A Moving Horizon Framework for Sound Zones

Møller, Martin Bo; Østergaard, Jan

Published in:
IEEE/ACM Transactions on Audio, Speech, and Language Processing

DOI (link to publication from Publisher):
10.1109/TASLP.2019.2951995

Publication date:
2020

Document Version
Accepted author manuscript, peer reviewed version

A Moving Horizon Framework for Sound Zones

Martin B. Møller, *Student Member, IEEE,* and Jan Østergaard, *Senior Member, IEEE*

***Abstract*—Sound zones are generated to provide independent audio reproduction to multiple people in the same room using loudspeakers. In this paper, sound zones are formulated in terms of a moving horizon framework. This framework allows the reproduction scenario to be time-varying and adapt to changes e.g. in the location of the zones or in the audio signal. The framework is tested using both simulated and measured room impulse responses from eight loudspeakers in a rectangular room. The performance is investigated using signals limited between 35 - 500 Hz, but the framework is not limited to a particular frequency range. The experimental results show that it is possible to gain on the order of 4 dB higher separation between the zones using the proposed framework, relative to a conventional time-invariant solution. This gain arises from knowledge about the audio content currently being reproduced in the zones, and it is obtained without deteriorating the reproduction accuracy or increasing the signal energy injected into the loudspeakers.**

***Index Terms*—Sound field control, sound zones, adaptive control, moving horizon control.**

I. Introduction

THE term *sound zones* refers to applying a set of loudspeakers to reproduce individual audio signals in separate spatial regions. The potential of the technology has been investigated in airplane and car cabins [1]–[6], for reducing low frequency noise from open air concerts [7], and in domestic rooms [8]–[10].

The simplest sound zone scenario consists of two separate spatial regions where different audio content is desired. The audio content A is desired in zone A and undesired in zone B, and vice versa. Content A should, thus, be reproduced in zone A while it should be suppressed as much as possible in zone B. Zone A is commonly referred to as the bright zone with respect to content A, while zone B is the dark zone. Reproduction of individual audio in the zones is achieved by linear superposition of solutions creating one bright and one dark zone for each audio signal.

A general approach to sound zones, is to use different loudspeaker layouts in different parts of the audible frequency range [8]. In this paper, a control method is introduced and demonstrated for low frequency application.[1] It is assumed that higher frequencies are controlled using e.g. beamforming as suggested in [8], [10], [12]–[15]. To control the sound field at low frequencies, it is necessary to know how each loudspeaker radiates sound to the bright and dark zone. Typically, this information is acquired by measuring the impulse responses from each loudspeaker to a number of microphones in each zone. The microphones are then removed, and control filters based on the measurements are implemented as a feed-forward control system. Often, the filters are time-invariant and are based on the implicit assumption that the audio signal is spectrally flat [13], [14]. Notable exceptions utilize the short-time Fourier transform to formulate sound zones as block-based processing [16]. A recent publication has shown the sound zones problem to be equivalent to the speech enhancement problem [17]. Thus, methods from the speech enhancement literature can be applied to sound zones by incorporating the time-varying cross-correlation of the audio signals to be reproduced in the zones [18], [19].

With time-invariant filters, a general concern is the time dependent variation of the measured room impulse responses (RIRs) after the microphones are removed. Typical degradation of the RIR accuracy include changes in the ambient temperature [6], [20], changes in sound zone location [21], [22], or changes in the electro-dynamic loudspeaker systems [23]. One way to reduce the sensitivity to the temperature and loudspeaker changes is by predicting the effect of the changes and designing the filters to be robust to such changes [24]–[27]. Likewise, the sensitivity to changes in the position could be reduced by increasing the size of the zones. The price paid for such robustness, is lower performance compared to accurate system knowledge.

If high separation between the zones is desired, accurate knowledge about the system state and the RIRs is required. Furthermore, if the state and the RIRs change over time, it is necessary to have a control system which can adapt the solution to such changes. In the literature, moving horizon control (MH), also referred to as model predictive control and receding horizon control, has been introduced as an effective framework for controlling dynamic systems [28]. This control framework enables adaptation to time-varying changes by predicting future states of the system given current control actions. In other audio-related fields, MH control has been used to improve audio transmission over networks in terms of the rate-distortion [29].

In the present paper, the MH control framework is adapted for a sound zones application. An application enabled by the framework is the creation of sound zones which follow listeners in a room, assuming the corresponding RIRs are known. While such a scenario is also possible using time-invariant filters, it would require pre-computation of all combinations of desired zone positions and sizes. Therefore, the MH framework adds flexibility to reconfigure the zones in response to time-varying reproduction scenarios. Latency is generally a

M. B. Møller is with the research department of Bang & Olufsen A/S, Struer, DK-7600, Denmark and the Signal and Information Processing Section, Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark. (e-mail: mim@bang-olufsen.dk).

J. Østergaard is with the Signal and Information Processing Section, Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark e-mail: (jo@es.aau.dk).

Manuscript received April 19, 2005; revised August 26, 2015.

[1]The work presented in this paper is introduced in the context of low frequency sound field control, but it could equivalently be translated to higher frequencies with appropriate definitions of bright and dark zones, as evident by comparing e.g. [11] to [12].

concern in audio applications. To keep the latency minimal, the proposed framework is only given information about changes in the RIRs and the input signal in the immediate future.[2] The framework is formulated in state-space form which allows for extensions of control objectives which can also be formulated in state-space form e.g. perceptual motivated objectives as desired in [16], [19], [31].

The performance of the moving horizon framework is investigated using simulated and measured room impulse responses. The experimental results are compared to the performance of a conventional time-invariant solution, using RIRs measured in a room with $T_{20} = 0.6$ s reverberation time, averaged over the $1/3$-octave bands with center frequencies between 31.5 and 315 Hz.

The remainder of the article is structured as follows. In section II, the general problem is introduced. The moving horizon framework is introduced in section III by modelling the audio reproduction in state-space form. The behavior of moving horizon control is investigated in a simulated environment in section IV. An experimental comparison between moving horizon and a conventional time-invariant solution is presented in section V, before the results are briefly discussed in section VI, and conclusions are presented in section VII.

II. Formulating the Problem

A. Notation

Throughout this paper, scalars are denoted by regular lowercase symbols, a. Vectors are denoted by lower case bold types $\mathbf{a}$, matrices are denoted by upper case bold types $\mathbf{A}$, and $\cdot^T$ denotes the regular matrix transpose. The double-sided sequence $\{a[-\infty], \ldots, a[k-1], a[k], a[k+1], \ldots, a[\infty]\}$ is denoted $\{a[k]\}$. Subscript k is used to denote time-dependence, $\mathbf{a}_k$. Arrows above quantities are used to denote whether a quantity is projected into the future or the past, relative to the subscripted time-index e.g. $\overleftarrow{\mathbf{a}}_k = [a[k], a[k-1], \ldots, a[k-n+1]]^T$ if $\overleftarrow{\mathbf{a}}_k \in \mathbb{R}^n$ and $\overrightarrow{\mathbf{a}}_k = [a[k], a[k+1], \ldots, a[k+n-1]]^T$ if $\overrightarrow{\mathbf{a}}_k \in \mathbb{R}^n$.

B. Problem Formulation

The problem treated in this paper is controlling the sound field inside two regions in space. The purpose is to reproduce a desired audio signal in the bright zone and suppress it in the dark zone. The sound field is controlled using L loudspeakers and a conceptual setup is illustrated in Fig. 1.

To control the sound field within the regions, the pressure is assumed known at M_b and M_d points in the bright and dark zone respectively. The pressure at the m^{th} point is the combined output from each of the L loudspeakers convolved with the room impulse responses. If the audio signal $\{u[k]\}$, is prefiltered by a finite impulse response (FIR) filter for each loudspeaker, the pressure at point m can be expressed as

$$p^{(m)}[k] = \sum_{\ell=1}^{L} (\mathbf{h}_k^{(\ell,m)} * \mathbf{w}_k^{(\ell)} * \mathbf{u})[k] \quad (1)$$

$$= \sum_{\ell=1}^{L} \sum_{i=0}^{n-1} \sum_{j=0}^{n_w-1} h_k^{(\ell,m)}[i] w_k^{(\ell)}[j] u[k-i-j]. \quad (2)$$

In the above, $\mathbf{h}_k^{(\ell,m)} = [h_k^{(\ell,m)}[0], h_k^{(\ell,m)}[1], \cdots, h_k^{(\ell,m)}[n-1]]^T$ is the RIR between loudspeaker ℓ and the point m, $\mathbf{w}_k^{(\ell)} = [w_k^{(\ell)}[0], w_k^{(\ell)}[1], \cdots, w_k^{(\ell)}[n_w-1]]^T$ is the control filter for the ℓ^{th} loudspeaker, and m is a point in either the bright or the dark zone. The subscript k indicates that both RIRs and control filters can change with time.[3]

Besides suppressing the reproduced sound in the dark zone, it is also of interest to control the reproduced sound within the bright zone. To accommodate the effect of the room, the target response in the bright zone could be determined as the room equalized response of the loudspeaker system [32]. In this paper, the target response is the delayed response of a single loudspeaker at the control points, which at a single point is expressed as

$$p_t^{(m)}[k] = (\mathbf{h}_k^{(t,m)} * \mathbf{w}_k^{(\text{eq})} * \mathbf{u})[k] \quad (3)$$

where $\mathbf{w}_k^{(eq)}$ is the room equalization filter (in this case an integer sample delay). The signal flow from the input audio signal to a point m_b in the bright zone and m_d in the dark zone is illustrated in Fig. 2.

The purpose of the control is to accurately reproduce the target audio signal in the bright zone, while suppressing it in the dark zone. This can be expressed as the simplified optimization problem,

$$\min_{\mathbf{w}_k} \sum_{m_b=1}^{M_B} \|p_t^{(m_b)}[k] - p^{(m_b)}[k]\|^2 + \sum_{m_d=1}^{M_D} \|p^{(m_d)}[k]\|^2, \quad (4)$$

where $\mathbf{w}_k = [\mathbf{w}_k^{(1)T} \cdots \mathbf{w}_k^{(L)T}]^T$ are the concatenated control filters. This optimization problem is extended in section III-A

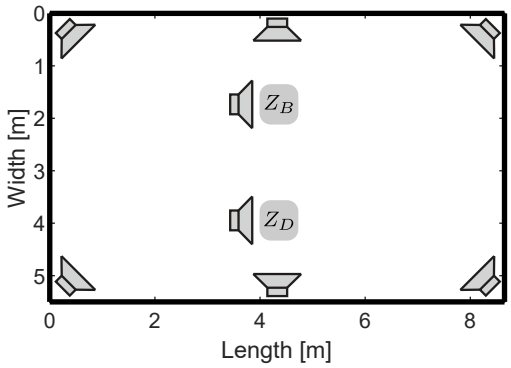


Fig. 1. Sketch of the experimental scenario. Here eight loudspeakers are shown, to control the sound field in the bright zone, Z_B, and the dark zone, Z_D.

[2]In the present paper, the RIRs are assumed known throughout the room. They could be measured in situ to all potential points of interest as a setup step, before the system is used to reproduce audio. Alternatively, a reduced number of measurements could be combined with methods for extrapolating the RIRs to the remaining points as suggested in e.g. [30]. During reproduction, the relevant RIRs could be retrieved from a database and used to update the control filters corresponding to the desired size and location of the sound zones.

[3]The room impulse responses are assumed known. They could be measured in situ to all potential points of interest with microphones which are removed prior to reproducing audio.

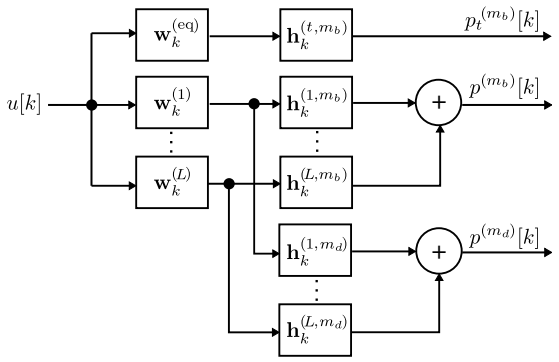


Fig. 2. Schematic of the signal flow from the input $u[k]$ to the target pressure in the bright zone $p_t^{(m_b)}[k]$, the reproduced bright zone pressure $p_b^{(m_b)}[k]$ and the reproduced dark zone pressure $p_d^{(m_d)}[k]$.

to include additional terms. Note that the solution is updated at each time index k and depends on the input signal and the RIRs representing the bright and dark zone.[4] The solution at time index k depends on sound previously emitted by the loudspeakers. By predicting the resulting pressure at future time-steps, the dependence on previous solutions can be incorporated in the optimization. This is the foundation of the moving horizon framework presented in the following.

III. Moving Horizon Framework

The following expressions describe the sound field in the bright zone. Equivalent expressions for the sound field in the dark zone can be obtained by substituting the RIRs.

The system in (1) can be written in state-space form, which is a convenient representation for time-varying, dynamic systems [33]. In order to do so, $[h_k^{(\ell,m_b)}[1],\ldots,h_k^{(\ell,m_b)}[n-1]]^T$ is introduced as a FIR filter in the z-domain $\tilde{H}_k^{(\ell,m_b)}(\mathrm{z})$ of the following form:

$$\tilde{H}_k^{(\ell,m_b)}(\mathrm{z}) = \sum_{i=1}^{n'} h_k^{(\ell,m_b)}[i]\mathrm{z}^{-i}, \tag{5}$$

where $n' = n-1$, $\mathrm{z} \in \mathbb{C}$, and where the first coefficient $h_k^{(\ell,m)}[0]$ of the RIR is omitted. The minimal state-space representation for $\tilde{H}_k^{(\ell,m_b)}(\mathrm{z})$ is [33][5]

$$\tilde{H}_k^{(\ell,m_b)}(\mathrm{z}) = \mathbf{h}_k^{'(\ell,m_b)T}(\mathrm{z}\mathbf{I}-\mathbf{A})^{-1}\mathbf{b}, \tag{6}$$

where $\mathbf{A} \in \mathbb{R}^{n'\times n'}$, $\mathbf{b} \in \mathbb{R}^{n'\times 1}$, and $\mathbf{h}_k^{'(\ell,m_b)} \in \mathbb{R}^{n'}$ are given by

$$\mathbf{A} = \begin{bmatrix} 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{7}$$

$$\mathbf{h}_k^{'(\ell,m_b)} = \begin{bmatrix} h_k^{(\ell,m_b)}[1] & \ldots & h_k^{(\ell,m_b)}[n'] \end{bmatrix}^T. \tag{8}$$

[4]In the case of moving sound zones, the set of RIRs can be updated between time steps to reflect the updated location of the zones.

[5]In (6) z^{-1} denotes the delay operator, see [33] for details.

It follows that

$$h_k^{(\ell,m_b)}[i] = \mathbf{h}_k^{'(\ell,m_b)T}\mathbf{A}^i\mathbf{b}, \quad i = 1,2,\ldots,n', \tag{9}$$

where $\mathbf{A}^i$ denotes i matrix multiplications with $\mathbf{A}$.

We can now express the sound pressure (1) in state-space form

$$p^{(\ell,m_b)}[k] = \mathbf{h}_k^{'(\ell,m_b)T}\overleftarrow{\mathbf{z}}_k^{(\ell)} + h_k^{(\ell,m_b)}[0]y^{(\ell)}[k] \tag{10}$$

where $y^{(\ell)}[k] = (\mathbf{w}_k^{(\ell)} * \overleftarrow{\mathbf{u}}_k)[k]$ is the filtered input signal to the ℓ^{th} loudspeaker and $\overleftarrow{\mathbf{z}}_k^{(\ell)} \in \mathbb{R}^{n'}$ defines the current state vector given by

$$\overleftarrow{\mathbf{z}}_k^{(\ell)} = [y^{(\ell)}[k-1],\ldots,y^{(\ell)}[k-n']]^T, \tag{11}$$

$$\overleftarrow{\mathbf{z}}_{k+1}^{(\ell)} = \mathbf{A}\overleftarrow{\mathbf{z}}_k^{(\ell)} + \mathbf{b}y^{(\ell)}[k]. \tag{12}$$

If the future RIRs and input samples are known or predicted, the future pressure samples can be predicted by applying the recursive relationships (10) and (12)

$$p^{(\ell,m_b)}[k] = \mathbf{h}_k^{'(\ell,m_b)T}\overleftarrow{\mathbf{z}}_k^{(\ell)} + h_k^{(\ell,m_b)}[0]y^{(\ell)}[k] \tag{13}$$

$$p^{(\ell,m_b)}[k+1] = \mathbf{h}_{k+1}^{'(\ell,m_b)T}\overleftarrow{\mathbf{z}}_{k+1}^{(\ell)} + h_{k+1}^{(\ell,m_b)}[0]y^{(\ell)}[k+1] \tag{14}$$

$$= \mathbf{h}_{k+1}^{'(\ell,m_b)T}(\mathbf{A}\overleftarrow{\mathbf{z}}_k^{(\ell)} + \mathbf{b}y^{(\ell)}[k]) + h_{k+1}^{(\ell,m_b)}[0]y^{(\ell)}[k+1] \tag{15}$$

$$p^{(\ell,m_b)}[k+2] = \mathbf{h}_{k+2}^{'(\ell,m_b)T}(\mathbf{A}(\mathbf{A}\overleftarrow{\mathbf{z}}_k^{(\ell)} + \mathbf{b}y^{(\ell)}[k]) + \mathbf{b}y^{(\ell)}[k+1]) + h_{k+2}^{(\ell,m_b)}[0]y^{(\ell)}[k+2] \tag{16}$$

$$= \mathbf{h}_{k+2}^{'(\ell,m_b)T}\mathbf{A}^2\overleftarrow{\mathbf{z}}_k^{(\ell)} + \mathbf{h}_{k+2}^{(\ell,m_b)T}\mathbf{A}\mathbf{b}y^{(\ell)}[k] + \mathbf{h}_{k+2}^{'(\ell,m_b)T}\mathbf{b}y^{(\ell)}[k+1] + h_{k+2}^{(\ell,m_b)}[0]y^{(\ell)}[k+2] \tag{17}$$

$$= \mathbf{h}_{k+2}^{'(\ell,m_b)T}\mathbf{A}^2\overleftarrow{\mathbf{z}}_k^{(\ell)} + h_{k+2}^{(\ell,m_b)}[2]y^{(\ell)}[k] + h_{k+2}^{(\ell,m_b)}[1]y^{(\ell)}[k+1] + h_{k+2}^{(\ell,m_b)}[0]y^{(\ell)}[k+2] \tag{18}$$

$$\vdots$$

$$p^{(\ell,m_b)}[k+I_p-1] = \mathbf{h}_{k+I_p-1}^{'(\ell,m_b)T}\mathbf{A}^{I_p-1}\overleftarrow{\mathbf{z}}_k^{(\ell)} + h_{k+I_p-1}^{(\ell,m_b)}y^{(\ell)}[k] + \cdots + h_{k+I_p-1}^{(\ell,m_b)}[0]y^{(\ell)}[k+I_p-1]. \tag{19}$$

In the following, the future time horizon is divided in two parts: A control horizon of length I_c and a prediction horizon extending from the control horizon to a total length of I_p. The purpose of the division is to react to changes in the immediate future but predict the consequences of those changes further ahead to ensure stable control [28].[6] This leads to the following expression for the loudspeaker inputs at future time steps,

$$y^{(\ell)}[k+i] = \begin{cases} \overleftarrow{\mathbf{u}}_{k+i}^T\mathbf{w}_{k+i}^{(\ell)}, & i \in [0,1,\cdots,I_c-1] \\ \overleftarrow{\mathbf{u}}_{k+i}^T\mathbf{w}_{k+I_c-1}^{(\ell)}, & i \in [I_c,I_c+1,\cdots,I_p-1]. \end{cases} \tag{20}$$

With this definition of the inputs, it is seen that the future tentative control filters only change within the control horizon

[6]This division into control and prediction horizon is a key point. It enables systems with latency equal to the short control horizon, where the consequences of the control are predicted further into the future. Furthermore, the number of filter coefficients determined at each time-step is proportional to I_c rather than I_p, where $I_c \leq I_p$.

and are kept fixed at the end of the control horizon for the remainder of the prediction horizon.[7]

Since $y^{(\ell)}[k] = \overleftarrow{\mathbf{u}}_k^T \mathbf{w}_k^{(\ell)} = \mathbf{w}_k^{(\ell)T} \overleftarrow{\mathbf{u}}_k$, we can rewrite the recursions (13) - (19) as

$$\vec{\mathbf{p}}_k^{(\ell,m_b)} = \begin{bmatrix} p^{(\ell,m_b)}[k] \\ p^{(\ell,m_b)}[k+1] \\ \vdots \\ p^{(\ell,m_b)}[k+I_p-1] \end{bmatrix} \tag{21}$$

$$= \Psi(\vec{\mathbf{h}}_k^{(\ell,m_b)}) \vec{\mathbf{U}}_k \vec{\mathbf{w}}_k^{(\ell)} + \Gamma(\vec{\mathbf{h}}_k^{'(\ell,m_b)}) \overleftarrow{\mathbf{z}}_k^{(\ell)}, \tag{22}$$

where $\vec{\mathbf{p}}_k^{(\ell,m_b)} \in \mathbb{R}^{I_p}$ and

$$\Psi(\vec{\mathbf{h}}_k^{(\ell,m_b)}) = \tag{23}$$
$$\begin{bmatrix} h_k^{(\ell,m_b)}[0] & 0 & \cdots & 0 \\ h_{k+1}^{(\ell,m_b)}[1] & h_{k+1}^{(\ell,m_b)}[0] & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ h_{k+I_p-1}^{(\ell,m_b)}[I_p-1] & h_{k+I_p-1}^{(\ell,m_b)}[I_p-2] & \cdots & h_{k+I_p-1}^{(\ell,m_b)}[0] \end{bmatrix},$$

$$\vec{\mathbf{U}}_k = \begin{bmatrix} \overleftarrow{\mathbf{u}}_k^T & \mathbf{0}^T & \cdots & \mathbf{0}^T \\ \mathbf{0}^T & \overleftarrow{\mathbf{u}}_{k+1}^T & \cdots & \mathbf{0}^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}^T & \mathbf{0}^T & \cdots & \overleftarrow{\mathbf{u}}_{k+I_c-1}^T \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}^T & \mathbf{0}^T & \cdots & \overleftarrow{\mathbf{u}}_{k+I_p-1}^T \end{bmatrix}, \tag{24}$$

$$\vec{\mathbf{w}}_k^{(\ell)} = \begin{bmatrix} \mathbf{w}_k^{(\ell)} \\ \mathbf{w}_{k+1}^{(\ell)} \\ \vdots \\ \mathbf{w}_{k+I_c-1}^{(\ell)} \end{bmatrix}, \quad \Gamma(\vec{\mathbf{h}}_k^{'(\ell,m_b)}) = \begin{bmatrix} \mathbf{h}_k^{'(\ell,m_b)T} \\ \mathbf{h}_{k+1}^{'(\ell,m_b)T} \mathbf{A} \\ \vdots \\ \mathbf{h}_{k+I_p-1}^{'(\ell,m_b)T} \mathbf{A}^{I_p-1} \end{bmatrix}. \tag{25}$$

The dimensions of the above matrices are $\Psi(\vec{\mathbf{h}}_k^{(\ell,m_b)}) \in \mathbb{R}^{I_p \times I_p}$, $\vec{\mathbf{U}}_k \in \mathbb{R}^{I_p \times I_c n_w}$, $\vec{\mathbf{w}}_k^{\ell} \in \mathbb{R}^{I_c n_w}$, and $\Gamma(\vec{\mathbf{h}}_k^{'(\ell,m_b)}) \in \mathbb{R}^{I_p \times n'}$.

To include the contribution from all L loudspeakers at the m_b^{th} control point, the state-space form is written as

$$\vec{\mathbf{p}}_k^{(m_b)} = \bar{\Psi}(\vec{\mathbf{h}}_k^{(m_b)}) \bar{\mathbf{U}}_k \bar{\mathbf{w}}_k + \bar{\Gamma}(\vec{\mathbf{h}}_k^{'(m_b)}) \bar{\mathbf{z}}_k \in \mathbb{R}^{I_p}, \tag{26}$$

where

$$\bar{\Psi}(\vec{\mathbf{h}}_k^{(m_b)}) = \begin{bmatrix} \Psi(\vec{\mathbf{h}}_k^{(1,m_b)}) & \dots & \Psi(\vec{\mathbf{h}}_k^{(L,m_b)}) \end{bmatrix}, \tag{27}$$

$$\bar{\Gamma}(\vec{\mathbf{h}}_k^{'(m_b)}) = \begin{bmatrix} \Gamma(\mathbf{h}_k^{'(1,m_b)}) & \dots & \Gamma(\mathbf{h}_k^{'(L,m_b)}) \end{bmatrix}, \tag{28}$$

[7]At time-step k the tentative filters do not change after the end of the control horizon, $k + I_c - 1$. At time-step $k+1$ the tentative filters are fixed after the updated end of the control horizon, $k + I_c$.

$$\bar{\mathbf{w}}_k = \begin{bmatrix} \vec{\mathbf{w}}_k^{(1)} \\ \vec{\mathbf{w}}_k^{(2)} \\ \vdots \\ \vec{\mathbf{w}}_k^{(L)} \end{bmatrix}, \bar{\mathbf{U}}_k = \begin{bmatrix} \vec{\mathbf{U}}_k & & & \\ & \vec{\mathbf{U}}_k & & \\ & & \ddots & \\ & & & \vec{\mathbf{U}}_k \end{bmatrix}, \tag{29}$$

$$\bar{\mathbf{z}}_k = \begin{bmatrix} \overleftarrow{\mathbf{z}}_k^{(1)T} & \overleftarrow{\mathbf{z}}_k^{(2)T} & \dots & \overleftarrow{\mathbf{z}}_k^{(L)T} \end{bmatrix}^T, \tag{30}$$

with $\bar{\Psi}(\vec{\mathbf{h}}_k^{(m_b)}) \in \mathbb{R}^{I_p \times LI_p}$, $\bar{\mathbf{U}}_k \in \mathbb{R}^{LI_p \times LI_c n_w}$, $\bar{\mathbf{w}}_k \in \mathbb{R}^{LI_c n_w}$, $\bar{\Gamma}(\vec{\mathbf{h}}_k^{'(m_b)}) \in \mathbb{R}^{I_p \times Ln'}$, and $\bar{\mathbf{z}}_k \in \mathbb{R}^{Ln'}$. The notation $\bar{\cdot}$ is used to highlight the dependence on all L loudspeakers.

Finally, the sound pressure at all the M_b control points in the bright zone can be expressed as

$$\vec{\mathbf{p}}_{b,k} = \begin{bmatrix} \vec{\mathbf{p}}_k^{(1)T} & \dots & \vec{\mathbf{p}}_k^{(M_b)T} \end{bmatrix}^T \tag{31}$$

$$= \bar{\Psi}_b(\vec{\mathbf{h}}_k) \bar{\mathbf{U}}_k \bar{\mathbf{w}}_k + \bar{\Gamma}_b(\vec{\mathbf{h}}_k^{'}) \bar{\mathbf{z}}_k \in \mathbb{R}^{M_b I_p}, \tag{32}$$

where

$$\bar{\Psi}_b(\vec{\mathbf{h}}_k) = \begin{bmatrix} \bar{\Psi}(\vec{\mathbf{h}}_k^{(1)}) \\ \vdots \\ \bar{\Psi}(\vec{\mathbf{h}}_k^{(M_b)}) \end{bmatrix} \in \mathbb{R}^{M_b I_p \times LI_p}, \tag{33}$$

$$\bar{\Gamma}_b(\vec{\mathbf{h}}_k^{'}) = \begin{bmatrix} \bar{\Gamma}(\vec{\mathbf{h}}_k^{'(1)}) \\ \vdots \\ \bar{\Gamma}(\vec{\mathbf{h}}_k^{'(M_b)}) \end{bmatrix} \in \mathbb{R}^{M_b I_p \times Ln'}. \tag{34}$$

Here, the subscript b is utilized to denote the concatenation related to the responses at all M_b control points in the bright zone.

A. Optimization Problem

With the introduced state-space form, it is possible to express the optimization problem (4) using the moving horizon framework. As described in the introduction, the proposed framework can accommodate a wide variety of objectives if they can be described in state-space form. The effectiveness of the framework depends on the objectives describing aspects of the application which should be controlled. In the following, several objectives extending (4) are introduced before being combined into the cost-function used in this example. However, the framework can easily be modified to include other objectives according to application requirements.

1) Mean Square Pressure in the Dark Zone: The goal of the sound zones system is to reduce the leakage of sound, intended for the bright zone, into the dark zone. One way of characterizing this is the mean square pressure in the dark zone as used in [34]–[37]. The pressure at the control points in the dark zone can be defined similarly to (32) and the mean square pressure summed across the current and future time-steps up to the end of the prediction horizon becomes

$$f_d(\bar{\mathbf{w}}_k) = M_d^{-1} \|\vec{\mathbf{p}}_{d,k}\|_2^2 =$$
$$M_d^{-1} \|\bar{\Psi}_d(\vec{\mathbf{h}}_k) \bar{\mathbf{U}}_k \bar{\mathbf{w}}_k + \bar{\Gamma}_d(\vec{\mathbf{h}}_k^{'}) \bar{\mathbf{z}}_k\|_2^2. \tag{35}$$

2) Reproduction Error in the Bright Zone: To control the sound field in the bright zone, we have at least two options: We could use the negative mean square sound pressure [34] or we could use the ℓ_2-norm deviation from a target sound pressure as introduced in (4) and [35]–[39]. The advantage of expressing the bright zone in terms of the deviation from a target sound field, is that this term is convex (where a negative mean square pressure would be concave). Thus, by choosing the squared ℓ_2-norm deviation for controlling the bright zone, we preserve the convexity of the combined cost-function. The target sound pressure at the m_b^{th} point in the bright zone, for the current and future time-steps of the prediction horizon, can be determined as

$$\vec{\mathbf{p}}_k^{(t,m_b)} = \Psi(\vec{\mathbf{h}}_k^{(t,m_b)})\vec{\mathbf{U}}_k\vec{\mathbf{w}}_k^{(\text{eq})} + \Gamma(\vec{\mathbf{h}}_k^{\prime(t,m_b)})\vec{\mathbf{z}}_k^{(t)}, \quad (36)$$

with $\Psi(\vec{\mathbf{h}}_k^{(t,m_b)}) \in \mathbb{R}^{I_p \times I_p}$, $\vec{\mathbf{w}}_k^{(\text{eq})} \in \mathbb{R}^{I_c n_w}$, and $\Gamma(\vec{\mathbf{h}}_k^{\prime(t,m_b)}) \in \mathbb{R}^{(I_p \times n')}$. Furthermore, the target pressures at all control positions can be concatenated as

$$\vec{\mathbf{p}}_k^{(t)} = \Psi_b(\vec{\mathbf{h}}_k^{(t)})\vec{\mathbf{U}}_k\vec{\mathbf{w}}_k^{(\text{eq})} + \Gamma_b(\vec{\mathbf{h}}_k^{\prime(t)})\vec{\mathbf{z}}_k^{(t)}, \quad (37)$$

where $\Psi_b(\vec{\mathbf{h}}_k^{(t)}) \in \mathbb{R}^{M_b I_p \times I_p}$ and $\Gamma_b(\vec{\mathbf{h}}_k^{\prime(t)}) \in \mathbb{R}^{(M_b I_p \times n')}$. The mean square deviation from the target pressure in the bright zone can thus be expressed as

$$f_b(\bar{\mathbf{w}}_k) = \|\vec{\mathbf{p}}_{b,k} - \vec{\mathbf{p}}_k^{(t)}\|_2^2 \quad (38)$$

$$= \|\bar{\Psi}_b(\vec{\mathbf{h}}_k)\bar{\mathbf{U}}_k\bar{\mathbf{w}}_k + \bar{\Gamma}_b(\vec{\mathbf{h}}_k^{\prime})\bar{\mathbf{z}}_k - \vec{\mathbf{p}}_k^{(t)}\|_2^2. \quad (39)$$

In this work, the target sound field in the bright zone is the delayed response of a single loudspeaker at the control points.[8]

3) Filter Energy: Another parameter to control is the signal energy of the filters, expressed as the ℓ_2-norm of the vector of filter coefficients $\bar{\mathbf{w}}_k$. This should be constrained in order to ensure the signal level to be reproduced by the loudspeakers is not excessive. If that were the case, the loudspeaker would possibly distort, suffer mechanical damage or the amplifier would clip due to being unable to supply the requested electrical power to the loudspeaker. The typical metric, used in this scenario is the squared ℓ_2-norm of the loudspeaker filters for the current and future filters [13], [14], [24],

$$f_w(\bar{\mathbf{w}}_k) = \|\bar{\mathbf{w}}_k\|_2^2 = \bar{\mathbf{w}}_k^T\bar{\mathbf{w}}_k. \quad (40)$$

4) Filter Shape: In [40] it was shown that there are multiple time-invariant FIR filters which can provide similar results in terms of sound field control. Some of these filters might be preferred over others due to their associated pre- and post-ringing. To encourage a particular shape of the FIR filters, a cost-function term could be introduced to restrict the filters to start and end at zero. Such a term can be formulated as a weighted squared ℓ_2-norm with high weight at the beginning and end of the FIR filters (as proposed in [40]),

$$f_{\text{env}}(\bar{\mathbf{w}}_k) = \bar{\mathbf{w}}_k^T \operatorname{diag}(\mathbf{1}_{LI_c} \otimes \mathbf{w}_{\text{env}})\,\bar{\mathbf{w}}_k, \quad (41)$$

where $\mathbf{w}_{\text{env}}$ is the weights on the filter taps, $\otimes$ is the Kronecker product, and $\mathbf{1}_{LI_c}$ is the all-ones vector of length LI_c. The filter shape weighting functions used in this paper are summarized in Table I.

5) Frequency Weighting: Electrodynamic loudspeakers are inherently bandlimited in terms of the frequencies they can reproduce [41]. To reflect this in the cost-function, a penalty can be added to represent the energy reproduced outside of this range [14]. One way of introducing this penalty is a band-stop filter, with the stopband covering the active frequency range of the loudspeakers. To take the input signal into account, the band-stop filter is introduced after the loudspeaker control filters. The energy at these undesired frequencies can be predicted using the structure of (26) where the RIR is replaced by the band-stop FIR filter $\mathbf{s}_k$. The penalty can be expressed as

$$f_{\text{bs}}(\bar{\mathbf{w}}_k) = \|\bar{\Psi}(\vec{\mathbf{s}}_k)\bar{\mathbf{U}}_k\bar{\mathbf{w}}_k + \bar{\Gamma}(\vec{\mathbf{s}}_k^{\prime})\bar{\mathbf{z}}_k\|_2^2, \quad (42)$$

where $\mathbf{s}_k \in \mathbb{R}^{n_{\text{bs}}}$ is the band-stop filter and

$$\mathbf{s}_k^{\prime T} = \begin{bmatrix} s_k[1] & s_k[2] & \cdots & s_k[n_{\text{bs}}-1] \end{bmatrix} \in \mathbb{R}^{n_{\text{bs}}-1}. \quad (43)$$

The filter used in this work is a FIR approximation to a 4th order Butterworth band-stop filter with band-stop between 35 and 500 Hz.

6) Closed Form Solution: The combined cost-function can be written as a weighted sum of the different penalties

$$f_{\text{cost}}(\bar{\mathbf{w}}_k) = \alpha_d f_d(\bar{\mathbf{w}}_k) + \alpha_b f_b(\bar{\mathbf{w}}_k) + \alpha_w f_w(\bar{\mathbf{w}}_k) + \alpha_{\text{env}} f_{\text{env}}(\bar{\mathbf{w}}_k) + \alpha_{\text{bs}} f_{\text{bs}}(\bar{\mathbf{w}}_k) \quad (44)$$

where all the weighted terms are ℓ_2 norms and the αs are non-negative real numbers.[9] The filters minimizing (44) can for example be determined as the filters $\bar{\mathbf{w}}_k$ solving the system of normal equations

$$(\mathbf{Q}_d + \mathbf{Q}_b + \mathbf{Q}_w + \mathbf{Q}_{\text{env}} + \mathbf{Q}_{\text{bs}} + \mathbf{Q}_{\text{diff}})\,\bar{\mathbf{w}}_k = -(\mathbf{q}_d + \mathbf{q}_b + \mathbf{q}_{\text{bs}} + \mathbf{q}_{\text{diff}}), \quad (45)$$

where

$$\mathbf{Q}_d = \alpha_d\bar{\mathbf{U}}_k^T\bar{\Psi}_d(\vec{\mathbf{h}}_k)^T\bar{\Psi}_d(\vec{\mathbf{h}}_k)\bar{\mathbf{U}}_k \quad (46)$$

$$\mathbf{Q}_b = \alpha_t\bar{\mathbf{U}}_k^T\bar{\Psi}_b(\vec{\mathbf{h}}_k)^T\bar{\Psi}_b(\vec{\mathbf{h}}_k)\bar{\mathbf{U}}_k \quad (47)$$

$$\mathbf{Q}_w = \alpha_w\mathbf{I} \quad (48)$$

$$\mathbf{Q}_{\text{env}} = \alpha_{\text{env}} \operatorname{diag}(\mathbf{1}_{LI_c} \otimes \mathbf{w}_{\text{env}}) \quad (49)$$

$$\mathbf{Q}_{\text{bs}} = \alpha_{\text{bs}}\bar{\mathbf{U}}_k^T\bar{\Psi}(\vec{\mathbf{s}}_k)^T\bar{\Psi}(\vec{\mathbf{s}}_k)\bar{\mathbf{U}}_k \quad (50)$$

$$\mathbf{q}_d = \alpha_d\bar{\mathbf{U}}_k^T\bar{\Psi}_d(\vec{\mathbf{h}}_k)^T\bar{\Gamma}(\vec{\mathbf{h}}_k^{\prime})\bar{\mathbf{z}}_k \quad (51)$$

$$\mathbf{q}_b = \alpha_t\bar{\mathbf{U}}_k^T\bar{\Psi}_b(\vec{\mathbf{h}}_k)^T\left(\bar{\Gamma}(\vec{\mathbf{h}}_k^{\prime})\bar{\mathbf{z}}_k - \vec{\mathbf{p}}_{t,k}\right) \quad (52)$$

[8]Note that this target pressure is a design choice and other choices could be made. This target inherently includes the spatial variation of the sound field emitted by the chosen reference loudspeaker observed at the control points in the bright zone. This choice of target is used to ensure that the target pressure in the bright zone is in the range space of the filtered loudspeaker responses.

[9]In the present paper, the weights are determined by a coarse grid search over the individual weights. Thus, the weights might not generalize to other setups. In a practical implementation, it might be necessary to determine maximum tolerable values for each penalty term and reformulate the problem as a constrained optimization.

$$\mathbf{q}_{\mathrm{bs}} = \alpha_{bs}\bar{\mathbf{U}}_k^T\bar{\Psi}(\vec{\mathbf{s}}_k)^T\bar{\Gamma}(\vec{\mathbf{s}}_k^{'})\bar{\mathbf{z}}_k. \tag{53}$$

Note that the matrices (46)-(50) are positive definite or positive semidefinite as they are constructed as symmetric outer matrix products. The addition by $\mathbf{Q}_w$ can be interpreted as Tikhonov regularization [42].

IV. Simulation Results

The behavior of the moving horizon framework can be illustrated with simulation examples. In this section, the RIRs are simulated as point sources radiating sound in a rectangular room using the expression for Green's function in a lightly damped room [43]. The simulated room is 5 m by 5 m by 2.7 m and has a reverberation time of 0.6 s. Two effects are investigated in this section: The effect of different control and prediction horizon lengths as well as the effect of zones which change position over time.

A. Evaluation Parameters

The reduction of the sound pressure in the dark zone is expressed by the contrast between the zones. The results presented in this paper are analyzed in the time-domain, hence, the contrast is defined as

$$\mathrm{Contrast}[k] = 10\log_{10}\left(\frac{M_b^{-1}\|\mathbf{p}_b[k]\|_2^2}{M_d^{-1}\|\mathbf{p}_d[k]\|_2^2}\right), \tag{54}$$

where $\mathbf{p}_b[k] \in \mathbb{R}^{M_b}$ is the pressure at the M_b control points in the bright zone at time-sample k. Equivalently, $\mathbf{p}_d[k] \in \mathbb{R}^{M_d}$ represents the pressures at the control points in the dark zone.

The normalized mean square error is introduced as

$$\mathrm{nmse}[k] = \frac{\|\mathbf{p}_b[k] - \mathbf{p}_t[k]\|_2^2}{K^{-1}\sum_{k=1}^{K}\|\mathbf{p}_t[k]\|_2^2} \tag{55}$$

with $\mathbf{p}_t[k] \in \mathbb{R}^{M_b}$ representing the target pressure in the bright zone and K is the number of time-steps used to evaluate the normalized mean square error.

The last parameter of interest is the signal energy driving the loudspeakers, expressed as

$$\mathrm{energy}[k] = \sum_{\ell=1}^{L}(y_k^{(\ell)}[k])^2. \tag{56}$$

The interest in this parameter, is that the sound field is only controlled explicitly within the zones. As the target pressure includes a target level, the signal energy is used to indicate the efficiency of the solution and the general sound pressure level outside the controlled zones. A high signal energy would in general indicate a lower efficiency and a higher sound pressure level outside the controlled zones.

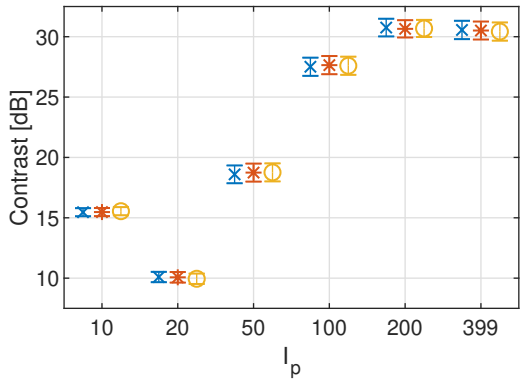
B. Control and Prediction Horizon Length

An important parameter choice for the MH framework is the choice of control and prediction horizon length. To reduce the latency in the implementation it is assumed that the input signal is only known ahead of time for the duration of the control horizon. After the control horizon and for the remainder of the prediction horizon, it is assumed that the input signal is switched off. This is done to ensure that the tentative filters at the end of the control horizon does not correspond to an average over the future unknown audio samples, but only control the decaying energy of the audio already reproduced in the room. Hereby, the maximal relevant length of the prediction horizon is the length of the linear convolution between the filters after the control horizon and the RIRs ($I_p \leq n + n_w + I_c - 1$).

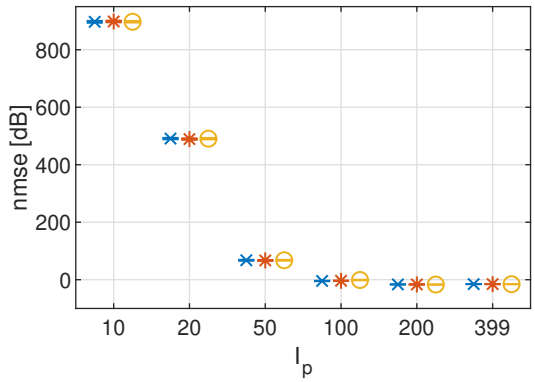
To illustrate the influence of the control and prediction horizon lengths, the performance of moving horizon in a simulated room is shown in Fig. 3 for various control and prediction horizons (the parameter choices for the framework are summarized in Table I and II). The sound zones were stationary in the initial positions shown in Fig. 4, where the RIRs in each zone were known at 9 points in a 3 by 3, planar array with 10 cm interelement separation. The input signal was white noise bandpass filtered between 35 and 500 Hz, with a 4th order Butterworth filter. The results in Fig. 3 show that for short prediction horizons, the moving horizon framework becomes unstable. This is understood as only a fraction of the signal energy stored in the sound field of the room as reverberation is represented for short prediction horizons. With a short prediction horizon, we seek to control all the energy stored in the state-vector without accurately modelling the consequence when choosing the current control filters. Therefore, the framework injects an increasing amount of energy at each time step to control the stored energy which is not predicted, and thus increases. The results in Fig. 3 indicate that it is not necessary to increase the prediction horizon to the maximal relevant length ($n + n_w + I_c - 1$), as long as the majority of the system energy is included in the prediction horizon. Furthermore, the control horizon length is of much lower importance than the length of the prediction horizon.

C. Moving Zones

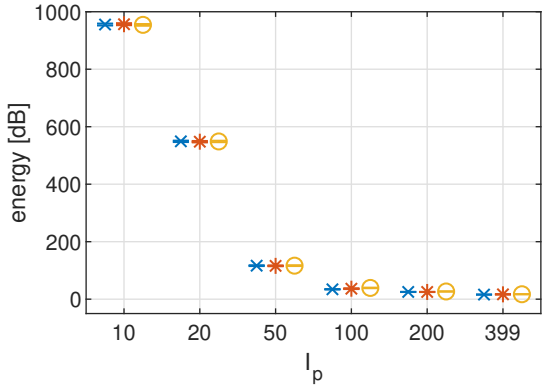
To exemplify moving sound zones, the following scenario is simulated: Two zones are moved at a steady pace of 0.5 m/s from one end of the room to the other, as sketched in Fig. 4. The RIRs are known along the path covered by the zones in a grid with 10 cm between adjacent control points. The mean square pressures in the moving bright and dark zone are plotted in Fig. 5 for two scenarios. In scenario 1) the MH framework updates the RIRs to follow the moving zones and predict future positions. At every time step, the future zone positions are predicted for the length of the prediction horizon, using the current zone position and directional velocity. In scenario 2) the MH framework utilizes RIRs centered around room length as sketched in Fig. 4 (the RIRs does not match the zone locations for the majority of the time steps). The results clearly show the advantage of updating the RIRs in the moving horizon framework to match the current zone positions. However, it is also observed that while the performance is optimal when the utilized RIRs match the zone position, the sound field is also controlled slightly outside the zone position.



(a) Contrast



(b) Normalized mean square error



(c) Summed input energy

Fig. 3. Mean and 95% confidence interval over 500 samples of white noise in the simulated room depicted in Fig. 4 with stationary zones. The results are reported for different prediction horizon lengths (I_p) and control horizon lengths (I_c). (×): $I_c = 1$. (*): $I_c = 2$. (○): $I_c = 5$.

TABLE I

FILTER TAP WEIGHTS USED TO CONTROL THE FILTER SHAPE FOR THE SIMULATED AND EXPERIMENTAL MOVING HORIZON RESULTS AS WELL AS THE EXPERIMENTAL REFERENCE STATIC FILTERS.

Sim. MH	$w_{\text{env}}[i] = 1,$	$i \in [0, 79]$
	$w_{\text{env}}[i] = \exp\{\ln(10^4)\frac{i-79}{20}\},$	$i \in [80, 99]$
Meas. MH	$w_{\text{env}}[i] = 1,$	$i \in [0, 199]$
	$w_{\text{env}}[i] = \exp\{\ln(10^4)\frac{i-199}{56}\},$	$i \in [200, 255]$
Static	$w_{\text{env}}[i] = \exp\{\ln(10^{12})\frac{5-i}{5}\},$	$i \in [0, 4]$
	$w_{\text{env}}[i] = 1,$	$i \in [5, 54]$
	$w_{\text{env}}[i] = \exp\{\ln(10^4)\frac{i-55}{201}\},$	$i \in [55, 255]$

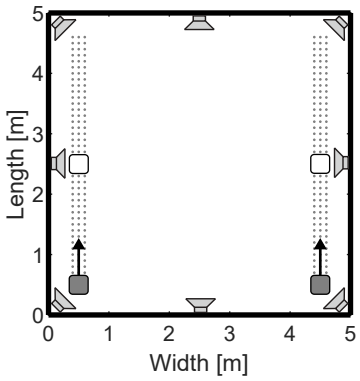


Fig. 4. Sketch of setup used to investigate moving zones in a simulated environment. The gray dots indicate positions where the room impulse responses are known. The gray-filled squares indicate zones moving in the direction of the arrows. The white-filled squares indicate zones which does not move.

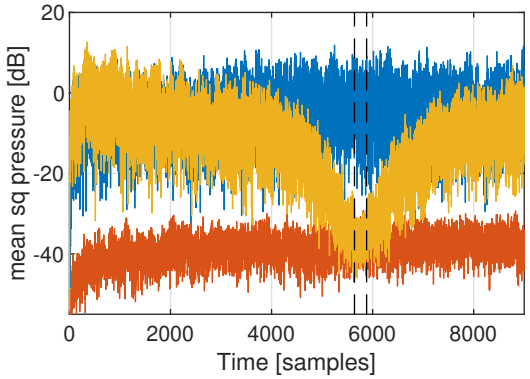


Fig. 5. Mean square pressure in the bright and dark zone in the moving zone scenario depicted in Fig. 4. (——): Mean square pressure in moving bright zone. (——): Mean square pressure in moving dark zone. (——): Mean square pressure in moving dark zone when the filters are optimized for a stationary position. The interval between the two black dashed lines (- - -) indicate where the stationary zone coincides with the moving zone.

TABLE II

COST-FUNCTION WEIGHTS FOR THE SIMULATED MOVING HORIZON WITH MOVING ZONES, THE EXPERIMENTAL MOVING HORIZON, AND THE EXPERIMENTAL CONVENTIONAL TIME-INVARIANT SOLUTION.

	Sim. MH	Exp. MH	Conventional
α_d	0.99	0.9915	0.97
α_b	0.01	0.0085	0.03
α_w	10^{-5}	5.5×10^{-5}	2.8×10^{-7}
α_{env}	0	5×10^{-6}	10^{-7}
α_{bs}	0	10^{-6}	10^{-5}
n	300	600	600
n_w	100	256	256
I_p	399	855	N/A
modelling delay	18 [samp.]	30 [samp.]	30 [samp.]

V. Experimental Results

In this section, the performance of the proposed framework is investigated using stationary zones and measured room impulse responses. The RIRs were measured in a 5.50 m by 8.65 m by 2.70 m raised ceiling room with an arithmetic average reverberation time $T_{20} = 0.6$ s over the $1/3$-octave bands between 31.5 and 315 Hz. The RIRs were measured using a 2 s logarithmic sweep from 0.1 Hz to 24 kHz [44]. The impulse responses were determined at a sampling frequency of 48 kHz and then downsampled to 1.2 kHz, to match the low frequency range of interest. Eight 10" loudspeakers, distributed as sketched in Fig. 1, were used to control the sound field in two zones. Each of the zones were sampled with a $3 \times 3 \times 3$ cubic array of 1/4" electret microphones with 10 cm between adjacent microphones. Two sets of RIRs were measured. The first set was used to determine the filters, while the second set was used to evaluate the resulting performance.

The purpose of sound zones is to generate multiple regions with different audio content playing in each zone. It is, therefore, of interest to compare the performance of the moving horizon to a conventional time-invariant solution using music as the input signals. The performance is investigated using two different audio excerpts as well as white noise. Due to the investigated frequency range, the audio excerpts are chosen for their significant low frequency content. The used excerpts are 10 s long and downsampled to 1200 Hz. All used audio signals were band-pass filtered between 35 and 500 Hz, as done for the simulation results.

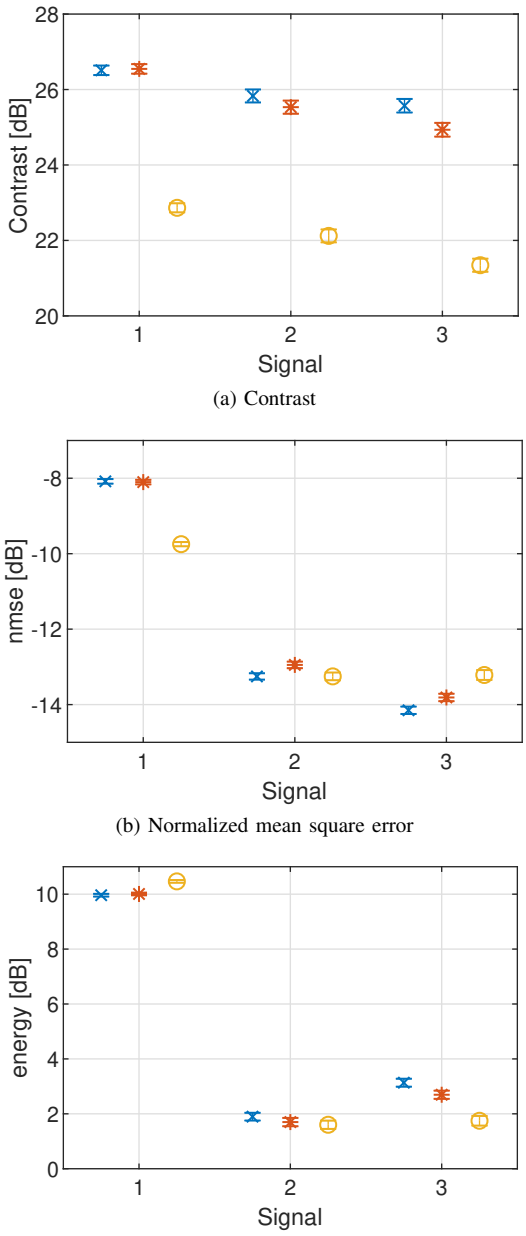
A. Conventional Time-Invariant Solution

A conventional method (presented in [40]) for generating sound zones was used to determine a set of time-invariant filters. The filters were based on the measured RIRs and employed a cost-function similar to (44), where the input signal was assumed to be a unit sample sequence.

B. Input Signal Dependence

Three different control methods are compared: MH with control horizon $I_c = 1$, MH with control horizon $I_c = 2$, and the conventional reference solution (the cost-function weights and settings used for the methods are summarized in Table I and II). To compare the results to the conventional time-invariant solution, the MH cost-function weights were adjusted to provide similar nmse and signal energy results to the conventional solution using input signal 2.

The results obtained using moving horizon and the conventional method with the different audio signals are displayed in Fig. 6. It is seen that the moving horizon results obtain about 4 dB higher contrast relative to the conventional method. The nmse-results are similar for the music excerpts while the error for the band-pass filtered white noise is increased when the MH solutions are used instead of the conventional solution. The loudspeaker signal energy required to attain the sound field control is seen to be similar for all the investigated methods. It is observed that the MH results with control horizon $I_c = 2$ performs slightly worse in terms of contrast and nmse, compared to control horizon $I_c = 1$. This is related to utilizing the same cost-function weights for both methods and the observation that the signal energy is lower for control horizon $I_c = 2$.



(a) Contrast

(b) Normalized mean square error

(c) Summed input energy

Fig. 6. Mean and 95% confidence interval over 11000 samples for three audio excepts. Signal 1: White noise. Signal 2: Electronic music excerpt. Signal 3: Rap music excerpt. (×): Moving horizon $I_c = 1$. (∗): Moving horizon $I_c = 2$. (○): Static Filters.

VI. Discussion

From the simulated results it is seen that it is possible to create moving sound zones, using the proposed MH framework. The solution relies on knowledge of the room impulse responses at the positions where the zones should move to. If these responses are known, it would also be possible to create multiple time-invariant solutions and cross-fade between them. However, to do so would require storing solutions for every

possible combination of bright and dark zone positions and sizes within the room. This is avoided in the MH framework at the expense of calculating the filters online.

The contrast results obtained using the proposed MH framework in Fig. 6 show that it is possible to obtain higher separation between the zones while attaining similar nmse and signal energy, compared to the time-invariant filters. This indicates that knowledge of the specific signal being reproduced in the zones enables more accurate control of the sound field. The difference between the conventional and the moving horizon solution can be related to the differences between the assumed input signals. As an example, consider a two-minute-long input signal and two control methods, A and B. In the first minute, the input signal has one spectrum and in the next minute the spectrum is different. Method A is based on the average spectrum over the two minutes, whereas method B applies different solutions for the first and second minute, matching the know signal spectra. Constrained to equal signal energy, method B will perform better than method A due to knowledge of the changing input spectrum. Moving horizon updates the control filters at each time-step to match the current input spectrum. Therefore, MH is able to attain better performance than the conventional method, which is based on an assumed white spectrum, when constrained to equal signal energy.

The downside with MH is that the specific behavior of the results depends on the audio content. This indicates that specific sound field control performance is not guaranteed for a given set of cost-function weights. If a particular performance is desired, the optimization should be redefined as a constrained optimization with appropriate constraints on the individual terms in the cost-function e.g. inspired from [45]. Furthermore, it might be of interest to extend the cost-function to directly control perceptual aspects of sound zones as suggested in e.g. [16], [19], [31], [46]. Similarly, the time-dependent changes in the reproduction system and scenario could be expressed and controlled by extending the framework to including these as subsystems in state-space form.

VII. Conclusion

In this paper, an adaptive framework based on moving horizon control was proposed for sound zones applications. It was seen that the framework can accommodate moving sound zones and input signals which change over time. Furthermore, it was seen that applying the framework in an experimental scenario with stationary zones, it is possible to increase the separation by about 4 dB, compared to a conventional time-invariant solution. This difference is attributed to the moving horizon framework being able to take the specific audio reproduced in the bright zone into account, whereas the conventional solution is designed to optimize the expected performance assuming a spectrally flat input signal.

Future work includes extending the framework to multiple audio signals reproduced simultaneously and explicitly constraining the sound field control for application relevant prior knowledge e.g. human perception or the available reproduction system.

Acknowledgment

This work was fully funded by Bang & Olufsen a/s.

References

[1] S. J. Elliott and M. Jones, "An active headrest for personal audio," *The Journal of the Acoustical Society of America*, vol. 119, no. 5, pp. 2702–2709, 2006.

[2] M. Jones and S. J. Elliott, "Personal audio with multiple dark zones," *The Journal of the Acoustical Society of America*, vol. 124, no. 6, pp. 3497–3506, December 2008.

[3] J. Cheer, S. Elliott, and M. F. S. Gálvez, "Design and implementation of a car cabin personal audio system," *J. Audio Eng. Soc.*, vol. 61, no. 6, pp. 412–424, June 2013.

[4] J. Cheer and S. J. Elliott, "A comparison of control strategies for a car cabin personal audio system," in *Audio Engineering Society Conference: 52nd International Conference: Sound Field Control - Engineering and Perception*, September 2013.

[5] J.-W. Choi, "Subband optimization for acoustic contrast control," in *22nd International Congress on Sound & Vibration*, July 2015.

[6] M. Olsen and M. B. Møller, "Sound zones: on the effect of ambient temperature variations in feed-forward systems," in *Audio Engineering Society Convention 142*, May 2017.

[7] F. Heuchel, D. C. Nozal, and F. T. Agerkvist, "Sound field control for reduction of noise from outdoor concerts," in *Audio Engineering Society Convention 145*, Oct 2018.

[8] W. F. Druyvesteyn and J. Garas, "Personal sound," *J. Audio. Eng. Soc.*, vol. 45, no. 9, pp. 685–701, September 1997.

[9] T. Betlehem and T. D. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms," *The Journal of the Acoustical Society of America*, vol. 117, no. 4, pp. 2100–2111, 2005.

[10] J. Rämö, S. Bech, and S. H. Jensen, "Validating a real-time perceptual model predicting distraction caused by audio-on-audio interference," *The Journal of the Acoustical Society of America*, vol. 144, no. 1, pp. 153–163, July 2018.

[11] J.-W. Choi and Y.-H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1695–1700, 2002.

[12] J.-W. Choi, Y. Kim, S. Ko, and J.-H. Kim, "Super-directive loudspeaker array for the generation of personal sound zone," in *Audio Engineering Society Convention 125*, 2008.

[13] Y. Cai, M. Wu, and J. Yang, "Design of a time-domain acoustic contrast control for broadband input signals in personal audio systems," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 341–345.

[14] M. F. S. Gálvez, S. J. Elliott, and J. Cheer, "Time domain optimization of filters used in a loudspeaker array for personal audio," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 11, pp. 1869–1878, November 2015.

[15] F. Olivieri, F. M. Fazi, P. A. Nelson, M. Shin, S. Fontana, and L. Yue, "Theoretical and experimental comparative analysis of beamforming methods for loudspeaker arrays under given performance constraints," *Journal of Sound and Vibration*, vol. 373, pp. 302–324, July 2016.

[16] J. Donley, C. Ritz, and W. B. Kleijn, "Multizone soundfield reproduction with privacy- and quality-based speech masking filters," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 6, pp. 1041–1055, June 2018.

[17] J. K. Nielsen, "Sound zones as an optimal filtering problem," in *Asilomar Conference on Signals, Systems and Computers. Conference Record*, 2018.

[18] T. Lee, J. K. Nielsen, J. R. Jensen, and M. G. Christensen, "A unified approach to generating sound zones using variable span linear filters," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2018, pp. 491–495.

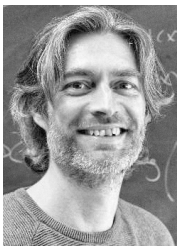
[19] T. Lee, J. K. Nielsen, and M. G. Christensen, "Towards perceptually optimized sound zones: A proof-of-concept study," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2019, pp. 136–140.

[20] P. Coleman, P. J. B. Jackson, M. Olik, M. Møller, M. Olsen, and J. A. Pedersen, "Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array," *J. Acoust. Soc. Am.*, vol. 135, no. 4, pp. 1929–1940, April 2014.

[21] J. Mourjopoulos, "On the variation and invertibility of room impulse response functions," *Journal of Sound and Vibration*, vol. 102, no. 2, pp. 217–228, 1985.

[22] J.-Y. Park, J.-W. Choi, and Y.-H. Kim, "Acoustic contrast sensitivity to transfer function errors in the design of a personal audio system," *The Journal of the Acoustical Society of America*, vol. 134, no. 1, pp. EL112–EL118, July 2013.
[23] X. Ma, P. J. Hegarty, J. A. Pedersen, and J. J. Larsen, "Impact of loudspeaker nonlinear distortion on personal sound zones," *J. Acoust. Soc. Am.*, vol. 143, no. 1, pp. 51–59, January 2018.
[24] S. J. Elliott, J. Cheer, J. Choi, and Y. Kim, "Robustness and regularization of personal audio systems," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 7, pp. 2123–2133, Sep. 2012.
[25] Q. Zhu, P. Coleman, M. Wu, and J. Yang, "Robust personal audio reproduction based on acoustic transfer function modelling," in *Audio Engineering Society Conference: 2016 AES International Conference on Sound Field Control*, 2016.
[26] ——, "Robust acoustic contrast control with reduced in-situ measurement by acoustic modeling," *J. Audio Eng. Soc.*, vol. 65, no. 6, pp. 460–473, June 2017.
[27] X. Ma, P. J. Hegarty, and J. J. Larsen, "Mitigation of nonlinear distortion in sound zone control by constraining individual loudspeaker driver amplitudes," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2018, pp. 456–460.
[28] G. C. Goodwin, M. M. Seron, and J. A. de Doná, *Constrained Control and Estimation*. Springer, 2005.
[29] J. Østergaard, D. E. Quevedo, and J. Jensen, "Real-time perceptual moving-horizon multiple-description audio coding," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4286–4299, September 2011.
[30] E. Fernandez-Grande, "Sound field reconstruction in a room from spatially distributed measurements," in *Proceedings of the 23rd International Congress on Acoustics*, September 2019, pp. 4961–4968.
[31] J. Francombe, P. Coleman, M. Olik, K. Baykaner, P. J. B. Jackson, R. Mason, M. Dewhirst, S. Bech, and J. A. Pederson, "Perceptually optimized loudspeaker selection for the creation of personal sound zones," in *Audio Engineering Society Conference: 52nd International Conference: Sound Field Control - Engineering and Perception*, September 2013.
[32] S. Cecchi, A. Carini, and S. Spors, "Room response equalization - a review," *Applied Sciences*, vol. 8, no. 1, 2018.
[33] G. C. Goodwin and K. S. Sin, *Adaptive filtering prediction and control*. Prentice-Hall, 1984.
[34] M. Shin, S. Q. Lee, F. M. Fazi, P. A. Nelson, D. Kim, S. Wang, K. H. Park, and J. Seo, "Maximization of acoustic energy difference between two spaces," *The Journal of the Acoustical Society of America*, vol. 128, no. 1, pp. 121–131, July 2010.
[35] M. B. Møller, M. Olsen, and F. Jacobsen, "A hybrid method combining synthesis of a sound field and control of acoustic contrast," in *Audio Engineering Society Convention 132*, 2012.
[36] J.-H. Chang and F. Jacobsen, "Sound field control with a circular double-layer array of loudspeakers," *The Journal of the Acoustical Society of America*, vol. 131, no. 6, pp. 4518–4525, June 2012.
[37] ——, "The effect of scattering on sound field control with a circular double-layer array of loudspeakers," in *Audio Engineering Society Convention 132*, 2012.
[38] O. Kirkeby and P. A. Nelson, "Reproduction of plane wave sound fields," *The Journal of the Acoustical Society of America*, vol. 94, no. 5, pp. 2992–3000, 1993.
[39] M. Poletti, "An investigation of 2-d multizone surround sound systems," in *Audio Engineering Society Convention 125*, Oct 2008.
[40] M. B. Møller and M. Olsen, "Sound zones: on envelope shaping of fir filters," in *24th International Congress on Sound & Vibration*, July 2017.
[41] W. M. Leach, *Introduction to Electroacoustics and Audio Amplifier Design, 3rd edt.* Kendall/Hunt, 2003.
[42] P. C. Hansen, *Discrete inverse problems: insights and algorithms*. Society for Industrial and Applied Mathematics, 2010.
[43] F. Jacobsen and P. M. Juul, *Fundamentals of General Linear Acoustics*. Wiley, 2013.
[44] A. Farina, "Advancements in impulse response measurements by sine sweeps," in *Audio Engineering Society Convention 122*, May 2007.
[45] Y. Cai, M. Wu, and J. Yang, "Sound reproduction in personal audio systems using the least-squares approach with acoustic contrast control constraint," *The Journal of the Acoustical Society of America*, vol. 135, no. 2, pp. 734–741, 2014.
[46] D. Wallace and J. Cheer, "Optimisation of personal audio systems for intelligibility contrast," in *144th Audio Engineering Society Convention*, May 2018.

Martin Bo Møller (S'17) received the M.Sc. degree in engineering acoustics from the Technical University of Denmark in 2011 and was employed by Bang & Olufsen A/S (B&O). Since 2013, he has been studying part-time towards the Ph.D. degree, sponsored by B&O, at the Department of Electronic Systems at Aalborg University, Denmark. His research interests include sound field control, loudspeaker array processing, as well as loudspeaker and room interaction.



Jan Østergaard Jan Østergaard (S'98-M'99-SM'11) received the M.Sc.E.E. degree from Aalborg University, Aalborg, Denmark, in 1999 and the PhD degree (*with cum laude*) from Delft University of Technology, Delft, The Netherlands, in 2007. From 1999 to 2002, he worked as an R&D Engineer at ETI A/S, Aalborg, and from 2002 to 2003, he was an R&D Engineer with ETI Inc., VA, USA. Between September 2007 and June 2008, he was a Postdoctoral Researcher at The University of Newcastle, NSW, Australia. He has been a Visiting Researcher at Tel Aviv University, Israel, and at Universidad Técnica Federico Santa María, Valparaíso, Chile. Dr. Østergaard is currently a Professor (WSR) in Information Theory and Signal Processing, and Head of the Centre on Acoustic Signal Processing Research (CASPR), at Aalborg University. He has received a Danish Independent Research Council's Young Researcher's Award, a Best PhD Thesis award by the European Association for Signal Processing (EURASIP), and fellowships from the Danish Independent Research Council and the Villum Foundation's Young Investigator Programme. He is an Associate Editor of EURASIP Journal on Advances in Signal Processing.