is a full professor in information theory and signal processing, head of the section on artificial intelligence and sound, and head of the Centre for Acoustic Signal Processing Research at Aalborg University. He has received the Danish Independent Research Council's Young Researcher's Award, the Best Ph.D. Thesis Award from the European Association for Signal Processing, and fellowships from the Danish Independent Research Council and the Villum Foundation's Young Investigator Programme. He is an associate editor for signal processing and source coding for the IEEE Transactions on Information Theory.