Aalborg Universitet

AALBORG UNIVERSITY

Spatial Sampling versus Acquisition Time of Room Impulse Responses for Low-Frequency Sound Zones

Cadavid, José; Møller, Martin; van Waterschoot, Toon; Bech, Søren; Østergaard, Jan

Published in:
AES Europe Spring 2024 - 156th Audio Engineering Society Convention 2024

Creative Commons License
CC BY 4.0

Publication date:
2024

Document Version
Publisher's PDF, also known as Version of record

Link to publication from Aalborg University

Citation for published version (APA):
Cadavid, J., Møller, M., van Waterschoot, T., Bech, S., & Østergaard, J. (2024). Spatial Sampling versus Acquisition Time of Room Impulse Responses for Low-Frequency Sound Zones. In *AES Europe Spring 2024 - 156th Audio Engineering Society Convention 2024* Article 10706 Audio Engineering Society. https://aes2.org/publications/elibrary-page/?id=22519

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal -

Take down policy
If you believe that this document breaches copyright please contact us at vbn@aub.aau.dk providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from vbn.aau.dk on: January 05, 2026

Audio Engineering Society

Convention Paper

Presented at the AES 156th Convention
2024 June 15–17, Madrid, Spain

This paper was peer-reviewed as a complete manuscript for presentation at this convention. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Spatial Sampling versus Acquisition Time of Room Impulse Responses for Low-Frequency Sound Zones

José Cadavid[1], Martin Bo Møller[2], Toon van Waterschoot[3], Søren Bech[1,2], and Jan Østergaard[1]

[1]*Dept. of Electronic Systems, Aalborg University, Fredrik Bajers Vej 7B, Aalborg, 9000, Denmark*
[2]*Bang & Olufsen a/s, Peter Bangs Vej 15, Struer, 7600, Denmark*
[3]*Dept. of Electrical Engineering (ESAT), KU Leuven, Kasteelpark Arenberg 10, Leuven, 3001, Belgium*

Correspondence should be addressed to José Cadavid (jmct@es.aau.dk)

ABSTRACT

Sound zone methods aim to create multiple listening areas within a room, allowing independent playback of different audio content. Achieving this requires a loudspeaker array, multiple microphones per sound zone and a set of control filters, the design of which involves the Room Impulse Responses (RIRs) between loudspeakers and microphones. Under certain acoustic conditions, acquiring these RIRs is possible using very short acquisition times. At lower frequencies, longer wavelengths enable spatial sampling of the sound zones with fewer points, reducing the required number of microphone positions and amount of data to process. This study exploits this characteristic to render and evaluate low-frequency sound zones in two real rooms, with few spatial sampling points. The performance is assessed by objectively evaluating multiple spatial sampling configurations and RIR acquisition times. Furthermore, the system's inherent ability to control sound beyond the sampled regions is evaluated by assessing the sound field over wider areas. For sound zones roughly the size of a human head, satisfactory performance is achieved in low reverberation conditions with at least four microphones, even for RIR acquisition times as short as 150 milliseconds. Conversely, the performance degradation in highly reflective environments cannot be compensated using additional microphones nor longer acquisition times. Finally, reducing these two factors enables faster control filter calculations.

1 Introduction

Many scenarios require playing a specific audio content at certain locations of a room, while keeping a low sound pressure level (SPL) in other locations. Such regions are known, respectively, as Bright and Dark zones. To achieve this, exact information about the sound radiation from loudspeakers is required at these locations, which is typically obtained as RIRs [1]. To ensure high signal-to-noise ratio (SNR) and exciting a wide range of frequencies when acquiring RIRs, signals of several seconds are typically used [2]. On the other hand, gathering information over large areas require more sampling points: the denser the spatial sampling, the more information available [3].

One of the main limitations of methods for sound zones is the need of new measurements after any change occurs in the system [1, 4]. Therefore, both information acquisition and processing must be performed in very short instants of time. Regarding the first requirement, Cadavid et al. reported good performance when very fast RIRs acquisition is used, compared to using long

signals [5]. Fast processing, the second requirement, can be approached from the computational perspective by using efficient algorithms, as those implemented in the frequency domain or comprising reduced-rank matrices [6, 7]. Other approaches rely on reducing the amount of data to process, either by excluding information from the RIRs [8, 9, 10], or by using less sampling points. For example, single-microphone sampling enables active noise control over a region with diameter equal to $\frac{1}{10}$-th of the wavelength [11]. This concept was implemented for low frequencies in an automotive application [12]. For sound zones, using a low number of control points has been studied for personal head-tracked systems, or when supported by processing methods as sound field interpolation [4, 13].

This work explores the combination of reducing both RIRs acquisition time and number of microphones, for expedite sound zones rendering. The frequency range of interest spans from 30 Hz to 600 Hz, enabling the use of fewer control points [4]. Tests carried out include two contrasting acoustic conditions, where RIRs measured with different acquisition times and number of microphones were used to create two sound zones. Analyses on the performance of all cases allowed establishing key relations between measurement time and spatial sampling, and their impact on calculation time.

In addition, a dataset of RIRs for low-frequency sound field control and evaluation in two different acoustic conditions is presented.[1] In order to avoid misleading analyses, two different sets of RIRs were acquired: one for sound zones design, and another for performance evaluation [3]. Henceforth, these two sets will be referred to as $\mathbf{h^d}$ and $\mathbf{h^e}$ respectively.

Theory behind sound zones creation and evaluation is detailed in Section 2. Section 3 describes the experimental RIR acquisition and spatial sampling performed for the different analyses developed. Results from these evaluations are presented and discussed in Section 4, while general conclusions and future work are outlined in Section 5.

2 Sound zones rendering and evaluation

Methods for sound zones rendering have two main objectives: creating destructive interference in the dark zone(s), and constructive interference in the bright zone(s). Among these, the so-called weighted least-squares (WLS) method allows trading-off these two objectives according to the value of a weighting factor β [14]. The method requires defining a reference pressure $\mathbf{p}_r$, generated by a reference source, aiming to match the pressure in the bright zone $\mathbf{p}_b$ with $\mathbf{p}_r$ as accurately as possible.

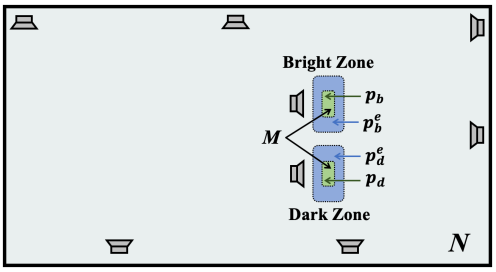


Fig. 1: Sketch of the system deployed for rendering two sound zones in Room 2 (see Section 3.1). Sampling regions are shown as in Figure 2, detailing their respective pressures.

For a system as the one depicted in Figure 1, the WLS method comprises an optimization problem aiming to find N optimal control filters $\mathbf{w}_n$ of length I, which, when applied to the respective N sources, allow rendering the intended sound zones. The concatenation of all $\mathbf{w}_n$ filters produces the $1 \times NI$ vector $\mathbf{w}$. To properly calculate it, each sound zone is spatially sampled by a number of microphones, being M for both zones in this study. These samplings allow acquiring the $J \times 1$ RIR at each microphone m by each loudspeaker n

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

where $\mathbf{h^d}_{mn}$ is one of the $N \cdot M$ RIRs composing $\mathbf{h^d}$, the set for control filters and reference pressure calculation. In the time domain, the most basic form of this optimization problem can be expressed as the cost function

$$\mathrm{J_{WLS}} = (1-\beta)[(\mathbf{H}_b\mathbf{w}-\mathbf{p}_r)^{\mathrm{T}}(\mathbf{H}_b\mathbf{w}-\mathbf{p}_r)] + \beta(\mathbf{H}_d\mathbf{w})^{\mathrm{T}}\mathbf{H}_d\mathbf{w}\,, \quad (2)$$

where $\mathbf{H}_b$ and $\mathbf{H}_d$ are block matrices formed by convolutional matrices containing the design RIRs in Eq. (1) for the bright and dark zones, respectively [15]. The

[1] The measured RIRs and complementary information are openly available with CC BY 4.0 license at: https://doi.org/10.5281/zenodo.11098543.

first term in J_{WLS} corresponds to the objective of matching $\mathbf{p}_b$ and $\mathbf{p}_r$, based on a least-squares approach known as the pressure matching (PM) method [16]. The second term corresponds to the objective of minimizing the pressure in the dark zone $\mathbf{p}_d$. Both terms in Eq. (2) form a convex combination ruled by the weight β.

In order to find the optimal control filters $\mathbf{w}$ by minimizing the cost function J_{WLS} in Eq. (2), its derivative is equated to zero, leading to

$$\mathbf{w} = \left[(1-\beta)\mathbf{H}_b^{\mathrm{T}}\mathbf{H}_b + \beta\mathbf{H}_d^{\mathrm{T}}\mathbf{H}_d\right]^{-1}(1-\beta)\mathbf{H}_b^{\mathrm{T}}\mathbf{p}_r\,. \quad (3)$$

In this work, three additional constraints were added to J_{WLS} for increased robustness through regularisation, to penalize frequencies outside of the range of interest, and to reduce pre- and post-ringing in the control filters. Details on these are not included for brevity, but can be found in references [5, 15, 17].

While $\mathbf{w}$ and $\mathbf{p}_r$ are based on the set of RIRs $\mathbf{h^d}$, the set $\mathbf{h^e}$ was used to calculate the evaluation pressures in the bright and dark zones: $\mathbf{p^e}_b$ and $\mathbf{p^e}_d$. For this, the corresponding RIRs are convolved with the signal of the loudspeakers, i.e., the input signal filtered by the control filters $\mathbf{w}$. In addition, the evaluation reference pressure, $\mathbf{p^e}_r$, is the pressure generated by the reference source in the bright zone when no control is exerted.

The performance can be objectively evaluated in terms of the ratio of the squares of $\mathbf{p^e}_b$ and $\mathbf{p^e}_d$, also known as Acoustic Contrast Ratio (ACR), and the Normalized Mean Squared Error (NMSE) between $\mathbf{p^e}_b$ and $\mathbf{p^e}_r$ [18]. For an input signal with K samples, and taking the M microphones in the bright zone, the ACR and NMSE averaged over time and space are expressed, respectively, as

$$\mathrm{ACR_t} = 10\log_{10}\left[\frac{\sum_{k=1}^{K}\sum_{m=1}^{M}|\mathbf{p^e}_{b_m}[k]|^2}{\sum_{k=1}^{K}\sum_{m=1}^{M}|\mathbf{p^e}_{d_m}[k]|^2}\right], \quad (4)$$

$$\mathrm{NMSE_t} = 10\log_{10}\left[\frac{\sum_{k=1}^{K}\sum_{m=1}^{M}|\mathbf{p^e}_{r_m}[k]-\mathbf{p^e}_{b_m}[k]|^2}{\sum_{k=1}^{K}\sum_{m=1}^{M}|\mathbf{p^e}_{r_m}[k]|^2}\right]. \quad (5)$$

3 Data acquisition

As mentioned before, this work explores the influence of both the acquisition time of the RIRs, and the spatial sampling performed at the regions of interest. This section details the methods followed and the tests carried out to characterise the effect of these two variables in the overall performance of the sound zones.

3.1 RIRs acquisition

The same system deployed in [5] was used in this study for sound zones rendering. It is composed by eight 10-in. custom-made subwoofers, with six of them distributed asymmetrically over the four walls of the room and surrounding both sound zones, and two subwoofers placed next to each sound zone. The sound zones were spatially sampled with an array of 15 microphones arranged in three rows and spaced 10 cm from each other.

The RIRs were acquired using synchronized swept-sines (SSS) [19] comprising the 30 Hz to 600 Hz range. To avoid artefacts during playback, the signals were faded at the extremes with a raised-cosine window. This has different effects on the RIRs [20] which, compared to other methods using swept-sines, are reduced with the SSS method by the analytical calculation of the deconvolution filter. Long to very short acquisition times were used, comprising both the length of the swept-sine, and a subsequent silence kept to capture the reverberant response of the room [2]. Hereafter, the different acquisition times evaluated will be written in bold, expressed as the duration of the SSS and the silence with their respective time units. For example, **100ms + 50ms** denotes a 100 milliseconds SSS signal followed by 50 milliseconds silence. The shorter the acquisition time, the shorter the resulting RIR and, the lower the SNR. Thus, noisy scenarios still require long signals. Two sets of measurements were performed:

- **RIRs for design and reference, $\mathbf{h^d}$.** At each sound zone position and at one metre height, RIRs were acquired specifically to design the control filters according to Eq. (3). These, shown in Figure 2 as position 5, correspond to the $\mathbf{h^d}_{mn}$ RIRs in Eq. (1). Sets of 15 RIRs were acquired using different acquisition times and window lengths, as detailed in Table 1. Since windowing the SSS signals in the time domain implies an additional filtering in the frequency domain [5], and in order to prevent reducing excessively their energy, very short

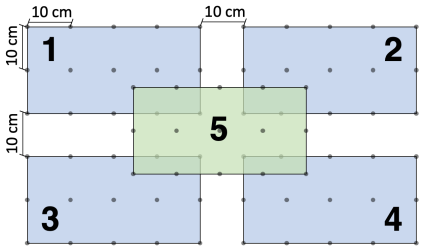


Fig. 2: Positions for RIRs measurement in both rooms. The rectangles depict the 15 microphone array, placed at the location of the sound zones (position 5) and around them (positions 1-4).

fading windows were chosen for the two shortest SSS signals: **50ms+25ms** and **36ms+18ms**. Conversely, the same window length was used for the other SSS signals to prevent introducing additional sources of variation between them.

- **RIRs for evaluation, $\mathbf{h}^{\mathbf{e}}$.** The evaluation set was acquired at a height of 1.05 m, at the five positions in Figure 2, with a **15s+1s** acquisition time. Position 5 corresponds to the exact location of the sound zones, but it is placed 5 cm above. It is surrounded by positions 1 to 4, included to assess the sound field around the sound zones. Henceforth, these subsets will be referred to as $\mathbf{h}^{\mathbf{e}}_{\mathbf{5}}$ and $\mathbf{h}^{\mathbf{e}}_{\mathbf{1-4}}$, respectively. Combined, both subsets sample a 90×50 cm area at and around each sound zone with a total of 75 RIRs.

Table 1: Duration of SSS signals, silences and fades evaluated. Values in milliseconds, unless specified.

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

The measurements were performed in two rooms of the Artificial Intelligence and Sound (AIS) Section laboratories, at Aalborg University. The first room, "Room 1", conforms to the ITU-R BS-1116-1 recommendation [21] and has dimensions $8.12 \times 7 \times 3$ m. The second room, "Room 2", complies with the IEC 268-13 report [22], has dimensions $7.8 \times 4.14 \times 3$ m and enabling variable acoustics, was used with all its absorbent panels. The system deployed in this room is drafted in Figure 1. The reverberation times of Room 1 and Room 2 were, respectively, $T_{20} = 0.29$ s and $T_{20} = 0.55$ s, averaged over the 50 Hz to 630 Hz 1/3 octave bands. Acquired at $f_s = 48$ kHz, the RIRs were downsampled to $f_s = 1.2$ kHz due to the frequency range of interest. In both rooms, the sound zones and their respective $\mathbf{h}^{\mathbf{e}}_{\mathbf{1-4}}$ sets were, respectively, 1.2 m and 30 cm apart.

3.2 Spatial sampling

In order to evaluate the effect of the number of spatial sampling points M, the control filters were designed using different numbers of RIRs in the set $\mathbf{h}^{\mathbf{d}}$. Thus, the performance was evaluated for each number M of RIRs, with $M = 1, \cdots, 15$. For each M, the overall performance is the average of the performance achieved with all possible combinations $\mathbf{c}$ of the M RIRs, with

$$\mathbf{c} = \binom{15}{M} = \frac{15!}{M!\,(15-M)!}\,, \tag{6}$$

where $\mathbf{c} = \left[m_1, \cdots, m_M\right]$ is the $1 \times M$ vector detailing a combination of M microphones used for a single evaluation. For example, for $M = 1$, there are 15 $\mathbf{c}$ vectors detailing each one microphone at the time. Conversely, for $M = 15$, the only possible combination is $\mathbf{c} = \left[1, \cdots, 15\right]$. Note that all existing combinations were used, not a random selection of them. And, for each evaluation, the design of the control filters $\mathbf{w}$ comprised the same combination $\mathbf{c}$ of M sampling points in both zones. This approach lead to broad insights and conclusions, irrespective of microphone combinations.

3.3 Sound zones simulation

Using the acquired data, a bright and a dark sound zone were simulated in both rooms, and the response of the system was evaluated using a Dirac impulse as input signal [15]. The N control filters $\mathbf{w}_n$ had each a length of $I = 100$ taps, were calculated with $\beta = 0.97$, and the reference source was the loudspeaker the closest to the bright zone. Thus, $\mathbf{p}_{r_m}$ and $\mathbf{p}^{\mathbf{e}}{}_{r_m}$ correspond to the convolution of the input signal with the corresponding RIRs in the $\mathbf{h}^{\mathbf{d}}$ and $\mathbf{h}^{\mathbf{e}}$ sets, respectively. In both cases, the RIRs were those from **15s+1s** time, ensuring the same target in the design, and including as much information as possible in the evaluation.

4 Results

The tests performed and results obtained based on the methods described before are presented in this section.

4.1 Sampling points and acquisition time

The first tests performed evaluate the effect of reducing both the acquisition time of the design RIRs $\mathbf{h}^{\mathbf{d}}$, and the number of sampling points, which translates in the number of RIRs included in the control filters calculation. Figures 3 and 4 compare the ACR_t and NMSE_t according to Eqs. (4) and (5), respectively, for both acoustic conditions considered. Evaluated with the $\mathbf{h}^{\mathbf{e}}_{\mathbf{5}}$ RIRs, each value is the average of all possible combinations of the respective number of microphones, for each of the six acquisition times evaluated.

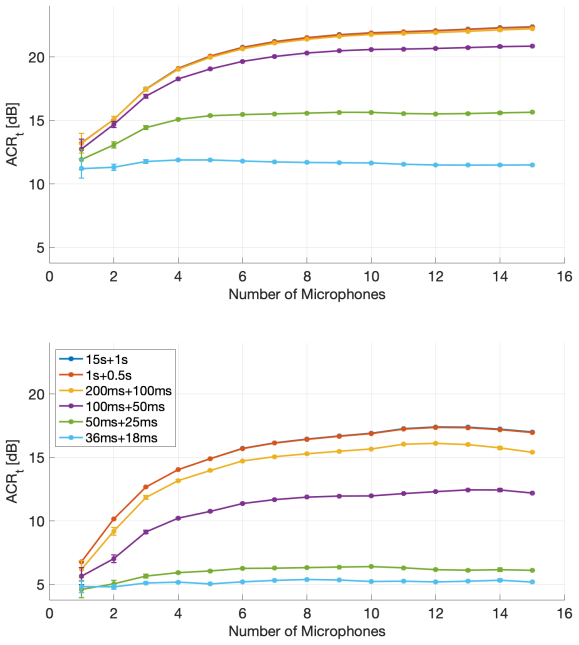


Fig. 3: ACR_t averaged over all combinations of each number of microphones, with RIRs from six acquisition times in Room 1 (top) and Room 2. Confidence intervals are between $\pm$ 0.8 dB. Same ACR_t values are achieved by different acquisition times and number of microphones.

The analysis of the ACR_t in Room 1 shows the same performance for the first three acquisition times over all number of microphones. **100ms+50ms** has a similar trend and values with few microphones, but when more microphones are used, the performance degrades on average up to 1.5 dB. In addition, these four acquisition times achieve the greatest improvements for less than six microphones, values from which the ACR_t increases less than 1.5 dB with respect to the 15-microphones case. These two observations suggest

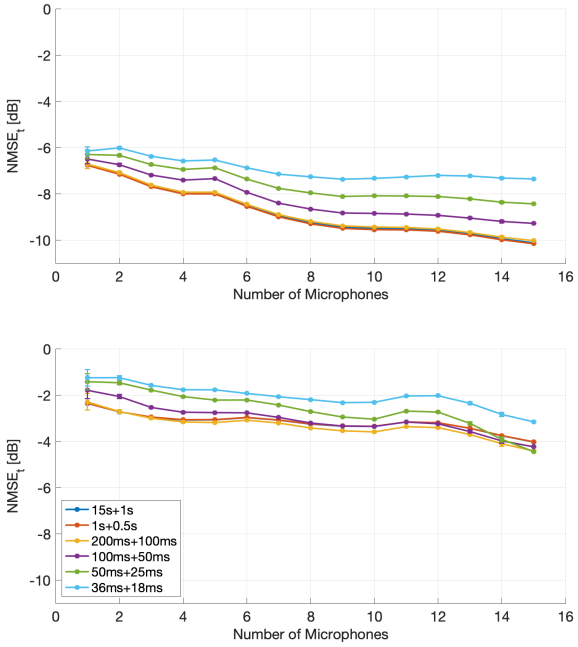


Fig. 4: NMSE_t averaged over all combinations of each number of microphones, with RIRs from six acquisition times in Room 1 (top) and Room 2. Confidence intervals are between $\pm$ 0.4 dB. Different acquisition times and number of microphones achieve equal NMSE_t values.

possible reductions in acquisition time and number of microphones which nonetheless, allow achieving high contrast values. Conversely, **50ms+25ms** and **36ms+18ms** greatly degrade the performance due to the very short acquisition time, the latter reaching almost the same value for all number of microphones. With respect to Room 1, the overall ACR_t values in Room 2 are lower, but in general terms, both behave almost the same: for long acquisition times, the more microphones included, the better the ACR_t. The main difference between rooms, due to the acoustic conditions, lies in the acquisition time from which ACR_t values start decreasing: being **100ms+50ms** for Room 1, it increases to **200ms+100ms** in Room 2.

The difference in the overall values for the ACR_t is also observed for the NMSE_t, where the highly reflective environment in Room 2 greatly degrades the NMSE [5]. In addition, both rooms show the same trend where, as expected, more control points improve the performance. However, differences are very small among all cases, both over acquisition times and over number of mi-

crophones. This is mainly due to the weighting factor chosen, $\beta = 0.97$, which favours ACR maximization.

Specific relations for certain acquisition times and number of microphones can be drawn from Figures 3 and 4. For example, for **100ms+50ms** in Room 1 and **200ms+100ms** in Room 2, the highest $\mathrm{ACR_t}$ achieved with four, five and six microphones, can be equalled by using, respectively, five, seven or twelve microphones. Moreover, the tight confidence intervals detailed in Figures 3 and 4 and obtained under confirmed normality, validate these relations for the majority of the combinations explored. Thus, it has been observed that 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.

4.2 Area of control

In this section, the region of influence of the control system is evaluated outside of the sound zones. Since such a control was not specifically intended in Eq. (2), the observed capability is inherent to the system. Discrete values of the NMSE and ACR were calculated using the $\mathbf{h^e_{1-4}}$ RIRs and, using MatLab's bilinear interpolation, maps of their spatial evolution were plotted. Based on the previous results, tests in both rooms comprise only certain acquisition times and the cases for 1, 4, 8 and 15 microphones. As detailed in Section 4.1, all their possible combinations were evaluated and averaged, in order to keep results in both sections relatable.

Figure 5 shows ACR and NMSE maps obtained with RIRs of Room 1 for **200ms+100ms**, **100ms+50ms**, **50ms+25ms** and **36ms+18ms** acquisition times, and 15, 8, 4 and 1 microphone. For both metrics, cases with the best performance appear in the top-left maps, corresponding to 15 microphones acquired during **200ms+100ms**. Conversely, the performance of a **36ms+18ms** acquisition with a single microphone is shown in the bottom-right corner. Maps between these two show clearly the effect of acquisition time and spatial sampling in the overall performance: it is substantially reduced by very fast RIR acquisition and/or reduced number of microphones. However, as seen in Section 4.1, high performance can be achieved even with **100ms+50ms** and four microphones.

As expected, the lowest NMSE and highest ACR values are achieved at the center of each map, at the sound zones location. However, all cases show that on average, the area of control extends approximately more than 10 cm beyond the bounds of sound zones, mainly due to the long wavelengths of low frequencies. Qiao et al. observed this effect from 200 Hz to 500 Hz in a sound zones context, where it is a double-edged sword by providing freedom of movement inside zones, but increasing interferences when close to each other [4].

These same tests were performed with the $\mathbf{h^e_{1-4}}$ RIRs of Room 2, Figure 6 showing the resulting maps. Based on results from Figures 3 and 4, the acquisition times evaluated for this room were **1s+0.5s**, **200s+100ms**, **100ms+50ms** and **50ms+25ms**. The trends observed in Room 1 regarding using shorter acquisition times and/or fewer microphones, and the reduced performance outside the sound zones positions, are reproduced in Room 2. However, due to the highly reflective environment, both the best performance values and the area where these are obtained are greatly reduced [9].

4.3 Effects on calculation time

Finally, the impact of the strategies explored on the computational burden was evaluated through the time taken to calculate the N control filters $\mathbf{w}_n$ [6]. Table 2 presents time averages of 100 different calculations performed with RIRs of Room 1 for **200s+100ms**, **100ms+50ms**, **50ms+25ms**, and **15s+1s** for reference. The widest confidence interval (CI), under verified normality, is also shown. Calculations were carried out in MatLab (R2022a, 64-bit) running on a MacBook Pro (13", 2020, 2.3 GHz Quad-Core Intel Core i7 processor) with macOS Sonoma 14.3.1. Times reported include the generation of the reference pressure $\mathbf{p}_r$, the block-convolutional matrices $\mathbf{H}_b$ and $\mathbf{H}_d$, and the calculation of Eq. (3). These two last processes were found to take almost the same time. Besides, each control filter $\mathbf{w}_n$ was $I = 100$ taps long, $f_s = 1.2$ kHz, and the matrix inversion in Eq. (3) was performed with the `mldivide` command (\ operator).

As expected, both using shorter RIRs and less control points result in faster calculations. Keeping in mind results in Sections 4.1 and 4.2, it is interesting noting that, on average, halving the number of microphones reduces the calculation time by a factor of 2 for **15s+1s**, and by a factor of up to 1.5 for the other acquisition times. With respect to **15s+1s**, the reduction on calculation time when using 15, 8, 4, 2 and 1 microphone is, respectively, 35, 28, 20, 13 and 6 times. These values were averaged over all the other acquisition times.

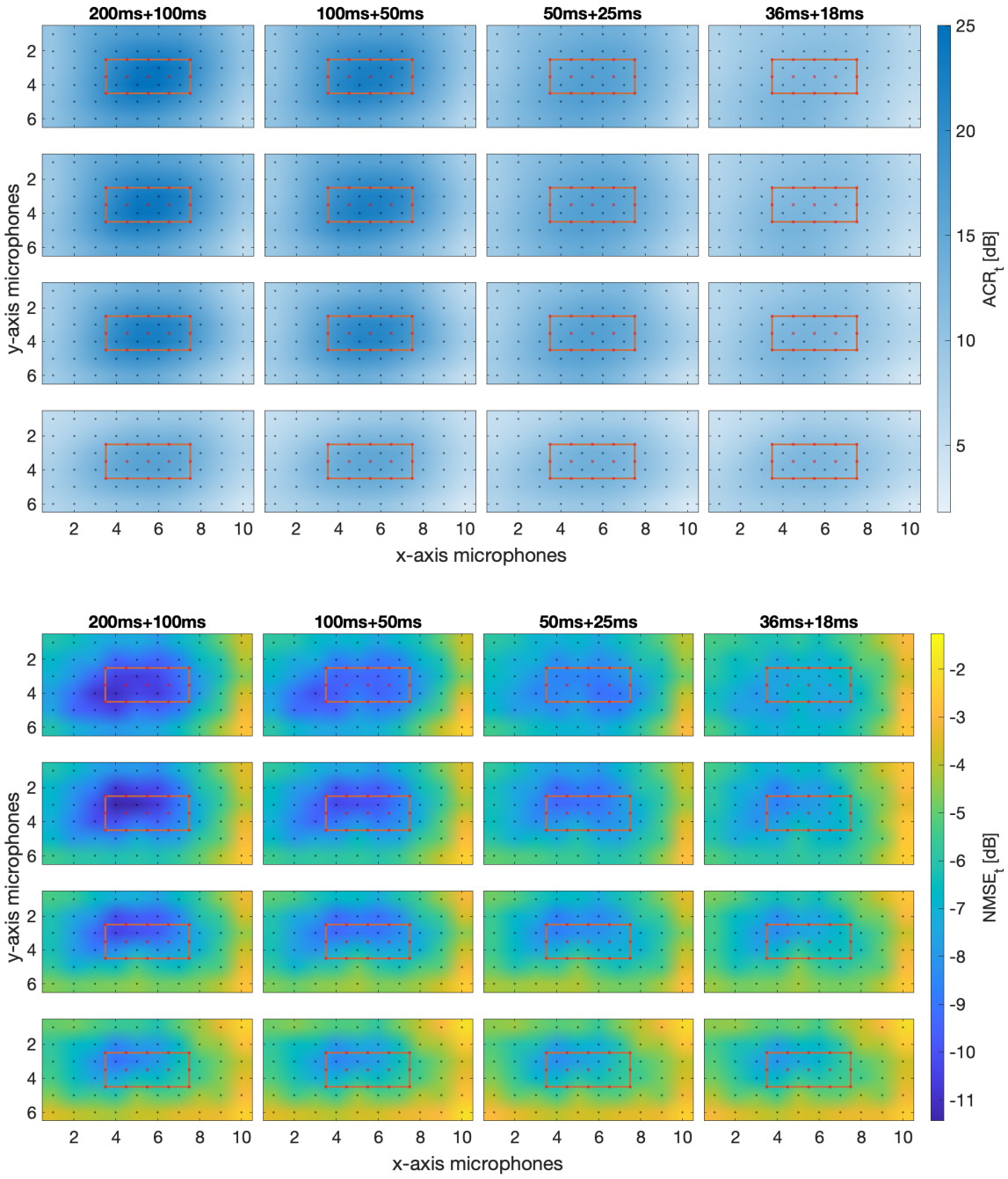


Fig. 5: Spatial distribution of the $\mathrm{ACR_t}$ (top) and $\mathrm{NMSE_t}$ (bottom) in Room 1 at the bright zone (red rectangle) and around it. From left to right, columns correspond to **200ms+100ms**, **100ms+50ms**, **50ms+25ms** and **36ms+18ms** acquisition times. From top to bottom, rows on each plot relates to cases with 15, 8, 4 and 1 microphone. The x and y axes count microphones on the $\mathbf{h^e_{1-4}}$ grid, shown as black dots. Microphones in the $\mathbf{h^e_5}$ grid are shown as red points. Color bars express decibels and their limits correspond to the minimum and maximum values obtained for each metric over all evaluations among all cases. The effect on the metrics of reducing the acquisition time and number of microphones is observable between the best (top-left) and worse (bottom-right) performance.

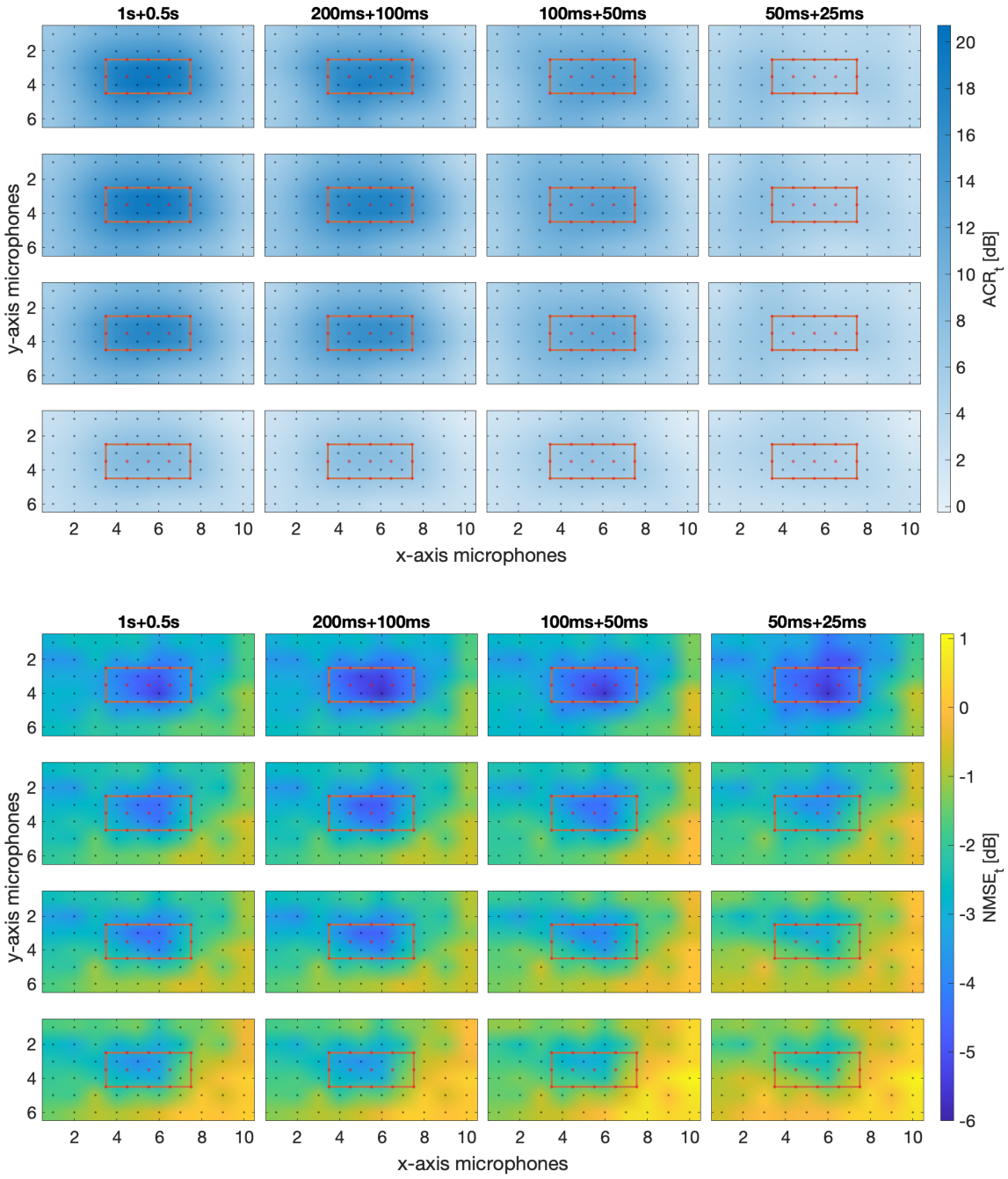


Fig. 6: Spatial distribution of the ACR_t (top) and $NMSE_t$ (bottom) in Room 2 at the bright zone (red rectangle) and around it. From left to right, columns correspond to **1s+0.5s**, **200ms+100ms**, **100ms+50ms** and **50ms+25ms** acquisition times. From top to bottom, rows on each plot relates to cases with 15, 8, 4 and 1 microphone. The *x* and *y* axes count microphones on the $\mathbf{h}^{\mathbf{e}}_{\mathbf{1-4}}$ grid, shown as black dots. Microphones in the $\mathbf{h}^{\mathbf{e}}_{\mathbf{5}}$ grid are shown as red points. Color bars express decibels and their limits correspond to the minimum and maximum values obtained for each metric over all evaluations among all cases. The effect on the metrics of reducing the acquisition time and number of microphones is observable between the best (top-left) and worse (bottom-right) performance.

Table 2: Average calculation times of control filters, and respective widest confidence intervals (CI). Values in milliseconds, unless specified.

	Acquisition Times			
N. Mics.	**15s+1s**	**200+100**	**100+50**	**50+25**
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
Max. CI	$\pm$ 21	$\pm$ 17	$\pm$ 11	$\pm$ 4

5 Conclusion

Two strategies for expedite sound zones update have been implemented: fast acquisition of the RIRs [5], and reduced number of microphones. The combined effect of these two approaches was evaluated for two zones, in two rooms with contrasting acoustics, over the range from 30 Hz to 600 Hz. The performance was expressed in terms of the ratio between energies in both zones, and in terms of the error between a desired pressure and the pressure rendered in the bright zone. The effect of both approaches on the calculation time of the control filters was also investigated. Measurements carried out included microphone positions inside and around the sound zones to also visualize the performance outside the regions of control.

The experiments performed clearly showed the effects, limits and relations of the two strategies explored when high performance is intended. For some cases and with respect to using most information available, each strategy individually allows obtaining maximum performance achievable, with reduced processing time. Under both acoustic conditions and provided low noise levels, little improvement was observed with RIR acquisition times longer than a quarter of a second, and with more than eight microphones. Moreover, four microphones showed to be enough to render satisfactory performance for short reverberation times. Furthermore, the effect of very short acquisition times can be compensated with increased number of microphones, and vice versa. Maps of the performance metrics over space showed good extension of the control outside of the sound zones, specially when more than four microphones are used. These effects, however, can be better exploited in rooms exhibiting low reverberation levels and fewer reflections.

Further work should include analyses of the inter-microphone distance and its impact in the frequency domain, as well as exploring optimal three-dimensional spatial sampling for sound zones. Finally, all the RIRs measured for this work have been made publicly available in the SOUNDS-ETN Zenodo repository. The data presented is well suited for simulations on room acoustics, multichannel audio, and among others, sound field analysis and control.

6 Acknowledgements

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 956369.

References

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

[2] Stan, G.-B., Embrechts, J.-J., and Archambeau, D., "Comparison of different impulse response measurement techniques," *Journal of the Audio engineering society*, 50(4), pp. 249–262, 2002.

[3] Møller, M. B. and Olsen, M., "Sound zones: On performance prediction of contrast control methods," in *Audio Engineering Society Conference: 2016 AES International Conference on Sound Field Control*, Audio Engineering Society, 2016.

[4] Qiao, Y. and Choueiri, E., "Optimal Spatial Sampling of Plant Transfer Functions for Head-Tracked Personal Sound Zones," in *Audio Engineering Society Convention 154*, AES, 2023.

[5] Cadavid, J., Møller, M. B., Pedersen, C. S., Bech, S., van Waterschoot, T., and Østergaard, J., "Performance of low-frequency sound zones with very fast room impulse response measurements," *The Journal of the Acoustical Society of America*, 155(1), pp. 757–768, 2024.

[6] Vindrola, L., Melon, M., Chamard, J.-C., Gazengel, B., and Plantier, G., "Personal sound zones: A comparison between frequency and time domain formulations in a transportation context," in *Audio Engineering Society Convention 147*, Audio Engineering Society, 2019.

[7] Shi, L., Lee, T., Zhang, L., Nielsen, J. K., and Christensen, M. G., "A fast reduced-rank sound zone control algorithm using the conjugate gradient method," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 436–440, IEEE, 2020.

[8] Ebri, M., Strozzi, N., Fazi, F. M., Farina, A., and Cattani, L., "Individual listening zone with frequency-dependent trim of measured impulse responses," in *Audio Engineering Society Convention 149*, Audio Engineering Society, 2020.

[9] Molés-Cases, V., Elliott, S. J., Cheer, J., Piñero, G., and Gonzalez, A., "Weighted pressure matching with windowed targets for personal sound zones," *The Journal of the Acoustical Society of America*, 151(1), pp. 334–345, 2022.

[10] Cadavid, J., Møller, M. B., Bech, S., van Waterschoot, T., and Østergaard, J., "Performance of low frequency sound zones based on truncated room impulse responses," in *Proceedings of the 17th International Audio Mostly Conference*, pp. 239–245, 2022.

[11] Elliott, S. J., Joseph, P., Bullmore, A., and Nelson, P. A., "Active cancellation at a point in a pure tone diffuse sound field," *Journal of sound and vibration*, 120(1), pp. 183–189, 1988.

[12] Chen, H., Samarasinghe, P., and Abhayapala, T. D., "In-car noise field analysis and multi-zone noise cancellation quality estimation," in *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, pp. 773–778, IEEE, 2015.

[13] Brunnström, J., Koyama, S., and Moonen, M., "Variable span trade-off filter for sound zone control with kernel interpolation weighting," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1071–1075, IEEE, 2022.

[14] Chang, J.-H. and Jacobsen, F., "Sound field control with a circular double-layer array of loudspeakers," *The Journal of the Acoustical Society of America*, 131(6), pp. 4518–4525, 2012.

[15] Gálvez, M. F. S., Elliott, S. J., and Cheer, J., "Time domain optimization of filters used in a loudspeaker array for personal audio," *IEEE/ACM Trans. Audio Speech Lang. Process.*, 23(11), pp. 1869–1878, 2015.

[16] Kirkeby, O., Nelson, P. A., Orduna-Bustamante, F., and Hamada, H., "Local sound field reproduction using digital signal processing," *The Journal of the Acoustical Society of America*, 100(3), pp. 1584–1593, 1996.

[17] Møller, M. and Olsen, M., "Sound Zones: On envelope shaping of FIR filters," in *Proc. 24th Int. Congr. on Sound and Vibration 2017 (ICSV 24)*, pp. 613–620, 2017.

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

[19] Novak, A., Lotton, P., and Simon, L., "Synchronized swept-sine: Theory, application, and implementation," *Journal of the Audio Engineering Society*, 63(10), pp. 786–798, 2015.

[20] Guidorzi, P., Barbaresi, L., D'Orazio, D., and Garai, M., "Impulse responses measured with MLS or Swept-Sine signals applied to architectural acoustics: an in-depth analysis of the two methods and some case studies of measurements inside theaters," *Energy Procedia*, 78, pp. 1611–1616, 2015.

[21] ITU-R-1116, "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems," Recommendation, International Telecommunication Union Radiocommunication Assembly, 1997.

[22] IEC TR 60268-13:1998, "Sound system equipment - Part 13: Listening tests on loudspeakers," Technical report, International Electrotechnical Commission, 1998.