







JANUARY 29 2024

Performance of low-frequency sound zones with very fast room impulse response measurements

José Cadavid ; Martin Bo Møller ; Christian Sejer Pedersen ; Søren Bech ; Toon van Waterschoot ; Jan Østergaard 



J. Acoust. Soc. Am. 155, 757–768 (2024)

<https://doi.org/10.1121/10.0024519>



View Online



Export Citation



WE BRING THE NOISE,
YOU BRING THE PRODUCTS

COMMITTED TO A SMARTER,
MORE CONNECTED FUTURE

 **ETS-LINDGREN**
An ESCO Technologies Company

Performance of low-frequency sound zones with very fast room impulse response measurements

José Cadavid,^{1,a)}  Martin Bo Møller,²  Christian Sejer Pedersen,¹  Søren Bech,^{1,2} 
 Toon van Waterschoot,³  and Jan Østergaard¹ 

¹Department of Electronic Systems, Aalborg University, Aalborg, 9000, Denmark

²Bang & Olufsen A/S, Struer, 7600, Denmark

³Department of Electrical Engineering (ESAT), KU Leuven, Leuven, 3001, Belgium

ABSTRACT:

Sound zone methods aim to control the sound field produced by an array of loudspeakers to render a given audio content in specific areas while making it almost inaudible in others. At low frequencies, control filters are based on information of the electro-acoustical path between loudspeakers and listening areas, contained in the room impulse responses (RIRs). This information can be acquired wirelessly through ubiquitous networks of microphones. In that case and for real-time applications in general, short acquisition and processing times are critical. In addition, limiting the amount of data that should be retrieved and processed can also reduce computational demands. Furthermore, such a framework would enable fast adaptation of control filters in changing acoustic environments. This work explores reducing the amount of time and information required to compute control filters when rendering and updating low-frequency sound zones. Using real RIR measurements, it is demonstrated that in some standard acoustic rooms, acquisition times on the order of a few hundred milliseconds are sufficient for accurately rendering sound zones. Moreover, an additional amount of information can be removed from the acquired RIRs without degrading the performance. © 2024 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

<https://doi.org/10.1121/10.0024519>

(Received 11 July 2023; revised 8 December 2023; accepted 7 January 2024; published online 29 January 2024)

[Editor: Yue Ivan Wu]

Pages: 757–768

I. INTRODUCTION

Given an array of loudspeakers and one or more arrays of microphones in a room, it is possible to control the sound at the positions of the microphones. That way, different sound fields known as sound zones can be created at those specific locations, achieving also low interference of the audio contents between them. This process relies on finding a set of optimal control filters, procedures for which several methods have been proposed in the last few decades.

Among these, one of the most used approaches is the weighted least-squares (WLS) method, based on two previous methods for sound zone creation: the so-called pressure matching (PM)¹ and the acoustic contrast control (ACC).² Initially proposed by Chang and Jacobsen³ in the frequency domain, the WLS was formulated in the time domain by Galves *et al.*⁴ It allows combining of the objectives of both the ACC and PM methods, respectively, reducing the pressure in the dark zone and matching the pressure in the bright zone with a reference pressure.

In many applications, personal sound zones are created in a room or car cabin by using beamforming techniques with a linear array of sources.^{5–7} Although practical, the effective reproducible frequency range is constrained in the

upper limit by the distance between adjacent sources and in the lower one by the maximum length of the array.⁸ This approach becomes infeasible for applications where low-frequency content reproduction is desired and sources must be distributed over the room. As an example, Druyvensteyn and Garas⁵ proposed delivering individualized sound using active noise control in the low frequencies, beamforming in the mid frequencies, and highly directive sources in the high frequencies. Thus, in general, the creation of low-frequency sound zones requires different approaches to, for example, handle very large wavelengths and low-frequency resonance modes.⁹

In such applications, the response of the room to the excitation of each source at each of the observation positions must be known.⁹ Therefore, the transfer functions (TFs) of all loudspeaker-microphone pairs, represented in the time domain as room impulse responses (RIRs), must be acquired. Irrespective of the method used among the many existing, this is usually a lengthy process, because long signals are used and/or several repetitions are performed to guarantee both high signal-to-noise ratio (SNR) and wide range excitation.^{10,11} However, expedited information acquisition and processing are required for applications where sound zone updating is intended: that is, when changes in the system and/or acoustical conditions are to be tracked and compensated.

^{a)}Email: jmct@es.aau.dk

In this context, it is worth mentioning different approaches involving using less information, as described by Betlehem *et al.*,⁸ aiming to improve the performance and to achieve faster implementations. For example, Molés-Cases *et al.* used a windowed version of the reference RIRs for the WLS method,¹² while Ebri *et al.* proposed frequency-dependent RIRs windowing for in-car applications.¹³ In both cases, dismissing the reverberant tail of the RIRs improved the performance of the system, especially in terms of the difference in energy between zones for mid and high frequencies. In addition, Cadavid *et al.*¹⁴ reported that low-frequency sound zones can achieve almost the same performance when using complete or truncated RIRs, i.e., removing part of the information each RIR contains. These results served as a basis to explore new ways to reduce the time and information needed to render low-frequency sound zones without compromising performance.

The present work validates and extends these results, previously obtained from simulations, with new experiments based on different measurements. Following Druyvensteyn and Garas' approach⁵ to reproduce each frequency range individually, only frequencies from 30 Hz to 600 Hz are considered, assuming that the remaining frequencies are reproduced using other methods. Two strategies are therefore evaluated: very fast RIR acquisition, and RIR truncation. The first one reduces the length of the measurement signal and the consequent silence to decrease the acquisition time. Keeping in mind the negative effects this has on the information acquired, including lower SNR and limited bandwidth, the main question is to what extent can these inaccuracies be tolerated before performance is degraded.

The second strategy dismisses information in the RIRs obtained, to reduce the computational complexity of determining the control filters. These two strategies are validated using RIR measurements performed under three different acoustic conditions.

In the following sections, it will be demonstrated how, under these conditions, data and acquisition time can be reduced while preserving high performance, i.e., performance close to the maximum achievable by long acquisition times and complete RIRs under the same specific settings. Section II details the methods implemented to generate, improve, and evaluate sound zones. Section III describes the two explored strategies to reduce sound zone rendering and update time, as well as the measurements performed for validation. Results from such tests are included and analyzed in Sec. IV, and Sec. V presents the main conclusions of the article.

II. DESIGN AND EVALUATION OF SOUND ZONES

A. WLS method

Figure 1 depicts an array of N loudspeakers surrounding two arrays of M microphones each. These are the control points of the so-called bright and dark zones, denoted, respectively, by the subscripts b and d . In the time domain, such zones can be created by filtering the audio signal $x[k]$

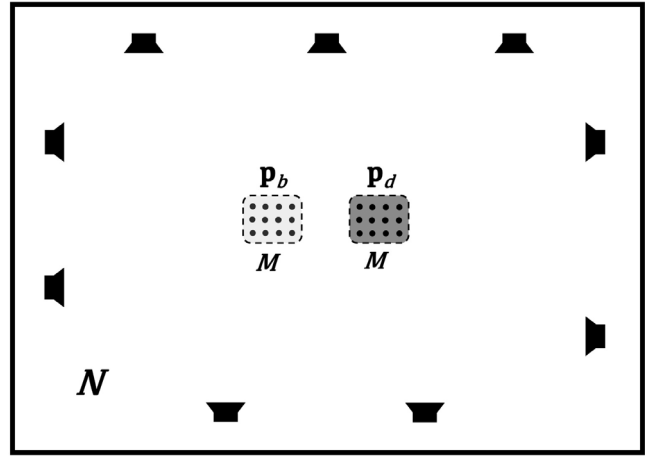


FIG. 1. Example of a setup for rendering two sound zones at the positions of the M -microphone arrays using the N -loudspeaker array.

by the carefully designed $NI \times 1$ multi-channel target control filter \mathbf{w} , composed by the N concatenated filters of length I ,

$$\mathbf{w} = [\mathbf{w}_1^T \ \mathbf{w}_2^T \ \cdots \ \mathbf{w}_N^T]^T, \quad (1)$$

with

$$\mathbf{w}_n = [w_n(0) \ w_n(1) \ \cdots \ w_n(I-1)]^T. \quad (2)$$

Taking the $J \times 1$ RIR vector of the microphone-loudspeaker pair (m, n) ,

$$\mathbf{h}_{mn} = [h_{mn}(0) \ h_{mn}(1) \ \cdots \ h_{mn}(J-1)]^T, \quad (3)$$

the $J + I - 1 \times I$ convolution matrix \mathbf{H}_{mn} can be constructed for both zones as

$$\mathbf{H}_{mn} = \begin{bmatrix} h_{mn}(0) & 0 & 0 \\ \vdots & \ddots & 0 \\ h_{mn}(J-1) & \ddots & h_{mn}(0) \\ 0 & \ddots & \vdots \\ 0 & 0 & h_{mn}(J-1) \end{bmatrix}. \quad (4)$$

Choosing the signal as a unit impulse, $x[k] = \delta[k]$, the pressure signal for all M microphones in any of the zones can be then expressed as the vector of length K ,

$$\mathbf{p} = \mathbf{H}\mathbf{w}, \quad (5)$$

where $K = M(J + I - 1)$ and \mathbf{H} is a $K \times NI$ block matrix composed of all convolution matrices \mathbf{H}_{mn} :

$$\mathbf{H} = \begin{bmatrix} \mathbf{H}_{11} & \cdots & \mathbf{H}_{1N} \\ \vdots & \mathbf{H}_{mn} & \vdots \\ \mathbf{H}_{M1} & \cdots & \mathbf{H}_{MN} \end{bmatrix}. \quad (6)$$

In this work, the control filters for the creation of the sound zones are calculated using the aforementioned WLS

method.⁴ It enables a trade-off between minimizing the mean square of the pressure signal in the dark zone, \mathbf{p}_d , and minimizing the mean square of the difference between the pressure signal in the bright zone, \mathbf{p}_b , and a desired reference pressure signal, \mathbf{p}_r . The weighting factor β is introduced in a convex combination of these two objectives in the WLS cost function J_{WLS} to control the trade-off between them:

$$J_{\text{WLS}} = (1 - \beta) \left[(\mathbf{p}_b - \mathbf{p}_r)^T (\mathbf{p}_b - \mathbf{p}_r) \right] + \beta \mathbf{p}_d^T \mathbf{p}_d. \quad (7)$$

In order to target a specific application and, in general, to increase the robustness of the system to noise and errors, further constraints can be included in the cost function in Eq. (7). For this study, three additional terms were added: the so-known Tikhonov regularization to ensure realistic values for the power of the filters,¹⁵ one term to shape their envelope and remove potential artefacts from pre- and post-ringing,¹⁶ and a term to penalise frequencies outside the intended frequency range.⁴ Respectively, these penalties are introduced and weighted in the cost function by λ and the $NI \times NI$ identity matrix \mathbf{I}_{NI} , ϵ and the matrix \mathbf{R}_E , and γ and the matrix \mathbf{R}_B . The interested reader can find detailed information about these penalties in the respective references. The final cost function for increased robustness is then

$$J_{\text{RWLS}} = (1 - \beta) \left[(\mathbf{H}_b \mathbf{w} - \mathbf{p}_r)^T (\mathbf{H}_b \mathbf{w} - \mathbf{p}_r) \right] + \beta (\mathbf{H}_d \mathbf{w})^T \mathbf{H}_d \mathbf{w} + \lambda \mathbf{w}^T \mathbf{I}_{NI} \mathbf{w} + \epsilon \mathbf{w}^T \mathbf{R}_E \mathbf{w} + \gamma \mathbf{w}^T \mathbf{R}_B \mathbf{w}. \quad (8)$$

Since $\Phi = [(1 - \beta)\mathbf{R}_b + \beta\mathbf{R}_d + \lambda\mathbf{I}_{NI} + \epsilon\mathbf{R}_E + \gamma\mathbf{R}_B]$ is a positive semi-definite matrix, the optimisation of J_{RWLS} with respect to \mathbf{w} is a convex problem. Therefore, setting its gradient equal to zero leads to the optimal filter \mathbf{w} :

$$\mathbf{w} = \Phi^{-1} (1 - \beta) \mathbf{H}_b^T \mathbf{p}_r. \quad (9)$$

Table III in the Appendix contains the values of the parameters used for the design of the control filters and their evaluation.

B. Performance evaluation

Given that the WLS method aims to decrease the pressure in the dark zone and to match the pressure in the bright zone to the reference, its performance can be evaluated based on these goals. The former is commonly assessed in terms of the acoustic contrast ratio (ACR), the ratio between the mean square pressure in the bright and dark zones, respectively, spatially sampled at the M microphones' positions. Averaging the pressures in the bright and dark zones at microphone m , \mathbf{p}_{b_m} and \mathbf{p}_{d_m} , both spatially and in time over the duration K ,¹⁷

$$\text{ACR}_t = 10 \log_{10} \left[\frac{\sum_{k=1}^K \sum_{m=1}^M |\mathbf{p}_{b_m}[k]|^2}{\sum_{k=1}^K \sum_{m=1}^M |\mathbf{p}_{d_m}[k]|^2} \right]. \quad (10)$$

Equivalently, the ACR averaged spatially in the frequency domain for the discrete frequency f is

$$\text{ACR}_f[f] = 10 \log_{10} \left[\frac{\sum_{m=1}^M |\bar{\mathbf{p}}_{b_m}[f]|^2}{\sum_{m=1}^M |\bar{\mathbf{p}}_{d_m}[f]|^2} \right]. \quad (11)$$

The second goal is usually quantified by the mean square difference between the pressure generated in the bright zone and the reference pressure, normalized by the total energy of the latter. After spatial and time averaging, this yields the normalized mean square error (NMSE):¹⁷

$$\text{NMSE}_t = 10 \log_{10} \left[\frac{\sum_{k=1}^K \sum_{m=1}^M |\mathbf{p}_{r_m}[k] - \mathbf{p}_{b_m}[k]|^2}{\sum_{k=1}^K \sum_{m=1}^M |\mathbf{p}_{r_m}[k]|^2} \right]. \quad (12)$$

Similarly, the spatial average in the frequency domain for F discrete frequencies f is

$$\text{NMSE}_f[f] = 10 \log_{10} \left[\frac{\sum_{m=1}^M |\bar{\mathbf{p}}_{r_m}[f] - \bar{\mathbf{p}}_{b_m}[f]|^2}{\frac{1}{F'} \sum_{f'=0}^{F'} \sum_{m=1}^M |\bar{\mathbf{p}}_{r_m}[f']|^2} \right], \quad (13)$$

where f' indicates the F' frequencies over which $\bar{\mathbf{p}}_{r_m}[f']$ is averaged. In this study, they correspond to f and F , respectively. These metrics were used to analyse the results in Sec. IV, obtained from the experiments detailed next.

III. MEASUREMENTS

The tests, results, and analyses in this study are based on RIRs acquired with the synchronized swept-sine (SSS) technique. This method, the measurements performed, the implications of very fast measurements, and the strategy of RIR truncation are detailed, respectively, in Secs. III A, III B, III C, and III D.

A. SSS technique

Introduced by Novak *et al.* in 2015, the SSS technique allows to measure both the impulse response and the harmonic distortion of non-linear systems.¹⁸ It is based on the exponential swept-sine (ESS) technique presented by Farina.¹⁹ In addition to robustness against non-linear

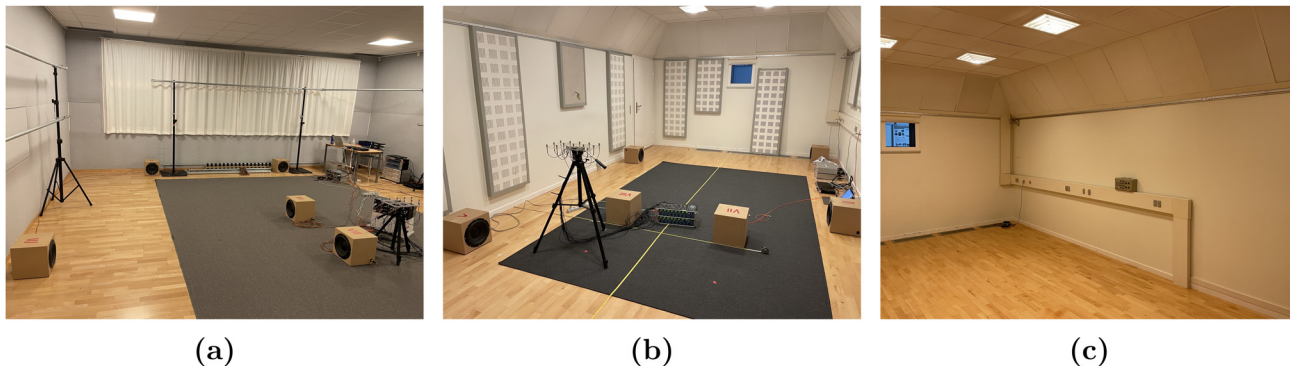


FIG. 2. (Color online) Multichannel room (a) and listening room with (b) and without (c) absorptive thick carpet and different sets of absorptive panels on the walls.

distortion, measurements performed with exponential swept sines have shown higher SNR values compared to other techniques.^{10,20}

The general idea is to capture the output of the system when excited with a sine of exponentially varying frequency and then deconvolve this observation with a specific inverse filter. The response of this inverse filter is usually calculated as the inverse of the Fourier transform of the input signal, or the Fourier transform of the input signal after being time-reversed and introduced a -3 dB/octave amplitude. Therefore, any process the original signal may have, e.g., fading windows, is applied twice. This can be avoided by using the analytical expression for the inverse filter response $\tilde{X}(f)$ introduced by Novak *et al.*,¹⁸

$$\tilde{X}(f) = 2\sqrt{\frac{f}{L}} \exp \left\{ -j2\pi fL \left[1 - \ln \left(\frac{f}{f_1} \right) \right] + j\frac{\pi}{4} \right\}, \quad (14)$$

with

$$L = \frac{T}{\ln \left(\frac{f_2}{f_1} \right)}, \quad (15)$$

where T is the time in seconds to vary from the initial frequency f_1 to the final frequency f_2 of the sweep.

As the aim in this study is only to retrieve the linear component of the RIRs and since it is not intended to characterise the non-linear components, the condition for $L \in \mathbb{Z}$ ¹⁸ that guarantees the synchronisation of the sweeps can be ignored. Moreover, it will be seen that the length T of the sweeps explored in this work makes this condition impossible to fulfill.

B. Experimental setup

The RIR measurements were performed in the Multichannel and Listening rooms of the AIS Section laboratories, at Aalborg University, as shown in Fig. 2. The multichannel room has dimensions $8.12 \times 7 \times 3$ m and is acoustically treated to comply with the ITU-R BS-1116-1 recommendation²¹ for subjective assessment of multichannel

sound systems. The listening room has dimensions $7.8 \times 4.14 \times 3$ m and conforms to the IEC 268-13 report,²² representing normal living rooms to assess loudspeakers for domestic use. Excluding the absorptive ceiling, the room has a thick carpet and 14 removable absorbing panels on the walls, adding around 18 m^2 of absorptive materials, which constitute 13% of the total inner surface. These elements were used and removed to create two different acoustic conditions in the listening room.

Since this work focuses on the range of frequencies below 600 Hz, the reverberation time evaluated over a decay of 20 dB (T_{20}) was measured in $\frac{1}{3}$ octave bands from 50 Hz to 630 Hz. Following the ISO 3382-2:2008 standard²³ for precision measurements, two source positions were used and six microphone (sound analyzer) positions for each, resulting in a total of 12 combinations. The measurements were performed with the interrupted noise method using a B&K 2270 sound analyzer with the reverberation time software BZ-7227 (Bruel & Kjaer, Virum, Denmark). The sound analyzer calculated the time of the 20 dB decay after exciting the rooms in the range of interest. Due the chosen range, a Genelec 1092A active subwoofer and a 1031A studio monitor were used as excitation source (Genelec, Iisalmi, Finland). On average, the multichannel room has $T_{20} = 0.29$ s, and the listening room with all panels and carpet has $T_{20} = 0.55$ s, and $T_{20} = 0.67$ s without these elements. The T_{20} curves by $\frac{1}{3}$ octave bands are shown in Fig. 3.

For the RIR acquisition and sound zone generation, the sources were eight 10-in. custom-made subwoofers

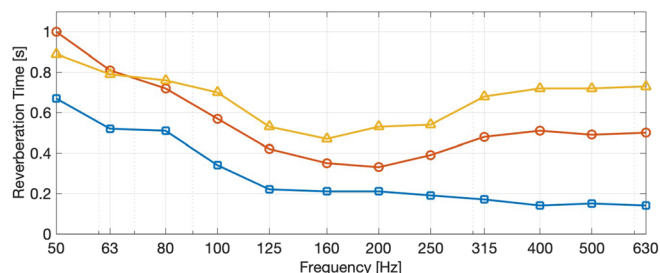


FIG. 3. (Color online) Comparison of the reverberation times T_{20} per $\frac{1}{3}$ octave bands from 50 Hz to 630 Hz, of the multichannel room (□) and the two setups created in the listening room with high (○) and low (△) total absorption. These reflect the different acoustic conditions considered.

distributed as detailed in Figs. 15 and 16 and Table II in the Appendix. The signals were captured by a rectangular array of 15 GRAS 40AZ microphones with GRAS 26CC preamplifiers, arranged in five rows and separated 10 cm from each other (GRAS, Holte, Denmark). An RME UFX+ audio interface was used with two 8-channel RME Mictasy preamplifiers and AD converters (RME Audio, Haimhausen, Germany). The array was placed in the two sound zone positions specified in the Appendix, and as advised by Møller *et al.*,⁹ two sets of measurements were taken at 1 m and 1.05 m height. The first one was used to design the control filters and define the reference pressure p_r , while the second was used to evaluate the performance in the room as detailed in Sec. II B. Though the analysis is not included in this article, it was seen that due to this 5 cm shift in the measurement positions, the values obtained for the performance metrics decrease around 1 to 2 dB. This means on the one hand, that values detailed in Sec. IV can still be slightly higher at some positions. On the other hand, that the sound field control excerpted by the system generalises well around the measurement points used for filter design.

The RIRs were measured at 48 kHz sampling frequency and re-sampled to $f_s = 1.2$ kHz, and to exclude the noise and non-linearities, the last $\frac{1}{3}$ of their samples was discarded. The reference pressure p_r was defined for all setups as the pressure generated by the seventh loudspeaker at the location of the bright zone. For this, the respective RIRs were used and a modelling delay, δ_{mod} , of 20 samples was introduced, corresponding to the maximum propagation time from a loudspeaker to a microphone for $f_s = 1.2$ kHz. In all the experiments, the weighting factor was chosen as $\beta = 0.97$, favouring ACR maximization. Finally, the amplitude of the SSS signals used in the listening room was half the amplitude used in the multichannel room. This was due to the smaller volume of the former with respect to the latter, in order to reach the same root mean square (RMS) input levels in the RME sound card on both rooms.

C. Performing very fast measurements

In order to capture completely the reverberation of the room when measuring RIRs, a short time of silence must be included in the recordings after the SSS signal is played back.^{10,19,20} Then is introduced the “acquisition time” as the total duration of each recording, composed by the duration of the SSS signal [T in Eq. (15)] and the silence afterwards. Hereafter, this quantity will be denoted in boldface, stating the duration of the two components and their respective units. For example, **100ms + 50ms** refers to a 100 ms SSS signal followed by a 50 ms silence.

With the idea of reducing both measurement and processing delays, the use of very short acquisition times was explored. Moreover, to further reduce the SSS duration, the sweeps were limited to the range of interest extended only in the low end, varying then from 15 Hz to 600 Hz. In order to avoid artefacts during playback, the beginning and end of all SSS signals were faded in and out with a raised cosine

TABLE I. Duration of SSS signals, silences, and fades evaluated.

Silences and fades	Duration of SSS signals					
	15 s	1 s	200 ms	100 ms	50 ms	36 ms
Silence 1	1 s	1 s	1 s	1 s	1 s	1 s
Silence 2		0.5 s	100 ms	50 ms	25 ms	18 ms
Fades	20 ms	20 ms	20 ms	20 ms	2 ms	2 ms

window.^{18,20} This further reduced the effective bandwidth of the sweeps.¹¹ As shown in Table I, a 20 ms fading window length was used for SSS signals of 100 ms and above. Conversely and given their very short length, 2 ms fading windows were used for the 50 ms and 36 ms sweeps, avoiding excessive reduction of their energy.

By limiting the temporal and spectral characteristics of the signals, these length, bandwidth, and windowing choices have important effects in both domains. Time-wise, the shorter the signal, the faster each frequency is reproduced in the room. This also implies transferring less energy to the room, translated in lower SNR.^{10,20} Additionally, it is well known that the narrow bandwidth creates ringing in the time domain. Frequency-wise, ripples arising from observing a limited interval of time also reduce the effective range, i.e., over which the spectrum of the SSS signal is flat.¹¹ Thus, in addition to the limited and windowed bandwidth, shorter lengths have as well a negative impact on the frequency response.

Most of these effects are shown in Fig. 4, where the deconvolved measurement signal, its frequency magnitude spectrum, and one of the RIRs acquired are compared for **15s + 1s** and **100ms + 50ms** acquisition times. Notice that the deconvolved signal and the RIR of the former exhibit, respectively, stronger ringing and pre-ringing than the latter case, even after re-sampling. This is due to the choice made for the fade-in window: being only 20 ms long, it is too short to have any effect over the low frequencies of such a long signal. Certainly, longer windows would be a better choice, creating a spectrum with less ripples and flat over a wider frequency range.¹¹ This, however, was a compromise made to avoid using extreme windowing on the shorter acquisition times, while having fewer parameters changing between all times evaluated.

The different lengths of the SSS evaluated, ranging from 36 ms to 15 s, are detailed in Table I. Measurements of 15 s, being the longest, are taken as reference for comparison. Note that 100 ms corresponds to half of the T_{20} of the multichannel room at most of the frequency bands included, and 36 ms is twice the longest propagation time from a loudspeaker to a sound zone among all conditions evaluated. The rest of the SSS lengths were included to also assess intermediate values. In addition, the acquisitions were performed for two lengths of silences: 1 s and half the duration of the SSS, except for 15 s. In Table I, these silences correspond, respectively, to Silence 1 and Silence 2.

D. Truncation of RIRs

The second strategy explored in this work aims to decrease computational effort, which would translate in

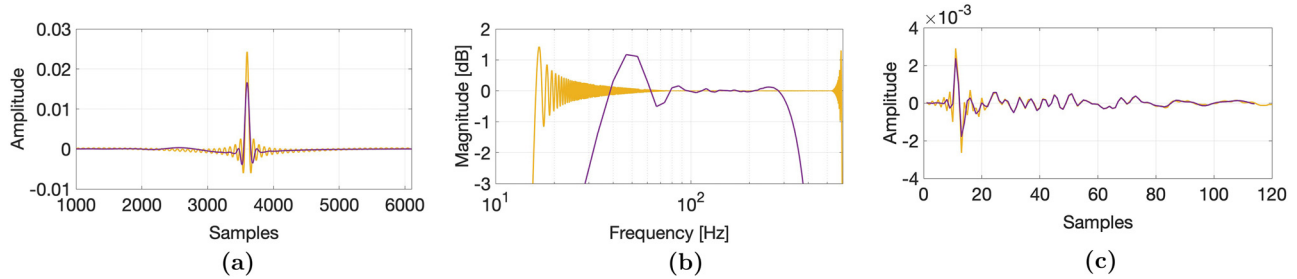


FIG. 4. (Color online) Comparison of the deconvolved measurement signal (a), the frequency magnitude spectrum (b), and a RIR acquired (c) for **15s + 1s** (light line) and **100ms + 50ms** (dark line) acquisition times. Curves in panel (a) are overlapped accordingly for comparison, and only the initial 120 samples are shown for the RIR for the **15s + 1s** in panel (c). Shorter SSS signals and window lengths influence the measurement signals and the RIRs acquired, in both the time and frequency domains.

faster processing. As described by Cadavid *et al.*, the RIRs are truncated by removing all information after a certain time, keeping only the initial samples.¹⁴ By choosing appropriately the truncation time, the minimum amount of information required to achieve a satisfactory performance is used. It is worth noting that results obtained by Cadavid *et al.* are based on simulated RIRs. In this work, these results are validated experimentally by using RIRs measured in rooms with different acoustic conditions.

Finally, this process must not be confounded with the truncation detailed in Sec. III B, necessary to remove unwanted noise and distortion components from the acquired RIRs.

IV. RESULTS

As mentioned before, this work explores two strategies to accelerate high-performance sound zone rendering: reduced RIR acquisition times and RIR truncation. In this section, the performance is evaluated objectively in terms of the ACR and NMSE metrics introduced in Sec. II B. First, in Sec. IV A, the influence of the RIR acquisition time is assessed for the sound zones in the multichannel room, having the lowest T_{20} . Section IV B evaluates the dependency of this first strategy on the reverberation time under the two

acoustic conditions created in the listening room. Finally, the use of truncated RIRs is discussed in Sec. IV C.

A. Influence of acquisition time

As a proof of concept, the effect of long and short acquisition times is initially evaluated under the least reverberant condition, using the RIRs obtained in the multichannel room with $T_{20} = 0.29$ s. Figures 5 and 6 show, respectively, the average ACR_t and $NMSE_t$ obtained for each acquisition time evaluated, both for silence lengths of 1 s and half the length of the SSS. It can be seen that both metrics remain almost unchanged from the best value achievable when SSS signals lasting 100ms or more are used, irrespective of the duration of the silence afterwards. Below this value, the performance degrades considerably when the silence is shorter than 50ms: while the ACR_t decays to 12 dB in Fig. 5, the $NMSE_t$ increases in Fig. 6 more than 2 dB above the minimum obtained. Conversely, both the ACR_t and $NMSE_t$ values remain close to the best value when enough silence is left after the SSS.

In summary, it is clear that RIRs acquired in this room for **100ms + 50ms** perform basically as well as those obtained over **15s + 1s**. Losing 1.7 dB in acoustic contrast and 0.6 dB in error, these values still represent a good performance. Moreover, this may indicate that, under similar

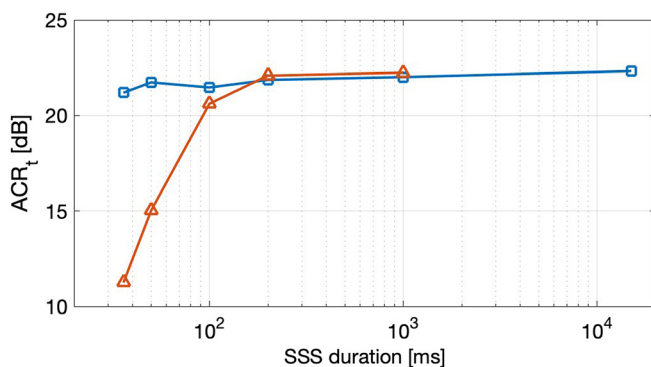


FIG. 5. (Color online) Average acoustic contrast ratio (ACR_t) for each SSS duration with silence lengths of 1 s (\square) and half the length of the SSS (\triangle). Values were averaged over time and positions for RIRs acquired in the multichannel room with $T_{20} = 0.29$ s. Certain acquisition times allow obtaining similar ACR_t values.

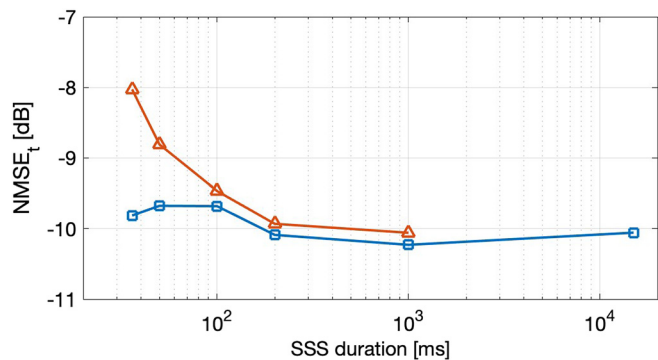


FIG. 6. (Color online) Average normalized mean square error ($NMSE_t$) for each SSS duration with silence lengths of 1 s (\square) and half the length of the SSS (\triangle). Values were averaged over time and positions for RIRs acquired in the multichannel room with $T_{20} = 0.29$ s. As noted in Fig. 5, similar performance can be obtained with different acquisition times.

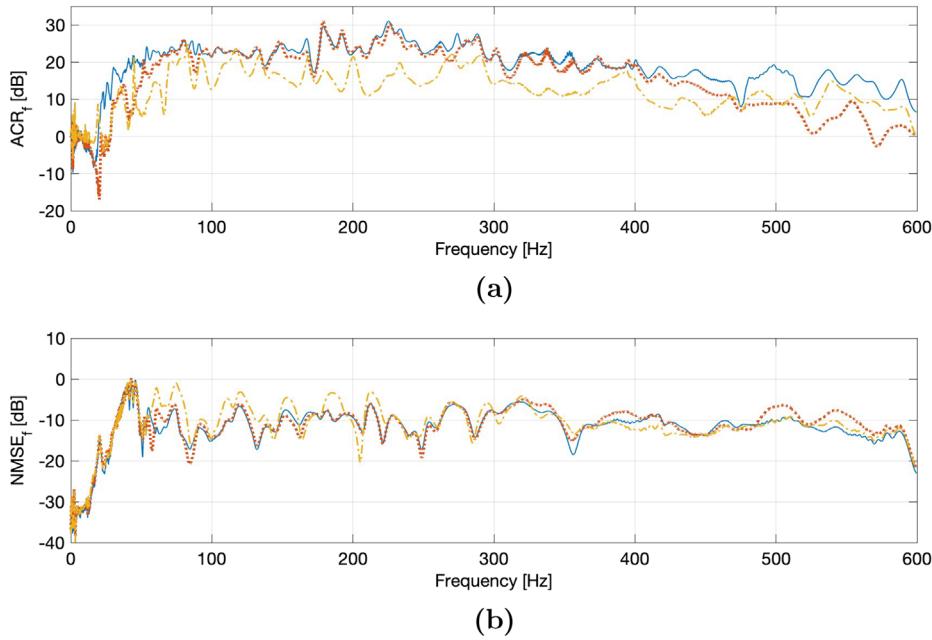


FIG. 7. (Color online) Acoustic contrast ratio (ACR_f) (a) and normalized mean square error ($NMSE_f$) (b) in the frequency domain for **15s + 1s** (solid lines), **100ms + 50ms** (dotted lines), and **50ms + 25ms** (dashed-dotted lines) acquisition times. Values were averaged over microphones for the tests performed in the multichannel room with $T_{20} = 0.29$ s. The first two acquisition times perform almost the same over a wide range of frequencies, while the latter introduces considerable degradation.

acoustic conditions, a single observation can be even shortened to **36ms + 50ms** without compromising the maximum performance.

Based on the previous results, further analyses were performed for **15s + 1s**, **100ms + 50ms**, and **50ms + 25ms** acquisition times. The first case is taken as reference, the second case corresponds to the shortest acquisition time, where performance is almost unaltered, and the third case exhibits a reduced performance. First, the ACR_f and $NMSE_f$ were evaluated using Eqs. (11) and (13), as depicted in Fig. 7. As observed previously, it is clear that the shorter **100ms + 50ms** acquisition follows closely the performance of the long reference **15s + 1s** acquisition, losing some contrast above 400 Hz and degrading the error above 480 Hz. This is mainly due to the fadeout applied with the raised cosine window, which, having the same length for both cases, becomes effective from a lower frequency for the **100ms + 50ms** acquisition. Conversely, the shortest **50ms + 25ms** acquisition time degrades the contrast over the entire frequency range of interest, mainly due to the very short silence afterwards, as deduced from Fig. 5 and other analyses not included. Notice also that no roll-off above 400 Hz is observed in the ACR_f for this acquisition time, due to the very short 2 ms fadeout window.

Similarly, and for the same three acquisition times, the frequency magnitude spectrum of the pressure in the bright zone was compared with the expected reference pressure, calculated with RIRs from **15s + 1s** acquisition time. As shown in Fig. 8 the three acquisition times show little differences with the target between approximately 45 Hz and 450 Hz. Above this range, the **100ms + 50ms** acquisition time presents a greater deviation from the reference pressure, again, due to the low-pass filtering introduced by windowing. In addition, the low-frequency roll-off exhibited by the reference below 45 Hz, inherent to the subwoofers, is also present under all conditions but at lower levels. It was

observed that such an effect is mainly due to the choice of the weight $\beta = 0.97$ and that for this setup, it occurs for values above $\beta = 0.83$. This means that choosing β to prioritize decreasing the energy in the dark zone may narrow the response in the low frequencies. In the lowest range, differences below 25 Hz are due to the lack of control of the system outside its working range.

Finally, it should be kept in mind that the normalization factor of the NMSE in Eq. (13) is the total energy of \mathbf{p}_r , i.e., averaged both spatially and over frequency. Due to this, the $NMSE_f$ in Fig. 7 below 45 Hz has very low values, even though clear differences in magnitude are observed in Fig. 8. In other words, the magnitudes are decreasing and despite visually noticeable, their difference in the $NMSE_f$ is negligible after normalization.

B. Influence of reverberation time

Based on the previous analyses, further experiments were performed the same way in the listening room. More specifically, RIRs were obtained with the same acquisition times as before, and the performance of sound zones based on that information was assessed in terms of the ACR and NMSE. These tests were performed with and without absorption materials on the floor and walls, yielding reverberation times of $T_{20} = 0.55$ s and $T_{20} = 0.67$ s, respectively.

Figures 9 and 10 compare respectively the ACR_t and $NMSE_t$ obtained for both cases. In order to better compare with the multichannel room, curves from Figs. 5 and 6 are also included. At first sight, the resemblance between the curves obtained in both rooms makes clear that the very fast measurement strategy still holds for these new acoustical conditions. That is, certain acquisition times result in performances as good as those obtained with the longest acquisition time. More detailed observation shows that, on the

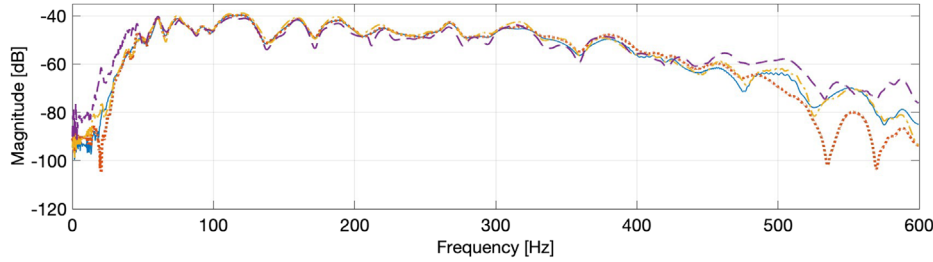


FIG. 8. (Color online) Frequency responses of the pressure in the bright zone for **15s + 1s** (solid lines), **100ms + 50ms** (dotted lines), and **50ms + 25ms** (dashed-dotted lines) acquisition times, compared with the reference pressure (dashed lines). Spectra were evaluated for experiments performed in the multichannel room with $T_{20} = 0.29$ s. Differences in the extremes of the range evaluated are mainly due to the choices of the fading windows and the weighting factor β .

one hand, the performance starts drastically changing below **200ms + 100ms** acquisition time, contrasting with **100ms + 50ms** for the multichannel room. On the other hand, the values obtained in both cases in the listening room differ only slightly from each other, being worse for the second scenario as expected. Notice how these values differ from those obtained for the less reverberant case of the multichannel room, which, having a greater volume and absorption, enables higher performance. The influence of the reverberation time is then clearly observed both in the best achievable

performance metric values, and in the minimum acquisition time required to achieve these. Finally, it is worth noticing the **100ms + 1s** case, presenting a small dip both in the ACR and NMSE curves under both reverberant conditions. Analyses not included in this work confirmed that this is also due to the choice of the length of the fades. More specifically, comparing the **100ms + 1s** acquisition time with **15s + 1s** and **50ms + 1s**, the choice of the fade-out window worsens the ACR and improves NMSE from around 400 Hz, impacting the overall values observed in Figs. 9 and 10.

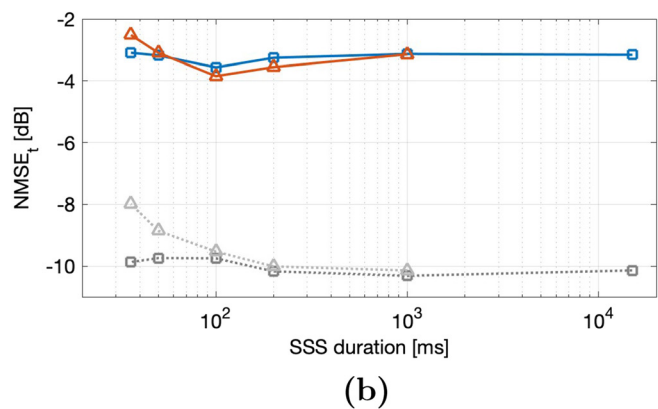
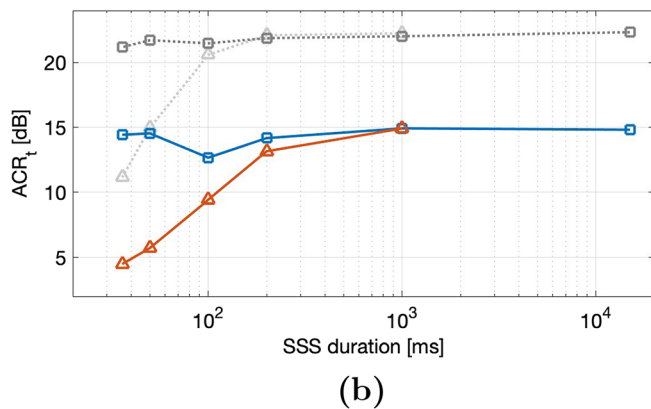
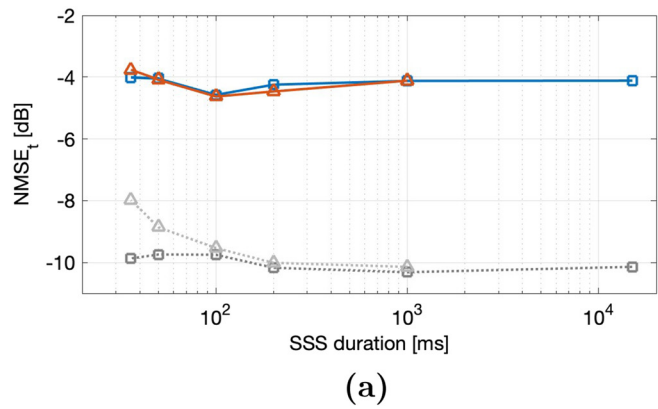
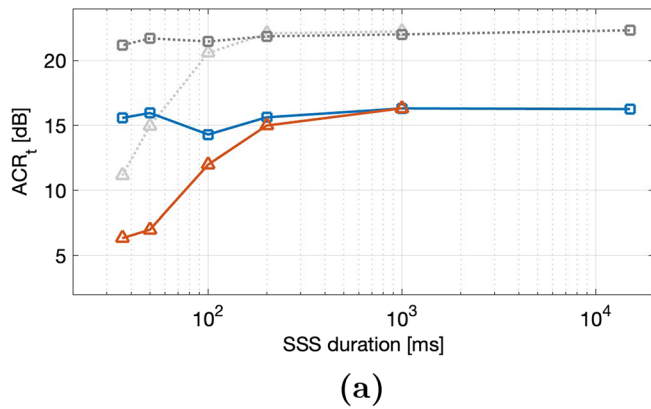


FIG. 9. (Color online) Average acoustic contrast ratio (ACR_t) for each SSS duration with silence lengths of 1 s (\square) and half the length of the SSS (\triangle). Averages were calculated over time and positions for RIRs acquired in the listening room with (a) $T_{20} = 0.55$ s and (b) $T_{20} = 0.67$ s. Curves in the gray dotted line are taken from Fig. 5 and included for comparison between rooms. Despite the differences in the values, similar trends can be observed across acquisition times in the average ACR_t .

FIG. 10. (Color online) Average normalized mean square error ($NMSE_t$) for each SSS duration with silence lengths of 1 s (\square) and half the length of the SSS (\triangle). Averages were calculated over time and positions for RIRs acquired in the listening room with (a) $T_{20} = 0.55$ s and (b) $T_{20} = 0.67$ s. Curves in the gray dotted line are taken from Fig. 5 and included for comparison between rooms. The different acoustic conditions in the listening room introduce considerable degradation of the reproduction error.

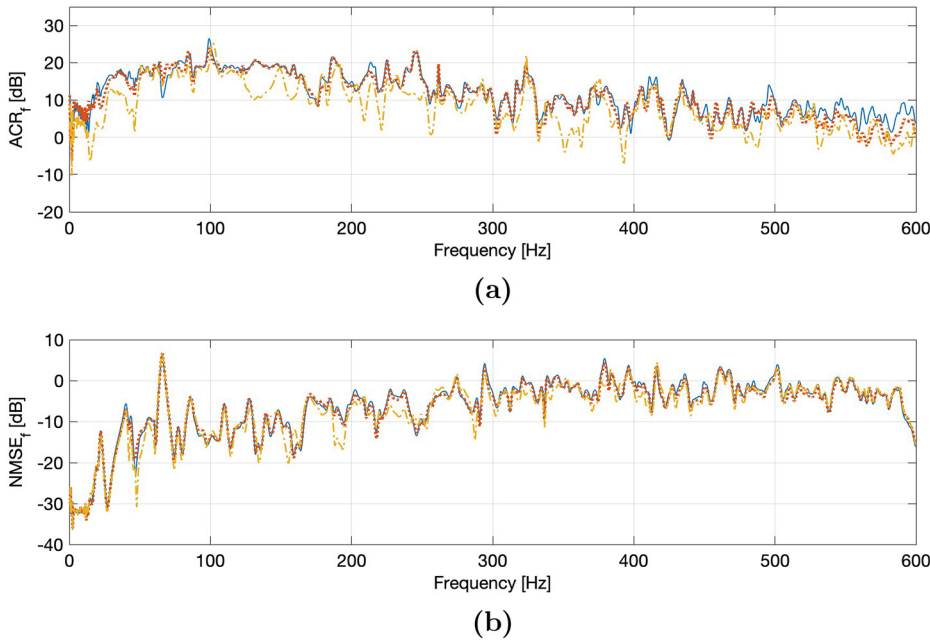


FIG. 11. (Color online) Acoustic contrast ratio (ACR_f) (a) and normalized mean square error ($NMSE_f$) (b) in the frequency domain for **15s + 1s** (solid lines), **200ms + 100ms** (dashed lines), and **100ms + 50ms** (dashed-dotted lines) acquisition times. Values were averaged over microphones for the experiments performed in the listening room with low absorption and $T_{20} = 0.67$ s. The effect of the new acoustic conditions, manifested in the uneven behaviour and overall performance degradation, can be clearly observed.

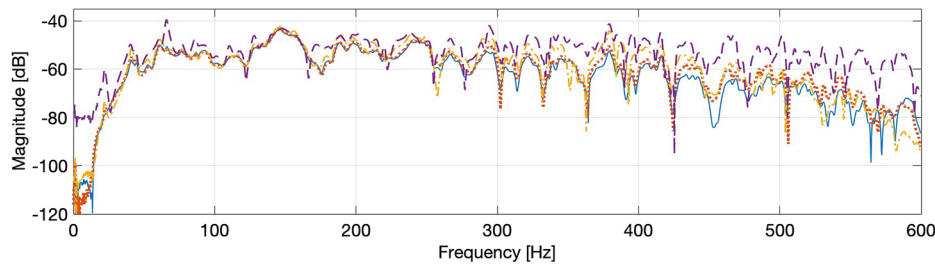


FIG. 12. (Color online) Frequency responses of the pressure in the bright zone for **15s + 1s** (solid lines), **200ms + 100ms** (dotted lines), and **100ms + 50ms** (dashed-dotted lines) acquisition times, compared with the reference pressure (dashed lines). Spectra were evaluated for experiments performed in the listening room with low absorption and $T_{20} = 0.67$ s. Above around 60 Hz and due to the highly reflective acoustic environment, the higher the frequency, the greater the differences between the four frequency responses.

Correspondingly, the same analyses performed in the frequency domain for the ACR_f , $NMSE_f$, and frequency magnitude spectrum in the bright zone are shown in Figs. 11 and 12. Based on the previous results, these experiments were performed with **15s + 1s**, **200ms + 100ms**, and **100ms + 50ms** acquisition times, the longest time being the reference and the shorter times corresponding to values above and below the value at which considerable changes in the metrics start occurring. In addition, the graphs included in Figs. 11 and 12 correspond only to the second case, with the highest T_{20} , because, despite the difference in the reverberation time, the results obtained in the two cases showed to be very similar.

As observed in Fig. 11, the ACR_f obtained for **200ms + 100ms** acquisition follows closely the values obtained with **15s + 1s** acquisition for almost all frequencies, decaying slightly above 500 Hz. Conversely, **100ms + 50ms** acquisition presents different values over the entire frequency range, deviating sometimes more than 10 dB from

the reference. This is not the case for the error, where differences among the acquisition times are negligible over most of the frequency range, deviating up to 5 dB at specific frequencies.

Regarding the frequency magnitude spectrum in the bright zone, it can be seen in Fig. 12 that from 40 Hz to 300 Hz, the pressure obtained with the three acquisition times deviates from the reference pressure with up to 10 dB and only at specific frequencies. Above approximately 300 Hz, the deviation for all acquisition times gradually increases and differences of more than 20 dB are observed. This is due to the higher number of reflections in the room, clearly observable in the number of notches in the spectra, including the reference. Moreover, note the differences between these spectra and the multichannel room case in Fig. 8.

Finally, it is observed in Fig. 11 and 12 how the performance worsens below approximately 50 Hz and above

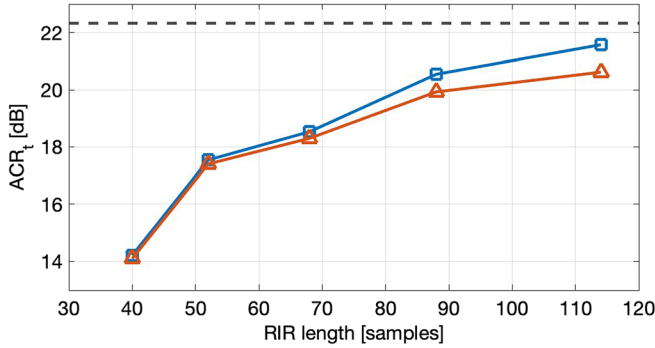


FIG. 13. (Color online) Average acoustic contrast ratio (ACR_t) of sound zones designed with RIRs from **15s + 1s** (\square) and **100ms + 50ms** (\triangle) acquisition times, both truncated at different times from 40 to 114 samples. Values were compared with the reference $ACR_t = 22.3$ dB obtained with the complete RIRs from **15s + 1s** acquisition time (dotted line). High ACR_t values can be achieved after dismissing a certain amount of information in the RIRs.

200 Hz, presenting the smallest deviations inside that range. This contrasts again with the least reverberant scenario in the multichannel room, where the performance starts degrading only above 400 Hz. This indicates two things. On the one hand, lower-frequency ranges allow the creation of higher-quality sound zones in different acoustic scenarios, showing higher robustness even under highly reflective conditions. On the other hand, shorter reverberation times are required to achieve higher performance.

C. Truncation of measured RIRs

In these final experiments, the second strategy described in Sec. III D is validated with some of the RIRs used previously to evaluate the effect of very short acquisition times. More specifically and based on the results detailed in Sec. IV A, the RIRs obtained in the multichannel room from **15s + 1s** and **100ms + 50ms** acquisition times are in addition truncated, in order to assess the impact of further reduced information in realistic conditions.

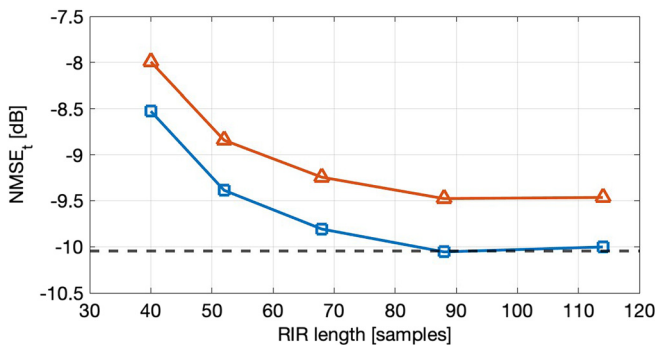


FIG. 14. (Color online) Average normalized mean square error ($NMSE_t$) of sound zones designed with RIRs from **15s + 1s** (\square) and **100ms + 50ms** (\triangle) acquisition times, both truncated at different times from 40 to 114 samples. Values were compared with the reference $NMSE_t = -10$ dB obtained with the complete RIRs from **15s + 1s** acquisition time (dotted line). $NMSE_t$ values equal or close to such a reference can be achieved while omitting certain information in the RIRs evaluated.

TABLE II. x and y coordinates of the source and sound zone positions in both rooms.^a

Source or zone	Multichannel room		Listening room	
	x	y	x	y
Source 1	0.35	1.48	7.6	3.94
Source 2	0.32	5.14	3.7	3.94
Source 3	4.10	6.67	0.3	3.94
Source 4	7.78	6.80	1.85	0.3
Source 5	7.78	0.25	5.55	0.3
Source 6	4.10	0.34	7.6	2.07
Source 7	3.54	4.42	4.7	2.67
Source 8	4.63	4.43	4.7	1.47
Bright zone	3.57	3.85	5.2	2.67
Dark zone	4.63	3.85	5.2	1.47

^aValues are given in meters.

The minimum truncation time is 40 samples, corresponding to twice the modelling delay, δ_m . The maximum truncation time was set to 114 samples, corresponding to the length of the **100ms + 50ms** RIRs, meaning that no truncation was performed in that case. Three additional truncation times are distributed logarithmically between these two values, being 52, 68, and 88 samples. Having the same truncation times for both cases ensures that the effects of truncation are identical in both RIRs, such that differences between performance metrics will be due to the difference in acquisition times.

Figures 13 and 14 show the average ACR_t and $NMSE_t$ obtained for sound zones designed with the selected RIRs for the five truncation times included in this evaluation. These are compared with the performance metrics achieved with the complete RIRs measured with **15s + 1s** acquisition time, evaluated earlier in Figs. 5 and 6. As expected, the shorter the truncation time, i.e., reduced information, the worse the performance. However, it is interesting to note that using only 88 samples from the RIRs, the ACR_t is still equal to or greater than 20 dB, and the $NMSE_t$ is equal to or smaller than -9.5 dB. In other words, for these tests and with respect to the reference without truncation, the

TABLE III. Parameters and values used for the design and evaluation of the control filters.

Name	Value	Description
c	340 m/s	Speed of sound
f_s	1200 Hz	Sampling frequency
I	100 samples	Control filters' length
δ_{mod}	20 samples	\mathbf{p}_r modelling delay
β	0.97	Effort weight
λ	1×10^{-3}	Regularization factor
γ	1×10^{-5}	Weight for frequency range
ϵ	1×10^{-7}	Weight for envelope shaping
μ_l^a	1×10^{12}	Pre-ringing envelope shaper
μ_u^a	1×10^7	Post-ringing envelope shaper
$(c_l, c_u)^a$	(6, 34) samples	Range with no envelope

^aName according to Ref. 16.

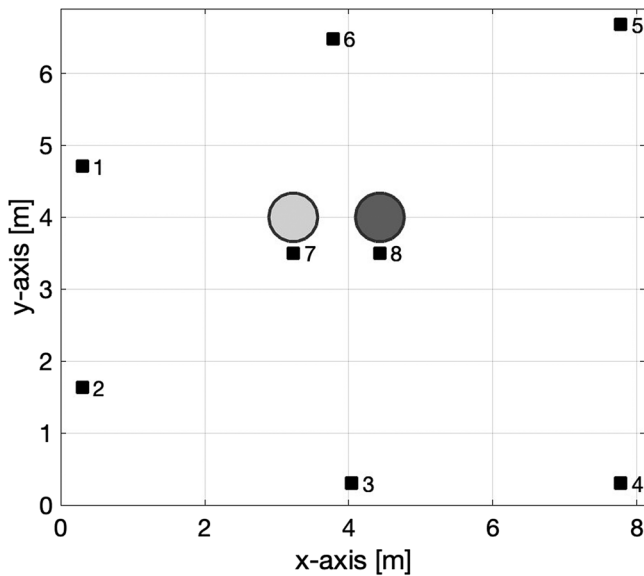


FIG. 15. Distribution of sources (□) and sound zone positions (○) set to acquire the RIRs in the multichannel room.

truncated RIRs result in a reduction in the ACR_t of 2.4 dB and 0.75 dB for the **100ms + 50ms** and **15s + 1s** acquisition times, respectively. Moreover, the $NMSE_t$ of the RIRs measured with **100ms + 50ms** acquisition time is 0.5 dB worse than the reference, while the RIRs measured with **15s + 1s** acquisition time, having the largest truncation, show no degradation. Finally, note that for $f_s = 1.2$ kHz, 88 samples are approximately 74 ms, around $\frac{1}{4}$ of the T_{20} of the multichannel room.

V. CONCLUSIONS

Two strategies to decrease the time and data required to render low-frequency sound zones were described and validated experimentally. In both cases, the performance was evaluated mainly in terms of the ACR and the NMSE.

The first one aims to reduce the acquisition time of the RIRs while achieving maximal performance. To evaluate this, measurements were performed using the SSS method

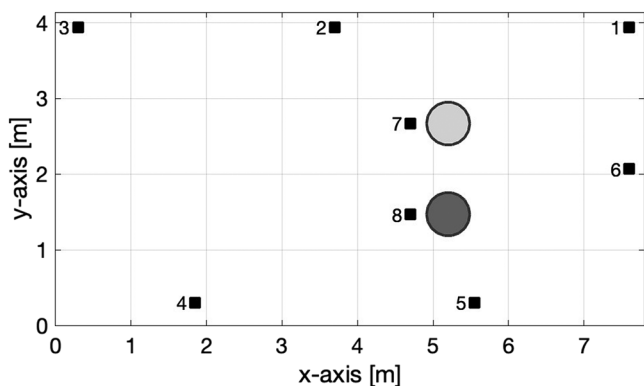


FIG. 16. Distribution of sources (□) and sound zone positions (○) set to acquire the RIRs in the listening room.

with narrow bandwidth and different lengths, both for the measurement signal and the silence afterwards. This validation was performed under three acoustic conditions, each presenting different absorption properties and therefore reverberation times.

It was observed that the maximum performance obtained with the longest RIR acquisition times can also be achieved with RIRs measured in a small fraction of that time. More specifically and under certain acoustic conditions, the SSS signals can be as short as twice the maximum distance between the loudspeakers and microphones. In addition, the silence after the SSS signal can be reduced even below the reverberation time of the room, but it still has to be long enough to capture a relevant amount of energy and reflections. This was experimentally shown to hold in different acoustic scenarios.

For the second strategy, the RIRs measured with the longest and shortest acquisition times reaching high performance were truncated at five lengths. The tests performed showed that, for short reverberation times, the RIRs can still be reduced to approximately one-fourth of the average reverberation time before considerably affecting the performance. This strategy proved to be highly effective both for RIRs obtained with long and short acquisition times. In any case, such a reduction of information may translate in faster processing, lower computational effort, and simpler hardware demands.

It can then be said that sound zones rendering can be highly optimised by performing very fast measurements and reducing the amount of information in the RIRs, while still achieving high performance. However, and despite these conclusions, creating sound zones still requires certain amount of time, making it infeasible for applications where changes in the system and/or environment are to be accounted for in real-time. Therefore, further work should be focused on faster estimation and tracking of the RIRs, increased robustness of the control filters, and the subjective assessment of the methods implemented. In addition, new experiments should be made regarding the fade windows, in order to find the optimal choice for different acquisition times.

ACKNOWLEDGMENTS

The authors would like to thank Antonin Novak and Claus Vestergaard for their valuable time and help. This project has received funding from the European Union's (EU) Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Actions Grant No. 956369. In addition, the research leading to these results has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program/ERC Consolidator Grant: SONORA (No. 773268). This paper reflects only the authors' views, and the Union is not liable for any use that may be made of the contained information.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

APPENDIX: SETUPS AND PARAMETERS

Figures 15 and 16 and Table II detail the positions of the sources and sound zones in the different rooms. Based on these, the modelling delay, δ_{mod} , was defined, corresponding to the delay introduced by the propagation of sound from the farthest loudspeaker to the bright zone. Table III details the values of the parameters required to design and evaluate the control filters. The weight β was chosen to favour ACR maximization, and in connection with the other parameters, their values were selected so that the best performance could be achieved with the same combination for the three acoustic conditions evaluated.

¹O. Kirkeby, P. A. Nelson, F. Orduna-Bustamante, and H. Hamada, "Local sound field reproduction using digital signal processing," *J. Acoust. Soc. Am.* **100**(3), 1584–1593 (1996).

²J.-W. Choi and Y.-H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," *J. Acoust. Soc. Am.* **111**(4), 1695–1700 (2002).

³J.-H. Chang and F. Jacobsen, "Sound field control with a circular double-layer array of loudspeakers," *J. Audio Eng. Soc.* **131**(6), 4518–4525 (2012).

⁴M. 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.* **23**(11), 1869–1878 (2015).

⁵W. Druyvesteyn and J. Garas, "Personal sound," *J. Audio Eng. Soc.* **45**(9), 685–701 (1997).

⁶L. Vindrola, M. Melon, J.-C. Chamard, and B. Gazengel, "Pressure matching with forced filters for personal sound zones application," *J. Audio Eng. Soc.* **68**(11), 832–842 (2020).

⁷M. B. Møller and M. Olsen, "On in situ beamforming in an automotive cabin using a planar loudspeaker array," in *23rd International Congress on Acoustics, Integrating 4th EAA Euroregio 2019*, Aachen, Germany (International Commission for Acoustics, Madrid, 2019), pp. 1109–1116.

⁸T. Betlehem, W. Zhang, M. A. Poletti, and T. D. Abhayapala, "Personal sound zones: Delivering interface-free audio to multiple listeners," *IEEE Signal Process. Mag.* **32**(2), 81–91 (2015).

⁹M. B. Møller, J. K. Nielsen, E. Fernandez-Grande, and S. K. Olesen, "On the influence of transfer function noise on sound zone control in a room," *IEEE/ACM Trans. Audio. Speech. Lang. Process.* **27**(9), 1405–1418 (2019).

¹⁰G.-B. Stan, J.-J. Embrechts, and D. Archambeau, "Comparison of different impulse response measurement techniques," *J. Audio Eng. Soc.* **50**(4), 249–262 (2002).

¹¹P. Guidorzi, L. Barbaresi, D. D'Orazio, and M. Garai, "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**, 1611–1616 (2015).

¹²V. 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.* **151**(1), 334–345 (2022).

¹³M. Ebri, N. Strozzi, F. M. Fazi, A. Farina, and L. Cattani, "Individual listening zone with frequency-dependent trim of measured impulse responses," in *149th Audio Engineering Society Convention*, Online. (Audio Engineering Society, New York, 2020).

¹⁴J. 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*, St. Pölten, Austria (Association for Computing Machinery, New York, 2022), pp. 239–245.

¹⁵S. J. Elliott, J. Cheer, J.-W. Choi, and Y. Kim, "Robustness and regularization of personal audio systems," *IEEE Trans. Audio. Speech. Lang. Process.* **20**(7), 2123–2133 (2012).

¹⁶M. Møller and M. Olsen, "Sound zones: On envelope shaping of fir filters," in *Proceedings of the 24th International Congress on Sound Vibration 2017 (ICSV 24)*, London, UK (International Institute of Acoustics and Vibration, Auburn, AL, 2017), pp. 613–620.

¹⁷M. B. Møller and J. Østergaard, "A moving horizon framework for sound zones," *IEEE/ACM Trans. Audio Speech Lang. Process.* **28**, 256–265 (2020).

¹⁸A. Novak, P. Lotton, and L. Simon, "Synchronized swept-sine: Theory, application, and implementation," *J. Audio Eng. Soc.* **63**(10), 786–798 (2015).

¹⁹A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society 108th Convention*, Paris France (Audio Engineering Society, New York, 2000).

²⁰S. Müller and P. Massarani, "Transfer-function measurement with sweeps," *J. Audio Eng. Soc.* **49**(6), 443–471 (2001).

²¹International Telecommunication Union Radiocommunication Assembly, "Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems," Recommendation ITU-R BS.1116-1 (10/97), International Telecommunication Union Radiocommunication Assembly, Geneva (1997).

²²International Electrotechnical Commission, "Sound system equipment—Part 13: Listening tests on loudspeakers," Technical report IEC 268-13, International Electrotechnical Commission, Geneva (1998).

²³International Organization for Standardization. *Acoustics—Measurement of Room Acoustic Parameters—part 2: Reverberation Time in Ordinary Rooms*, Standard ISO 3382-2:2008. (International Organization for Standardization, Geneva, 2008).