Performance of Low Frequency Sound Zones Based on Truncated Room Impulse Responses

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

Published in:
Proceedings of the 17th International Audio Mostly Conference

DOI (link to publication from Publisher):
10.1145/3561212.3561248

Publication date:
2022

Document Version
Accepted author manuscript, peer reviewed version

Performance of Low Frequency Sound Zones Based on Truncated Room Impulse Responses

José Cadavid
jmct@es.aau.dk
Dept. of Electronic Systems,
Aalborg University
Aalborg, Denmark

Martin Bo Møller
mim@bang-olufsen.dk
Bang & Olufsen a/s
Struer, Denmark

Søren Bech
sbe@bang-olufsen.dk
Bang & Olufsen a/s
Struer, Denmark

Toon van Waterschoot
toon.vanwaterschoot@esat.kuleuven.be
Dept. of Electrical Engineering
(ESAT), KU Leuven
Leuven, Belgium

Jan Østergaard
jo@es.aau.dk
Dept. of Electronic Systems,
Aalborg University
Aalborg, Denmark

ABSTRACT

Using spatially distributed loudspeakers and properly designed control filters, it is possible to generate sound zones that play different audio contents in different regions of the same room. For low frequency content, the design of control filters relies on the room impulse responses (RIRs) between each loudspeaker and the desired listening positions. Estimates of the RIRs can be obtained by distributing wireless microphones within the sound zones and thereby systematically acquiring sufficient knowledge about the acoustical characteristics of the room and loudspeakers. Longer acquisition times would generally lead to better estimates of the RIRs but would also introduce processing delays, which is undesirable in cases where time-varying RIRs are to be compensated. In addition, shorter RIRs may imply lower computational complexity.

In this work, control filters were calculated using truncated versions of the RIRs in order to simulate the effect of a reduced acquisition time. The performance was evaluated in terms of the acoustic contrast ratio (ACR) and it was seen that within a certain limit, doubling the acquisition time increases the ACR around 4 dB or more. Moreover, to keep the sound pressure low in the dark zone, only a limited part of the reverberation tail is required.

CCS CONCEPTS

• **Applied computing** → **Sound and music computing**; **Physics**;
• **Networks** → Network performance modeling.

KEYWORDS

sound zones, room impulse response, sound field control, wireless acoustic network

1 INTRODUCTION

One of the several approaches within sound field control aims to create small, well-defined audio regions at given locations, also known as sound zones [1]. The result of such control is two or more zones where different audio content is played clearly within each of them, so that it is almost inaudible in the rest of the zones [2, 3]. In general, sound zones are created by controlling the signals reproduced by an array of loudspeakers, while spatially sampling the pressure of the zones with several microphones distributed inside. The control signals result from the convolution of the audio content to play and a set of control filters, which are based on the specific room impulse response (RIR) obtained between each loudspeaker and each microphone. For each bright zone defined in the system, one control filter per loudspeaker is calculated. Usually, a specific audio content is intended to be reproduced in a so-called "bright" zone while keeping pressure levels low in a "dark" one. Using superposition, this configuration can be swapped to make the dark zone the bright one, so that a different audio content can be played at that location.

Currently, these systems can be implemented under a fixed or "closed-form" solution, meaning that no changes in the electro-acoustical chain can be accounted for. Thus, they are highly sensitive to all kinds of variations such as room temperature [4], loudspeaker non-linearities [5], position of objects and even more, listener movement. In addition, the emerging Internet of Sounds allows to set a wireless acoustical sensor network (WASN) to assist with the sound field control. In that case, the performance of the changing sound zones can also be affected by noise and estimation errors [6, 7], communication artefacts [8, 9] and computational resources [10].

In order to compensate for some of these changes in real-time, a low-latency RIRs acquisition is required for a fast update of the control filters. In general, the longer the RIRs, the more information about the room can be captured but, the greater the processing latency. Hence, it is necessary to know how much new information of the room must be acquired with the RIRs, so that the sound zones can be adapted satisfactorily. To answer part of this question, this paper studies the influence of the length of the RIRs used to create sound zones at the low frequencies, on their performance. This impact is evaluated under objective metrics, such as the Acoustic Contrast Ratio (ACR), the Normalized Mean Squared Error (NMSE) and the sound pressure levels.

The control method implemented for creating sound zones is explained in section **2**. Section **3** details the RIRs truncation evaluated in this work, as well as the different simulations and analyses performed. The results and conclusions derived are presented in sections **4** and **5** respectively.

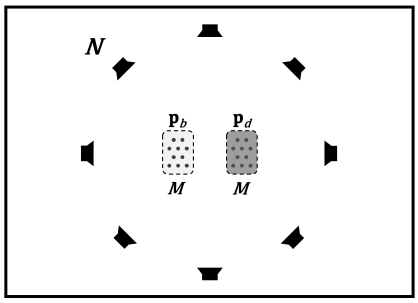

Figure 1: Illustration of a setup for creating two sound zones with N loudspeakers and M control microphones per zone.

2 WEIGHTED LEAST-SQUARES METHOD

Consider the system depicted in Figure 1: N loudspeakers distributed in a room are fed with N specific control signals, so that the resulting sound field comprises two well-defined sound zones: a "Bright" and a "Dark" one. A set of M microphones is placed inside each zone in order to monitor and control their sound pressure, $\mathbf{p}_b$ and $\mathbf{p}_d$ for the bright and dark zone respectively. For such system, $2MN$ Room Impulse Responses (RIRs) are required: one for each microphone-loudspeaker pair. These must be taken into account when designing the N control filters $\mathbf{w}_n$, $n = \{1, \cdots, N\}$, which convolved with the audio content, generate the control signals.

In the time domain and for control filters of length I, all the RIRs of the two zones can be arranged in the $M(J+I-1) \times NI$ matrices $\mathbf{H}_b$ and $\mathbf{H}_d$. For example, for the bright zone:

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

where each $\hat{\mathbf{H}}_{b_{mn}}$ is a $(J+I-1)\times I$ convolution matrix containing the RIR of length J between source n and control point m in the bright zone as:

$$\hat{\mathbf{H}}_{b_{mn}} = \begin{bmatrix} \mathbf{h}_{b_{mn}}(0) & 0 & 0 \\ \vdots & \ddots & 0 \\ \mathbf{h}_{b_{mn}}(J-1) & \ddots & \mathbf{h}_{b_{mn}}(0) \\ 0 & \ddots & \vdots \\ 0 & 0 & \mathbf{h}_{b_{mn}}(J-1) \end{bmatrix}. \tag{2}$$

In order to find a signal-agnostic solution, the input signal can be defined as an unitary impulse $\delta(n)$ [11]. Thus, the pressure inside the two zones can be expressed as:

$$\mathbf{p}_b = \mathbf{H}_b \mathbf{w}\,, \tag{3}$$

$$\mathbf{p}_d = \mathbf{H}_d \mathbf{w}\,, \tag{4}$$

with $\mathbf{w} = \begin{bmatrix} \mathbf{w}_1^T & \mathbf{w}_2^T & \cdots & \mathbf{w}_N^T \end{bmatrix}^T$ the $NI \times 1$ vector composed by the N control filters of length I. Both $\mathbf{p}_b$ and $\mathbf{p}_d$ are $M(J+I-1)\times 1$ vectors.

At this point, $\mathbf{w}$ can be found with the Acoustic Contrast Control method by Choi et al. [12] by defining a cost function that maximizes the ratio of the mean square pressure in the bright zone, with respect to that one in the dark zone. However, if a least-squares solution is to be implemented, a reference source must be defined for the bright zone and its corresponding IR must be obtained. The idea behind this is to reconstruct in the bright zone the sound field that the reference source would create in the same location [3, 13]. In this case, the cost function must minimize the difference between the sound field the reference source would create, and the one actually obtained by the system. This is the principle behind the Least-Squares or Pressure-Matching method and those based on it.

As proposed initially by Chang and Jacobsen [14] and adapted to the time domain by Galves et al. [11], it is possible to combine the optimization goals of both methods. By defining a cost function with both objectives and a parameter β to weight them, the control filters can be designed to perform anywhere between these two solutions. In other words and depending on the value of β, it is possible to design the control filters to only aim to reconstruct the reference sound field in the bright zone or, to only aim to minimize the pressure in the dark zone or, to do both under the trade-off defined by β and $(1-\beta)$. Notice that $0 \leqslant \beta \leqslant 1$.

$$J_{WLS} = (1-\beta)[(\mathbf{p}_b - \mathbf{p}_r)^T(\mathbf{p}_b - \mathbf{p}_r)] + \beta \mathbf{p}_d^T \mathbf{p}_d. \tag{5}$$

Equation (5) details the cost function described above, with β the weighting parameter, $\mathbf{p}_b$ and $\mathbf{p}_d$ the pressures in the bright and dark zones, and $\mathbf{p}_r$ the reference pressure. This target pressure is obtained by convolving the IR of the desired reference source with the input signal, an unitary impulse $\delta(n)$ in this formulation.

Since solving J_{WLS} to find $\mathbf{w}$ involves a matrix inversion, adding a regularization term λ will control the amplitude of the filters and increase the robustness of the solution [11, 15]. Such regularization can be emphasized over a given frequency range defined by a filter included in the cost function inside the block diagonal matrix $\mathbf{R}_F$, so that frequencies outside it are penalized depending on the factor γ [11, 16]. Moreover, as proposed by Møller and Olsen [16], the envelopes of the filters can be shaped to reduce their pre- and post-ringing, which could potentially avoid audible artifacts. Including such penalties as detailed in the above-mentioned publications and, substituting (3) and (4), the cost function would then be:

$$J_{WLSR} = (1-\beta)[(\mathbf{H}_b\mathbf{w} - \mathbf{p}_r)^T(\mathbf{H}_b\mathbf{w} - \mathbf{p}_r)] \; + \; \beta(\mathbf{H}_d\mathbf{w})^T\mathbf{H}_d\mathbf{w} + \lambda\mathbf{w}^T\mathbf{I}_D\mathbf{w} \; + \; \gamma\mathbf{w}^T\mathbf{R}_F\mathbf{w} \; + \; \epsilon\mathbf{w}^T\mathbf{R}_E\mathbf{w}\,, \tag{6}$$

where $\mathbf{I}_D$ is the $NI\times NI$ identity matrix and ϵ governs the penalty for shaping the filter envelope, defined in the diagonal matrix $\mathbf{R}_E$.

In order to find the global minimum of J_{WLSR}, its partial derivative with respect to $\mathbf{w}$ must be set equal to zero. As stated in [11], this will be a valid approach if the matrix

$$[(1-\beta)\mathbf{H}_b^T\mathbf{H}_b + \beta\mathbf{H}_d^T\mathbf{H}_d + \lambda\mathbf{I}_D + \gamma\mathbf{R}_F + \epsilon\mathbf{R}_E]$$

is positive definite. Solving such expression for $\mathbf{w}$, the analytical solution is found as:

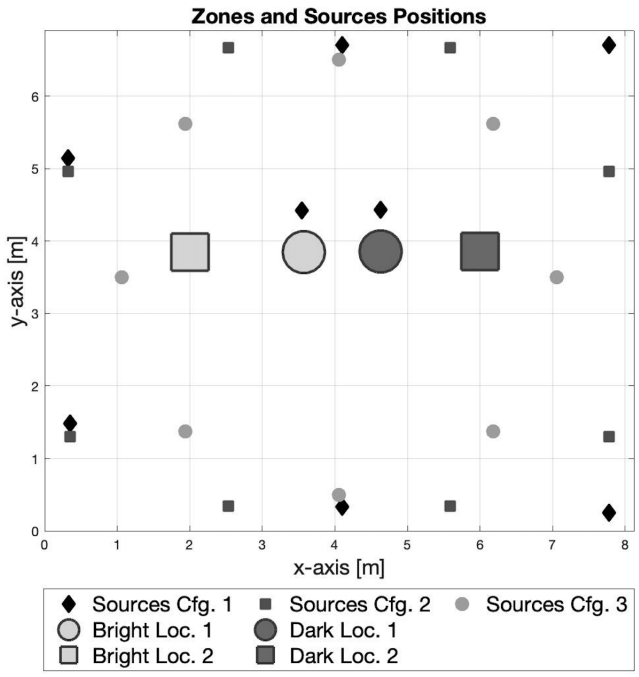

Figure 2: Room, source configurations and sound zone positions used to perform the different simulations and analyses.

$$\mathbf{w} = [(1-\beta)\mathbf{H}_b^T\mathbf{H}_b + \beta\mathbf{H}_d^T\mathbf{H}_d + \lambda\mathbf{I}_D + \gamma\mathbf{R}_F + \epsilon\mathbf{R}_E]^{-1}(1-\beta)\mathbf{H}_b^T\mathbf{p}_r. \quad (7)$$

As explained before, the control signals fed into the loudspeakers are obtained by filtering the signal intended to play, with the vector of control filters $\mathbf{w}$.

From the cost function in (6), two metrics can be derived to evaluate the performance of the solution obtained: the Acoustic Contrast Ratio (ACR) and the Normalized Mean Squared Error (NMSE). The former is the ratio of the mean square pressure in the bright zone with respect to the the same quantity in the dark one, when both are sampled under the same conditions. The NMSE quantifies the differences between the pressure in the bright zone and the reference pressure $\mathbf{p}_r$ [12, 17].

3 SIMULATIONS

This section describes the simulations performed to evaluate the effect of the duration of the RIRs on the performance of the sound zones. Henceforth, we will refer to loudspeakers and microphones as sources and control points respectively. Using a real room as model for simulations, six cases with different zones and source setups were considered. In addition, the same sequence of ten seconds of white noise $\mathbf{s}_w$ was used as input signal for all evaluations.

The room, depicted in Figure 2, has $8.12 \times 7 \times 3$ m dimensions, and is acoustically treated to have a $T20_{mid}$ of 0.2 s. However, since this work focuses in the range of frequencies below 600 Hz, the average value of 0.4 s from 50 to 315 Hz was considered for the simulations [18].

Table 1: Definition of the six cases for the different source configurations and zones locations.

	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6
Source	Cfg. 1	Cfg. 1	Cfg. 2	Cfg. 2	Cfg. 3	Cfg. 3
Zones	Loc. 1	Loc. 2	Loc. 1	Loc. 2	Loc. 1	Loc. 2

Table 2: X and Y coordinates of sources' position in the three different configurations.

	Sources Cfg. 1		Sources Cfg. 2		Sources Cfg. 3	
#	x	y	x	y	x	y
1	0.35	1.48	0.35	1.48	7.06	3.50
2	0.32	5.14	0.32	5.14	6.18	6.18
3	4.10	6.67	2.53	6.67	4.06	4.06
4	7.78	6.80	5.59	6.67	1.93	1.93
5	7.78	0.25	7.78	5.14	1.06	1.06
6	4.10	0.34	7.78	1.48	1.93	1.93
7	3.54	4.42	5.59	0.34	4.06	4.06
8	4.63	4.43	2.53	0.34	6.18	6.18

Table 3: X and Y coordinates of the centres of the zones in both different locations.

	Location 1		Location 2	
Zone	x	y	x	y
Bright	3.57	3.85	2.00	3.85
Dark	4.63	3.85	6.00	3.85

As shown in Table 1 and depicted in Figure 2, three different source configurations and two zone locations were combined to define the six cases evaluated. Table 2 details the coordinates of the sources for each configuration and Table 3, the two different locations for both zones. These values correspond to the centre of the set of control points.

In all cases and for the sake of comparison, $N = 8$ sources and $M = 20$ control points per zone were used. These monitoring points were distributed in two 10-points arrays configured geometrically as shown in Figure 3. Inside each zone, one array was set at 1.05 m and the other at 1.15 m height, sampling spatially a $20 \times 30 \times 15$ cm volume at typical seated ear height. All sources were set at 0.2 m from the ground.

Table 4: Parameters and values used for evaluations, RIR simulations and control filter calculations.

Name	Value	Description
General		
c	340 [m/s]	Speed of sound
fs	1200 [Hz]	Sampling frequency
K	12000 [samples]	Input signal length
J	1000 [samples]	RIR length
I	100 [samples]	Control filters' length
δ_m	20 [samples]	Modelling delay
RIRs simulation*		
beta	0.4 [s]	Reverberation time
mtype	Omnidirectional	Microphone type
order	-1	Maximum reflection order
dim	3	Room dimensions
orientation	[0 0]	Microphone's direction
hp_filter	1	High-pass filter enabled
Control filters calculation		
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
ζ_l **	1×10^{12}	Pre-ringing envelope shaper
ζ_u **	1×10^{7}	Post-ringing envelope shaper

** Names according with [19]*
*** Name according with [16]*

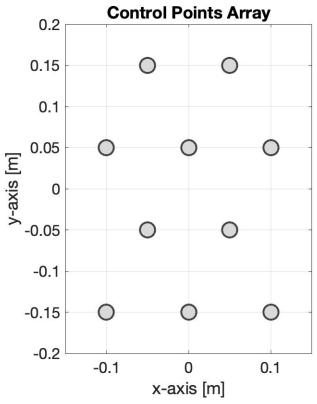

Figure 3: Geometric distribution of the control point arrays. Two arrays at different heights were simulated in each zone.

For each of the cases, RIRs of 1000 samples were calculated for both zones with the toolbox made by Habets, based on the image-source model [19]. Table 4 includes the parameters used for the RIR simulations. Given the frequency range of interest of this work, a sampling frequency of 1200 Hz was chosen. For this value, the 1000 samples of the RIRs correspond to a duration of 830 ms, more than twice as long as the average RT at the low frequencies. As seen in Figure 4, this acquisition time ensures a total decay of the reverberation within the measurement.

For each case, the simulation of the sound zones comprised the following steps:

(1) **RIRs truncation.** The RIRs used to determine the control filters were truncated, as seen in Figure 4, at 33, 74, 166, and 372 ms. This means that only the first samples were considered when calculating the filters: the samples after the above-mentioned lengths were removed. These truncation values were distributed logarithmically between twice the modelling delay δ_m and the complete duration of 830 ms, allowing to assess different moments in the RIR and therefore, to include different information. Steps 2-4 were performed for each set of truncated RIRs.

(2) **Control filters computation.** With each set of truncated RIRs, control filters $\mathbf{w}$ were calculated as stated in (7) using the same parameters, detailed in Table 4. In addition and for comparison, filters with the complete RIRs were also generated. It is worth to mention that, for every set of filters, the regularization factor λ was scaled by the norm of the cross-correlation matrix of RIRs in the bright zone $\| \mathbf{H}_b^T \mathbf{H}_b \|$. A 4th-order high-pass Butterworth filter was introduced in matrix $\mathbf{R}_F$, emphasizing the frequency-dependent regularization below 500 Hz. In addition, we introduce here $\mathbf{p}_{r_\tau}$, the reference pressure in (7) calculated using the corresponding RIRs truncated to length τ ms. Finally, despite the last two penalty terms in (7), no specific or intended effort was made to optimize the filter design with truncated RIRs.

(3) **Reference pressure calculation.** The reference pressure was defined to be the signal played in the bright zone by the 7th source, whose position changed depending on the case. This implies different distances between the bright zone and the reference source for all cases and therefore, different reference pressures. To take this into account, a modelling delay δ_m of 17 ms was added through a pure-delay filter [20]. This delay corresponds to the maximum propagation time of the furthest source to the bright zone, evaluated among all the cases. To compute the reference pressure on each case, the RIRs between the 7th source and the control points in the bright zone were convolved with the input signal. In this step, the complete, 830 ms RIRs were used. Finally, all the individual signals were averaged to find the reference pressure or reference sound field $\mathbf{p}_r$.

(4) **Evaluation pressures calculation.** The evaluation pressures at the control points in both zones were calculated as the sum of the convolutions of the input signal, the N control filters and the respective RIRs. For point m in the bright zone, for example:

$$\mathbf{p}_{\text{eval\_}b_m} = \sum_{n=1}^{N} \mathbf{h}_{b_{mn}} * \mathbf{w}_{\tau,n} * \mathbf{s}_w \,, \tag{8}$$

where $\mathbf{h}_{b_{mn}}$ is the RIR from source n to control point m, $\mathbf{s}_w$ is the white noise sequence and $\mathbf{w}_{\tau,n}$ is the FIR control filter calculated for the source n, based on RIRs truncated to length τ ms. Since the lengths of $\mathbf{s}_w$, $\mathbf{h}_{b_{mn}}$ and $\mathbf{w}_{\tau,n}$ are K,

J and I respectively, the resulting vector has a total length of $(K + J + I - 2)$. In this step, both $\mathbf{p}_{\text{eval\_}b_m}$ and $\mathbf{p}_{\text{eval\_}d_m}$ pressures were calculated for each of the five different durations of RIRs, for all the six cases. Notice again that, while the filters $\mathbf{w}_{\tau,n}$ used truncated RIRs, pressures $\mathbf{p}_{\text{eval\_}b_m}$, $\mathbf{p}_{\text{eval\_}d_m}$ and $\mathbf{p}_r$ were found with the complete RIRs $\mathbf{h}_{b_{mn}}$. Otherwise, the propagation inside the room and its contribution to the sound would not have been completely and correctly taken into account.

(5) **Objective metrics computation.** Based on the previous steps, the average ACR and NMSE were found for each RIR duration in each case, computed only along the K-samples of signal $\mathbf{s}_w$ and for the M control points. Since these quantities are equal for both zones, the averages can be expressed as:

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

$$\overline{NMSE} = 10\log_{10}\left[\frac{\sum_{k=1}^{K}\sum_{m=1}^{M}\|\mathbf{p}_r[k]-\mathbf{p}_{\text{eval\_}b_m}[k]\|^2}{\sum_{k=1}^{K}\sum_{m=1}^{M}\|\mathbf{p}_r[k]\|^2}\right]. \tag{10}$$

Notice that the tails of the convolutions in (8) were excluded from the calculations of the ACR and NSME. In addition, the latter was assessed using the reference pressure $\mathbf{p}_r$, obtained in step **3** with the complete RIRs. To evaluate the sound field inside the zones per case, the mean square pressures were found and compared for each case. This calculation included both spatial and temporal averages. Exemplifying once again for the bright zone:

$$\mathbf{p}_{b_{MS}} = \frac{1}{KM}\sum_{k=1}^{K}\sum_{m=1}^{M}\|\mathbf{p}_{\text{eval\_}b_m}[k]\|^2. \tag{11}$$

Finally, the overall ACR, NMSE and pressure averages between cases were also assessed, giving a general idea of the performance of the zones with respect to the duration of the RIRs.

4 RESULTS

Figure 5 shows the ACR in dB for each RIRs duration, for all six cases evaluated, and their overall average. Figure 6 shows the same analysis performed for the NSME.

In general, it can be seen that short RIR durations yield lower acoustic contrasts and higher errors, as one might expect. Increasing the truncation length from 33 ms to 74 ms and from 74 ms to 166 ms, increases the ACR around 6 dB and 4 dB respectively. For truncation lengths longer than 166ms, this value tends to stabilize close to 20 dB and only small changes are observed. Conversely, the NMSE presents no considerable changes for truncation lengths longer than 74ms and in some cases, it even worsens slightly. Hence, even very short RIRs approach the lowest error achievable.

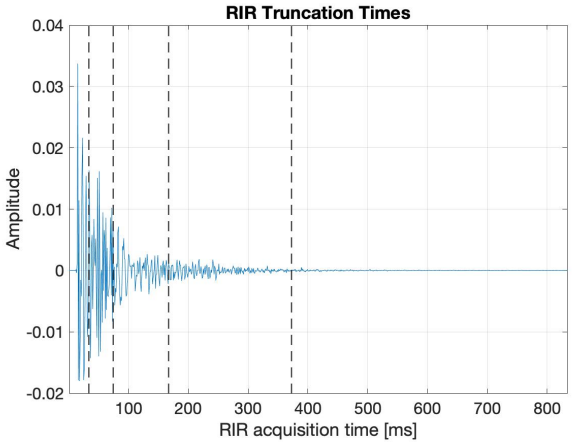

Figure 4: Truncation times of RIRs simulated for the estimation of the control filters. These times, shown by the vertical dashed lines, correspond to 33, 74, 166 and 372 milliseconds.

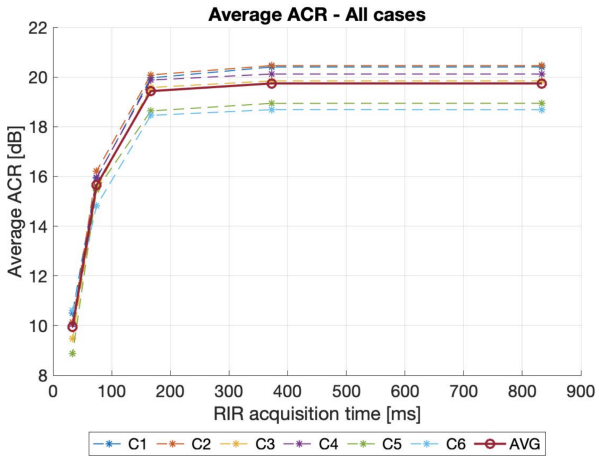

Figure 5: ACR evaluated for the six cases (C1-C6) for different RIR acquisition times and their average (AVG).

From the previous analysis, it can be said that after a certain acquisition time, sound zones created with longer RIRs behave almost the same. In other words, there is a limit where taking more information into account presents no big improvement in the performance. For the investigated scenarios and due to the ACR, such duration can be said to be around 160 ms. This truncation occurs at a point where the level of the RIRs has decreased almost 30 dB and provides little additional information.

The previous analyses assessed the performance of the sound zones in terms of the mean square pressure in the bright zone with respect to the dark one and, with respect to the reference. However, these give no information about the behaviour of each zone individually. In order to assess this, the spatial and temporal averaged square pressure of both zones for each truncation time was evaluated and plotted in Figures 7 and 8. These also show the overall average for all six cases. Notice how the level decreases

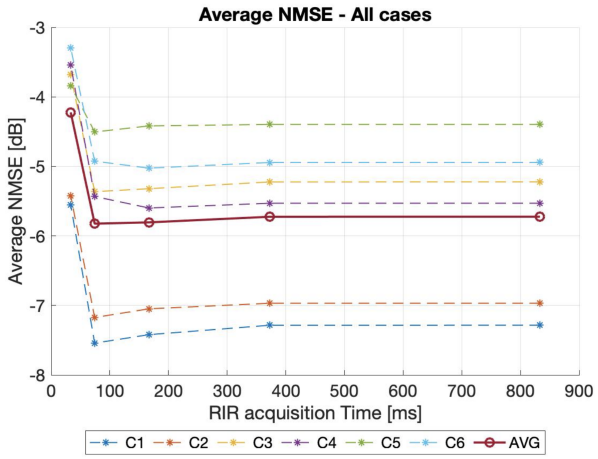

Figure 6: Normalized Mean Squared Error evaluated for the six cases (C1-C6) for different RIR acquisition times and their average (AVG).

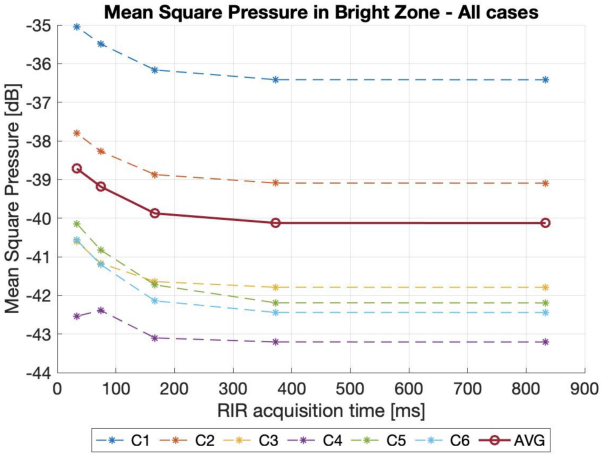

Figure 7: Frequency average of the mean square pressure in the bright zone, evaluated for the six cases (C1-C6) for different RIR acquisition times and their average between cases (AVG).

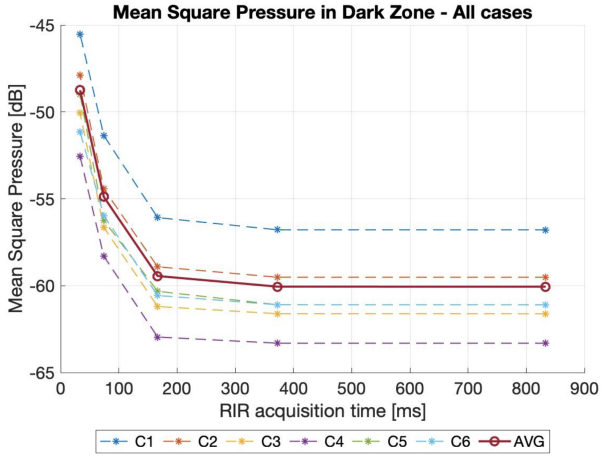

Figure 8: Frequency average of the mean square pressure in the dark zone, evaluated for the six cases (C1-C6) for different RIR acquisition times and their average between cases (AVG).

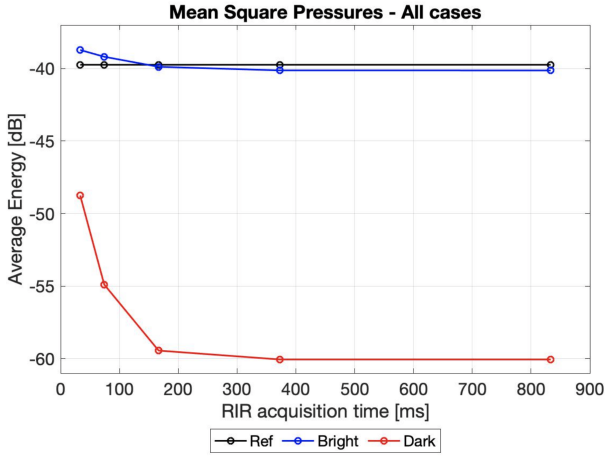

Figure 9: Comparison of the frequency averages of the mean square pressure in the bright and dark zones, and the reference pressure.

in both zones for longer RIRs but, while the average pressure in the bright zone decreases around 1.2 dB when increasing the RIRs duration, in the dark zone it decreases almost 12 dB. In fact, it can be said that the energy in the dark zone almost decreases with the ACR increase as the truncation time increases.

Such difference in the variations between zones can be better observed in Figure 9. It can be seen that shortening the RIRs have a considerably greater impact in the dark zone than in the bright one, and that the improvement in the ACR relies mostly on the energy reduction in the dark zone. In addition, as also observed from the previous analysis, the levels inside the sound zones remain almost constant after a given RIR duration, namely around 160 ms for this case. Finally, notice that all cases follow the same trend on each metric and pressure assessment but, their relative levels differ among them. This is mainly because the reference pressures are also different between cases which, as explained before, is due to different distances from the reference source and the bright zone.

5 CONCLUSIONS

The effect of the acquisition time of RIRs on the performance of sound zones was assessed by simulating six different configurations in the same room. Such performance was evaluated in terms of the Acoustic Contrast Ratio (ACR), the Normalized Mean Square Error (NSME) and the mean square pressures. It was observed that up to a certain acquisition time, the performance can be improved by using longer RIRs. After this time, improvements on the ACR and NSME are negligible. Linked to this, it was seen that the lowest

NMSE possible in a given setup can be achieved with very short RIRs. Moreover, analyses of the mean square pressures showed also that using truncated RIRs to design the control filters have a remarkable greater impact on the mean square pressure inside the dark zone than inside the bright one.

Therefore, it can be said that using more than a certain amount of information of the RIRs to design control filters for sound zones, can be unnecessary. This is because, depending on the application, the performance obtained with less information may be satisfactory enough. Nevertheless, such minimum acquisition time found in the simulated scenarios is still very large for real-time applications, especially for achieving proper sound control in the dark zone.

This work constitutes also a suitable basis for further tests and analyses of the influence of RIR characteristics on the performance of sound zones. These experiments may include using truncated RIRs of multiple rooms with higher reverberation times, assessing the impact of noise and other RIR inaccuracies, and evaluating real systems for sound zones creation, both under objective metrics and with subjective listening tests.

ACKNOWLEDGMENTS

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. 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.

REFERENCES

[1] Philip Coleman, Philip J. B. Jackson, Marek Olik, and Jan Abildgaard Pedersen. 2014. Personal audio with a planar bright zone. *J. Acoust. Soc. Amer.*, 136, 4, (October 2014), 1725–1735.

[2] WF Druyvesteyn and John Garas. 1997. Personal sound. *J. Audio Eng. Soc.*, 45, 9, 685–701.

[3] Terence Betlehem, Wen Zhang, Mark A. Poletti, and Thushara D. Abhayapala. 2015. Personal Sound Zones: Delivering interface-free audio to multiple listeners. *IEEE Signal Process. Mag.*, 32, 2, (March 2015), 81–91.

[4] Martin Bo Møller and Martin Olsen. 2017. Sound zones: on the effect of ambient temperature variations in feed-forward systems. In *Proc. AES 142th Convention*. Audio Engineering Society, 1009–1018.

[5] Xiaohui Ma, Patrick J Hegarty, Jan Abildgaard Pedersen, and Jakob Juul Larsen. 2018. Impact of loudspeaker nonlinear distortion on personal sound zones. *J. Acoust. Soc. Amer.*, 143, 1, 51–59.

[6] Martin Bo Møller, Jesper Kjær Nielsen, Efren Fernandez-Grande, and Søren Krarup Olesen. 2019. 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.

[7] Jin-Young Park, Jung-Woo Choi, and Yang-Hann Kim. 2013. Acoustic contrast sensitivity to transfer function errors in the design of a personal audio system. *J. Acoust. Soc. Amer.*, 134, 1, EL112–EL118.

[8] Christian Sejer Pedersen, Martin Bo Møller, and Jan Østergaard. 2022. Effect of wireless transmission errors on sound zone performance at low frequencies. In *EUROREGIO BNAM 2022 Joint Acoustics Conference*, 115–124.

[9] Gema Piñero, Carmen Botella, Maria de Diego, Miguel Ferrer, and Alberto González. 2017. On the feasibility of personal audio systems over a network of distributed loudspeakers. In *Proc. 25th European Signal Process. Conf. (EUSIPCO '17)*, 2729–2733.

[10] Peter Koch and Jan Østergaard. 2022. The effect of fixed-point arithmetic on low frequency sound-zone control. In *EUROREGIO BNAM 2022 Joint Acoustics Conference*, 125–134.

[11] Marcos F. Simón Gálvez, Stephen J Elliott, and Jordan Cheer. 2015. Time domain optimization of filters used in a loudspeaker array for personal audio. *IEEE/ACM Trans. Audio Speech Lang. Process.*, 23, 11, 1869–1878.

[12] Joung-Woo Choi and Yang-Hann Kim. 2002. Generation of an acoustically bright zone with an illuminated region using multiple sources. *J. Acoust. Soc. Amer.*, 111, 4, (April 2002), 1695–1700.

[13] Ole Kirkeby and Philip A Nelson. 1993. Reproduction of plane wave sound fields. *J. Acoust. Soc. Amer.*, 94, 5, 2992–3000.

[14] Ji-Ho Chang and Finn Jacobsen. 2013. Experimental validation of sound field control with a circular double-layer array of loudspeakers. *J. Acoust. Soc. Amer.*, 133, 4, 2046–2054.

[15] Ole Kirkeby and Philip A Nelson. 1999. Digital filter design for inversion problems in sound reproduction. *J. Audio Eng. Soc.*, 47, 7/8, 583–595.

[16] Martin Bo Møller and Martin Olsen. 2017. Sound zones: on envelope shaping of fir filters. In *Proc. 24th Int. Congr. on Sound and Vibration 2017 (ICSV 24)*. (July 2017), 613–620.

[17] Martin Bo Møller and Jan Østergaard. 2020. A Moving Horizon Framework for Sound Zones. *IEEE/ACM Trans. Audio Speech Lang. Process.*, 28, 256–265.

[18] Miklas Strøm Kristoffersen, Martin Bo Møller, Pablo Martínez-Nuevo, and Jan Østergaard. 2021. Deep Sound Field Reconstruction in Real Rooms: Introducing the ISOBEL Sound Field Dataset, (February 2021). Retrieved 01/26/2022 from http://arxiv.org/abs/2102.06455.

[19] Emanuël A.P. Habets. 2020. Room impulse response generator. (October 2020). Retrieved 11/26/2021 from https://www.audiolabs-erlangen.de/fau/professor/habets/software/rir-generator.

[20] Hironori Tokuno, Ole Kirkeby, Philip A Nelson, and Hareo Hamada. 1997. Inverse filter of sound reproduction systems using regularization. *IEICE Trans. Fundamentals*, 80, 5, 809–820.