Reduced complexity for sound zones with subband block adaptive filters and a loudspeaker line array

Møller, Martin B.; Martinez, Jorge; Østergaard, Jan

DOI
10.1121/10.0025464

Publication date
2024

Document Version
Final published version

Published in
The Journal of the Acoustical Society of America

Citation (APA)
Møller, M. B., Martinez, J., & Østergaard, J. (2024). Reduced complexity for sound zones with subband block adaptive filters and a loudspeaker line array. *The Journal of the Acoustical Society of America*, *155*(4), 2314–2326. https://doi.org/10.1121/10.0025464

The Journal of the Acoustical Society of America

APRIL 01 2024

Reduced complexity for sound zones with subband block adaptive filters and a loudspeaker line array

Martin B. Møller ; Jorge Martinez ; Jan Østergaard

Check for updates

J. Acoust. Soc. Am. 155, 2314–2326 (2024)
https://doi.org/10.1121/10.0025464

View Online

Export Citation

ASA ACOUSTICAL SOCIETY OF AMERICA

JASA ARTICLE

Reduced complexity for sound zones with subband block adaptive filters and a loudspeaker line array

Martin B. Møller,[1,a)] Jorge Martinez,[2] and Jan Østergaard[3]

[1]*Research Department, Bang & Olufsen A/S, Struer, DK-7600, Denmark*

[2]*Department of Intelligent Systems, Multimedia Computing Group, Delft University of Technology, Netherlands*

[3]*Department of Electronic Systems, Section on Artificial Intelligence and Sound, Aalborg University, 9220 Aalborg, Denmark*

ABSTRACT:
Sound zones are used to reproduce individual audio content to multiple people in a room using a set of loudspeakers with controllable input signals. To allow the reproduction of individual audio to dynamically change, e.g., due to moving listeners, changes in the number of listeners, or changing room transfer functions, an adaptive formulation is proposed. This formulation is based on frequency domain block adaptive filters and given room transfer functions. To reduce computational complexity, the system is extended to subband processing without cross-adaptive filters. The computational savings come from recognizing that sound zones consist of part-solutions which are inherently band limited, hence, several subbands can be ignored. To validate the theoretical findings, a 27-channel loudspeaker array was constructed, and measurements were performed in anechoic and reflective environments. The results show that the subband solution performs identically to a full-rate solution but at a reduced computational complexity.
© *2024 Acoustical Society of America*. https://doi.org/10.1121/10.0025464

(Received 29 September 2023; revised 9 March 2024; accepted 12 March 2024; published online 1 April 2024)

[Editor: James F. Lynch] Pages: 2314–2326

I. INTRODUCTION

The concept behind sound zones is to reproduce individual audio in separate regions of a room using loudspeakers, where the input signal can be adjusted by a controllable finite impulse response (FIR) filter (Betlehem *et al.*, 2015). The general strategy for sound zones is that different part-solutions are applied to cover different parts of the audible frequency range. An example is using active control at low frequencies, beamforming at mid frequencies, and passive directivity control at high frequencies (Druyvesteyn and Garas, 1997). The need for part-solutions is due to the loudspeakers and room interacting differently depending on the room dimensions relative to the wavelength of the reproduced sound. Another factor is that loudspeaker drivers are usually optimized for reproducing audio in specific frequency ranges. This naturally leads to sound zone processing happening in limited frequency bands. At low frequencies, the loudspeakers are typically large, and woofers are distributed around the room (Cheer *et al.*, 2013; Druyvesteyn and Garas, 1997; Møller *et al.*, 2019), whereas mid-to-high frequency solutions often come in the form of compact loudspeaker arrays (Elliott *et al.*, 2012; Gálvez *et al.*, 2015; Møller and Olsen, 2019). Throughout this work, we will utilize these observations to propose an adaptive formulation of sound zones with reduced computational complexity.

The intention with sound zones is to support the activities of multiple individuals in a room. As people are rarely stationary, it is of interest to dynamically adapt the processing relative to the context. Examples of dynamic changes are zones changing location in the room (Møller and Østergaard, 2020), changes in zone size (Jacobsen *et al.*, 2022), and changes in the ambient temperature (Olsen and Møller, 2017).

The sound zones processing consists of processing the zone-specific audio signals for each loudspeaker before the signals are reproduced. The processing is based on assumed knowledge of how each loudspeaker radiates sound to the spatial regions where zones should be created. To adapt to the changes, it is necessary to update the processing accordingly. In Møller and Østergaard (2020), a moving horizon approach was suggested to update the FIR filters for each loudspeaker for every new audio sample. While this method provides beneficial performance, it is also computationally intensive. In Moles-Cases *et al.* (2020), a set of static FIR filters was recalculated, when necessary, as the solution to a least squares problem using subband decomposition to reduce the complexity of the associated inverse problem. In Hu *et al.* (2023), Vindrola *et al.* (2021), and Zhao and Burnett (2022), adaptive procedures with microphones in the zones were suggested as a way of compensating for changes in the transfer functions. In the present work, it is assumed that estimates of the room transfer functions are made available by a secondary system, e.g., by premeasuring the transfer functions to multiple locations, via sound field extrapolation from remotely located microphones (Caviedes-Nozal *et al.*, 2021; Jin and Kleijn, 2015; Lluís *et al.*, 2020; Pham Vu and Lissek, 2020), or by

[a)] Email: mim@bang-olufsen.dk

assuming a point source radiation model. The focus is then on the speed of adaptation to this new information as well as the associated computational complexity.

To reduce computational complexity, it is possible to leverage that the part-solutions of a sound zone system are band limited through crossover networks. These are designed with respect to the frequency range where the loudspeaker drivers are effective, both in terms of their frequency response as well as their spatial position. It is desired to utilize this inherent band limitation to reduce the sampling frequency in each of the frequency bands. One approach for doing this is to introduce subband processing where the bands, that should not be reproduced by a given loudspeaker driver, can safely be skipped in the processing. This was investigated for the calculation of static filters in Moles-Cases *et al.* (2020), where determining the filters relied on solving a least squares problem with complexity increasing approximately with the cube of the number of loudspeakers times the length of the control filters. Given the complexity of the problem being solved, the authors of that paper observed a large reduction in computational complexity by dividing the problem into separable subbands (reducing the length of the control filter in the individual subbands). The premise of the present work is a system which constantly adapts to the input signal and transfer functions. To keep the complexity low, the adaptive system of choice is a gradient based adaptive filter in the form of a frequency domain block adaptive filter strategy.

The rest of this paper is structured as follows. Section III introduces the block-based data model used for frequency domain block adaptive filtering in a sound zones context, where the feedback is a prediction based on given room transfer functions. Section IV introduces how the problem can be separated into individual subbands. In Sec. V the algorithm is investigated in terms of computational complexity, tracking performance, and sound field control performance, while being evaluated using a purpose-built loudspeaker array. Potential audible artifacts and the latency of the proposed system are briefly discussed in Sec. VI before the conclusion in Sec. VII.

II. NOTATION

In this paper we apply lower- and upper-case bold letters **a**, **A** for vectors and matrices of time-domain samples. Transform domain (Fourier and Z-transform) vector and matrix quantities are denoted by italic lower and upper case roman letters, $\boldsymbol{a}$, $\boldsymbol{A}$. Parentheses super script $\boldsymbol{a}^{(m,\ell)}$ is used for indexing microphones and loudspeakers, respectively. Identity and zero matrices are denoted as $\mathbf{I}_M$ and $\mathbf{0}_{M\times N}$, where M and N denote the dimensions of the matrices. Superscript $(\cdot)^\mathsf{T}$ and $(\cdot)^\mathsf{H}$ are used to denote regular and Hermitian transpose, while $(\cdot)^*$ denotes complex conjugation.

III. BLOCK-BASED SOUND ZONE FILTERS

The basic situation, which is sought solved in this work, is to reproduce sound in a bright zone, while suppressing it in a dark zone using a given loudspeaker array. This situation is the basic building block for creating multiple sound zones through superposition of an additional solution where the designations of the bright and dark zones are swapped. The base case is sketched in Fig. 1, where the given loudspeakers form a line array. We can mathematically formulate the sound field control problem as the optimization problem

$$\min_{\boldsymbol{w}} \quad \sum_{m=1}^{M} ||\boldsymbol{p}^{(m)}(\boldsymbol{w}) - \boldsymbol{t}^{(m)}||_2^2 + \lambda \sum_{\ell=1}^{L} ||\boldsymbol{w}^{(\ell)}||_2^2, \tag{1}$$

where $\boldsymbol{p}^{(m)}(\boldsymbol{w})$ is the reproduced pressure which is a function of the concatenated control filters $\boldsymbol{w}$ and $\boldsymbol{t}^{(m)}$ is the target pressure. The vectors of frequency components $\boldsymbol{p}^{(m)}(\boldsymbol{w})$ and $\boldsymbol{t}^{(m)}$ represent the reproduced and target sound pressure at control point m of M, respectively. The control filter of the ℓ th loudspeaker is denoted $\boldsymbol{w}^{(\ell)}$ and $\boldsymbol{w}$ represents the concatenated filters for the L available loudspeakers. The parameter λ is a positive scalar adjusting the penalty on the norm of the concatenated filters. Thus, the optimization problem describes our desire to minimize the discrepancy between the reproduced and target sound fields, while penalizing filters with large coefficient amplitudes. The separation between two sound zones is introduced by defining $\boldsymbol{t}^{(m)} = \mathbf{0}$ for control points in the dark zone, which is similar to the general pressure matching method to calculate static filters for sound zones (Poletti, 2008). The target sound field in the bright zone can be chosen as, e.g., the delayed response of the centermost loudspeaker driver in the line array. The problem in Eq. (1) implies equal importance towards minimizing the reproduction error in the bright and dark zones. Emphasis on, e.g., reducing the sound pressure level in the dark zone can be introduced by changing the ℓ_2-norm to a weighted ℓ_2-norm having greater weights at the control points in the dark zone as described in Chang and Jacobsen (2012), Gálvez *et al.* (2015), and Shin *et al.* (2010).

A. Data model

In this section, the block-based data model is introduced and used to refine the cost-function of Eq. (1). The filters in the final formulation are expressed as leaky frequency domain block adaptive filters.

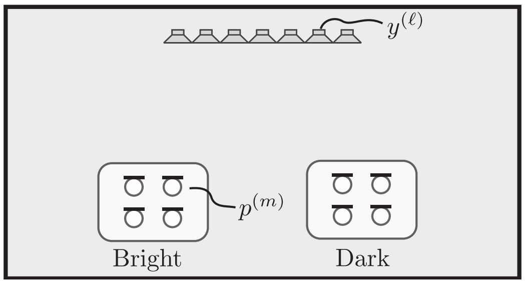


FIG. 1. Sketch of a sound zones system consisting of a line array and a bright and dark zone.

The sound pressure due to the filtered output of loudspeaker ℓ at microphone m at time sample n can be expressed as the discrete convolution

$$\mathrm{p}^{(m,\ell)}[n] = \sum_{i=0}^{I-1}\sum_{j=0}^{J-1} \mathrm{u}[n-j-i]\mathrm{r}^{(m,\ell)}[i]\mathrm{w}^{(\ell)}[j], \tag{2}$$

where $\mathrm{u}[n]$ is the nth sample of the input audio signal and define $\mathbf{w}^{\ell} := [\mathrm{w}^{(\ell)}[0]\cdots\mathrm{w}^{(\ell)}[J-1]]^{\mathsf{T}}$ as the length-J FIR filter for the ℓ th loudspeaker and $\mathbf{r}^{m,\ell} := [\mathrm{r}^{(m,\ell)}[0]\cdots\ \mathrm{r}^{(m,\ell)}[I-1]]^{\mathsf{T}}$ as the length-I room impulse response (RIR) from the ℓ th loudspeaker to the mth control point.[1]

Similar to the microphone array data model presented in Buchner *et al.* (2005), the reproduced pressure can be written in terms of block-based processing. Here, an overlap-save scheme with 50% overlap between blocks of time-domain samples is chosen. The block size is denoted $2B$ ($B \geq J+I-1$), i.e., the scheme produces B output samples at block index i. This is written as

$$\mathbf{p}_i^{(m,\ell)} = \mathbf{Z}^{01}\mathbf{W}_{2B}^{-1}\boldsymbol{U}_i\boldsymbol{R}^{(m,\ell)}\mathbf{W}_{2B}\mathbf{Z}^{10}\mathbf{w}^{(\ell)}, \tag{3}$$

where $\mathbf{W}_{2B}$ is the discrete Fourier transform (DFT) matrix with elements $e^{-j2\pi cg/(2B)}$ with $c, g = 0,\ldots,2B-1$ and

$$\boldsymbol{U}_i := \mathrm{diag}\left\{\mathbf{W}_{2B}\begin{bmatrix}\mathrm{u}[iB-B]\\ \vdots\\ \mathrm{u}[iB+B-1]\end{bmatrix}\right\}, \tag{4}$$

$$\boldsymbol{R}^{(m,\ell)} := \mathrm{diag}\left\{\mathbf{W}_{2B}\begin{bmatrix}\mathbf{r}^{(m,\ell)}\\ \mathbf{0}_{2B-I}\end{bmatrix}\right\}, \tag{5}$$

$$\mathbf{Z}^{01} := \begin{bmatrix}\mathbf{0}_{B\times B} & \mathbf{I}_B\end{bmatrix}, \quad \mathbf{Z}^{10} := \begin{bmatrix}\mathbf{I}_J & \mathbf{0}_{J\times 2B-J}\end{bmatrix}^{\mathsf{T}}. \tag{6}$$

Note that $\mathbf{W}_{2B}^{-1}\boldsymbol{U}_i\boldsymbol{R}^{(m,\ell)}\mathbf{W}_{2B}$ defines a circulant matrix, and this form expresses its diagonalization in terms of the DFT matrix explicitly. Furthermore, $\mathbf{Z}^{01}\mathbf{W}_{2B}^{-1}\boldsymbol{U}_i\boldsymbol{R}^{(m,\ell)}\mathbf{W}_{2B}\mathbf{Z}^{10}$ defines a Toeplitz matrix, which explicitly expresses the discrete convolution operation given in Eq. (2).

Multiplying both sides of Eq. (3) with the B-point DFT matrix we express the sound pressure in the frequency domain

$$\boldsymbol{p}_i^{(m,\ell)} = \mathbf{G}_{B\times 2B}^{01}\boldsymbol{U}_i\boldsymbol{R}^{(m,\ell)}\boldsymbol{w}_{2B}^{(\ell)}, \tag{7}$$

where

$$\boldsymbol{w}_{2B}^{(\ell)} := \mathbf{G}_{2B\times J}^{10}\mathbf{W}_J\mathbf{w}^{\ell}, \tag{8}$$

$$\mathbf{G}_{B\times 2B}^{01} := \mathbf{W}_B\mathbf{Z}^{01}\mathbf{W}_{2B}^{-1}, \tag{9}$$

$$\mathbf{G}_{2B\times J}^{10} := \mathbf{W}_{2B}\mathbf{Z}^{10}\mathbf{W}_J^{-1}. \tag{10}$$

The data model for the sound pressure at control point m due to the contributions from L loudspeakers can be written as

$$\boldsymbol{p}_i^{(m)} = \mathbf{G}_{B\times 2B}^{01}\boldsymbol{U}_i\boldsymbol{R}^{(m)}\boldsymbol{w}_{2BL}, \tag{11}$$

where

$$\boldsymbol{R}^{(m)} := \begin{bmatrix}\boldsymbol{R}^{(m,1)} & \cdots & \boldsymbol{R}^{(m,L)}\end{bmatrix}, \tag{12}$$

$$\boldsymbol{w}_{2BL} := \begin{bmatrix}\boldsymbol{w}_{2B}^{(1)\mathsf{T}} & \cdots & \boldsymbol{w}_{2B}^{(L)\mathsf{T}}\end{bmatrix}^{\mathsf{T}}. \tag{13}$$

The reproduction error for the ith block at control point m is given as

$$\boldsymbol{e}_i^{(m)} := \boldsymbol{t}_i^{(m)} - \mathbf{G}_{B\times 2B}^{01}\boldsymbol{U}_i\boldsymbol{R}^{(m)}\boldsymbol{w}_{2BL}, \tag{14}$$

where $\boldsymbol{t}_i^{(m)}$ are the B DFT-coefficients of the target pressure at control point m for block i. To ensure that the target sound field in the bright zone is achievable with the available loudspeakers, we choose to model the target pressure in terms of the available room impulse responses. The target is then expressed as

$$\boldsymbol{t}_i^{(m)} := \mathbf{G}_{B\times 2B}^{01}\boldsymbol{U}_i\boldsymbol{R}^{(m,\ell_t)}\boldsymbol{w}_{2B}^{(m,\ell_t)}, \tag{15}$$

where $\boldsymbol{R}^{(m,\ell_t)}$ is the transfer function from the target loudspeaker[2] (denoted by index ℓ_t) to control point m and $\boldsymbol{w}_{2B}^{(m,\ell_t)}$ is the corresponding filter modifying the input signal for the mth control point. Note that the target filter could be different for each control point. In this work, it is a modelling delay (Nelson *et al.*, 1992) for control points in the bright zone and zero for points in the dark zone. This data model is graphically represented in Fig. 2, where $\boldsymbol{R}$ represents the collection of RIRs as expressed in Eq. (A2).

B. Leaky frequency domain block adaptive filters

With the data model defined, it is now possible to restate the cost function from Eq. (1). The first step is to define the error as the difference between the reproduced and target sound fields. The desired outcome is to minimize the mean square reproduction error at the control points adaptively. Thus, the cost-function at time step i can be expressed as

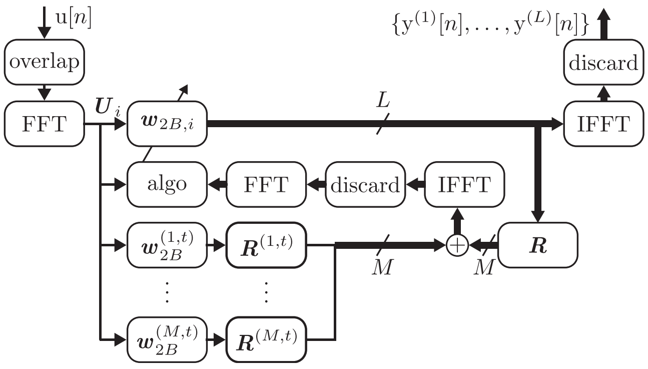


FIG. 2. Graphical representation of the overlap save based data model. As seen from the update Eqs. (20)–(22) the algorithm "algo" utilizes the given RIRs $\boldsymbol{R}$.

$$J(\boldsymbol{w}_i) := \sum_{m=1}^{M} ||\boldsymbol{e}_i^{(m)}||_2^2 + \lambda ||\boldsymbol{w}_i||_2^2, \quad \lambda \in \mathbb{R}^+. \tag{16}$$

The (leaky) term $\lambda||\boldsymbol{w}_i||^2$ has been added as a regularization term penalizing amplification introduced by the loudspeaker filters.

One approach to reduce this cost function (in order to minimize its expectation as $i \to \infty$) is as an adaptive filter with a gradient update step expressed as

$$\boldsymbol{w}_{i+1} = \boldsymbol{w}_i - \alpha \nabla J(\boldsymbol{w}_i), \quad \alpha \in \mathbb{R}^+. \tag{17}$$

In the above, $\nabla J(\boldsymbol{w}_i) := \partial J(\boldsymbol{w}_i)/\partial \boldsymbol{w}_i^*$ is the complex gradient [as introduced in Brandwood (1983)] and $\boldsymbol{w}_i$ denotes the filters at block index i, while $\boldsymbol{w}_{i+1}$ denotes the next block.

1. Update equations

To express the update equations and simplify the computation of the gradient step, it is desirable to express it in terms of the frequency spectra of the zero-padded control filters $\boldsymbol{w}_{2BL}$.

With the used definition of the complex gradient, the gradient of the cost function is determined as

$$\nabla J(\boldsymbol{w}_i) = (\mathbf{G}_{2BL\times JL}^{10})^{\mathsf{H}} \left(-\sum_{m=1}^{M} \boldsymbol{R}^{(m)\mathsf{H}} \boldsymbol{U}_i^{\mathsf{H}} \boldsymbol{e}_{2B,i}^{(m)} + \lambda \boldsymbol{w}_{2BL} \right), \tag{18}$$

where

$$\mathbf{G}_{2BL\times JL}^{10} := \mathbf{I}_L \otimes \mathbf{G}_{2B\times J}^{10}, \quad \boldsymbol{e}_{2B,i}^{(m)} := (\mathbf{G}_{B\times 2B}^{01})^{\mathsf{H}} \boldsymbol{e}_i^{(m)}. \tag{19}$$

In the above $\otimes$ denotes the Kronecker product. We can multiply both sides of Eq. (17) by $\mathbf{G}_{2BL\times JL}^{10}$ and introduce $\mathbf{G}_{2BL\times 2BL}^{10} := \mathbf{G}_{2BL\times JL}^{10}(\mathbf{G}_{2BL\times JL}^{10})^{\mathsf{H}}$ and $\mathbf{G}_{2B\times 2B}^{01} := (\mathbf{G}_{B\times 2B}^{01})^{\mathsf{H}}\mathbf{G}_{B\times 2B}^{01}$ to obtain the updated equations

$$\boldsymbol{e}_{2B,i}^{(m)} = \mathbf{G}_{2B\times 2B}^{01} \boldsymbol{U}_i \left(\boldsymbol{R}^{(m,t)} \boldsymbol{w}_{2B}^{(m,t)} - \boldsymbol{R}^{(m)} \boldsymbol{w}_{2BL,i} \right), \tag{20}$$

$$\begin{aligned} \nabla J_{2B}(\boldsymbol{w}) &:= \mathbf{G}_{2BL\times JL}^{10} \nabla J(\boldsymbol{w}) \\ &= \mathbf{G}_{2BL\times 2BL}^{10} \left(-\sum_{m=1}^{M} \boldsymbol{R}^{(m)\mathsf{H}} \bar{\boldsymbol{U}}_i^{\mathsf{H}} \boldsymbol{e}_{2B,i}^{(m)} + \lambda \boldsymbol{w}_{2BL,i} \right), \end{aligned} \tag{21}$$

$$\boldsymbol{w}_{2BL,i+1} = \boldsymbol{w}_{2BL,i} - \alpha \nabla J_{2B}(\boldsymbol{w}_i). \tag{22}$$

In the above update step, the input data block $\bar{\boldsymbol{U}}_i := \boldsymbol{\Lambda}_i^{-1}\boldsymbol{U}_i$ is introduced to potentially pre-whiten the audio data in the gradient step, as done in transform domain LMS filters [Sayed (2008), Chap. 26 and Yang *et al.* (2019)]. In this work, pre-whitening of the audio data is chosen and $\boldsymbol{\Lambda}_i = \gamma\boldsymbol{\Lambda}_{i-1} + (1-\gamma)\boldsymbol{U}_i^{\mathsf{H}}\boldsymbol{U}_i$ (with $0 < \gamma < 1$). Note, that all matrices besides $\mathbf{G}_{2BL\times 2BL}^{10}$ and $\mathbf{G}_{2B\times 2B}^{01}$ are compositions of diagonal matrices. Hence, the update steps can be implemented using element wise multiplications and fast Fourier transform (FFT) operations, rather than dense matrix-matrix multiplications. Thus, the complexity of the adaptive filter update scales with the $\mathcal{O}(2B\log_2(2B))$ of the FFT.

The maximum step size for which the resulting adaptive filters are stable in the mean square sense is considered in Appendix A. To improve the trade-off between convergence rate and misalignment error, a variable step-size algorithm can be applied. The step-size update rule applied in this work is described in Appendix B.

IV. SUBBAND DECOMPOSITION

In this section, decomposition of linear systems in subband components is introduced. The purpose for this is to express sound zones in terms of subband adaptive filters. The approach described in this section is based on the work presented in Moles-Cases *et al.* (2020) and Reilly *et al.* (2002), and will only be summarized here.

Sound zones processing generally requires that we predict the sound pressure at given positions in space through estimated room impulse responses or free-field radiation assumptions. These RIRs constitute a linear model, which can be used for feed-forward control as shown with Eq. (3). The motivation for introducing subband processing is to reduce the sampling rate at which we process the signals to match the target frequency range of each loudspeaker driver.

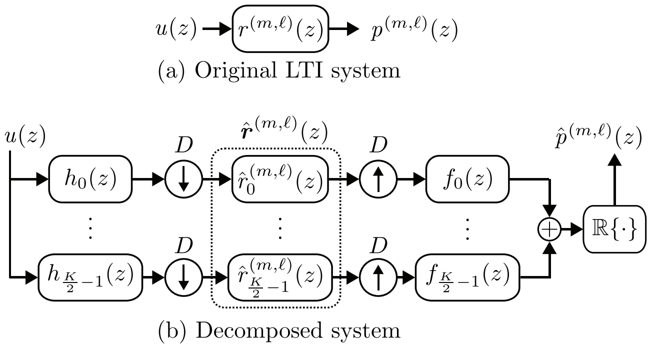
To process the loudspeaker signals in each subband separately (without mutual coupling between subbands), it is desired to approximate the RIRs by corresponding subband FIR filters, operating at the reduced sampling rate as suggested in Reilly *et al.* (2002). This requirement introduces some constraints on the applied analysis and synthesis filter banks. To obtain a filter bank with negligible "in-band" aliasing as well as (near) perfect reconstruction, it is necessary to design an oversampled filter bank (Harteneck *et al.*, 1998; Kellermann, 1988), since the mutual coupling between adjacent subbands is unavoidable for accurate system description using critically sampled filter banks (Gilloire and Vetterli, 1992).

A. System decomposition in subbands

The procedure for decomposing a linear time-invariant (LTI) system into subband systems, proposed in Reilly *et al.* (2002), utilizes a generalized DFT (GDFT) filter bank. The concept is illustrated in Fig. 3, where $r^{(m,\ell)}(z)$ is the original LTI system (a single RIR), and $\hat{\boldsymbol{r}}^{(m,\ell)}(z)$ is the decomposed subband approximation. The idea is that we would have $\hat{r}^{(m,\ell)}(z) = \beta z^{-d_0} p^{(m,\ell)}(z)$, where $\beta \in \mathbb{R}^+$ and d_0 is a delay such that $\hat{p}^{(m,\ell)}(z)$ is equal to $p^{(m,\ell)}(z)$ up to a scaling and a delay.

B. Filterbank requirements

The approach used in Reilly *et al.* (2002), relies on two sufficient conditions in order to approximate $r^{(m,\ell)}(z)$ by the subband components $\hat{r}_k^{(m,\ell)}(z), k \in 0, \ldots, K/2-1$, without mutual coupling between the subbands. The first condition is that there should be almost no frequency overlap between



(a) Original LTI system

(b) Decomposed system

FIG. 3. Schematic overview of original system and approximate subband decomposition.

the repeated (or modulated) spectra of the kth analysis filter $h_k(zW_D^d)$ (due to down sampling) and the corresponding synthesis filter $f_k(z)$. This can be described as the equation

$$f_k(z)h_k(zW_D^d)|_{z=e^{j\omega t}} \approx 0, \quad d=1,\dots,D-1, \quad 0\le\omega<2\pi, \tag{23}$$

where D is the down sampling factor and $W_D^d = e^{-j2\pi d/D}$. This property can be realized by a suitable choice of oversampling ratio and transition band of the lowpass prototype filter.

The second condition is that of a near perfect reconstruction of the filter bank, i.e.,

$$\frac{1}{2}\sum_{k=0}^{K/2-1}\mathbb{R}\{f_k(z)h_k(z)\}\approx\beta z^{-d_0}, \tag{24}$$

where $\mathbb{R}\{\cdot\}$ denotes taking the real part of the argument.

C. Decomposition into subsystems

The subband components are determined in such a way that the subband system approximates the target LTI system $r^{(m,\ell)}(z)$ in a least squares optimal way. If the in-band aliasing is kept small, the channels in the filter bank can be treated independently (Reilly *et al.*, 2002). The error between the true subband channel and approximated subband channel is then expressed as

$$e_k^{(m,\ell)}(z) := \frac{1}{D}\sum_{d=0}^{D-1}\Big(h_k(z^{1/D}W_D^d)\hat r_k^{(m,\ell)}(z) - h_k(z^{1/D}W_D^d)r^{(m,\ell)}(z^{1/D}W_D^d)\Big). \tag{25}$$

Hereby, one choice for the decomposed subband system is the one which minimizes the squared error

$$\hat r_{k,\mathrm{LS}}^{(m,\ell)}(e^{j\omega}) = \mathrm{argmin}\,\frac{1}{2\pi}\int_{-\pi}^{\pi}|e_k^{(m,\ell)}(e^{j\omega})|^2 d\omega. \tag{26}$$

Additionally, the least squares subband components can be computed as time-domain impulse responses as described in Moles-Cases *et al.* (2020) and Reilly *et al.* (2002).

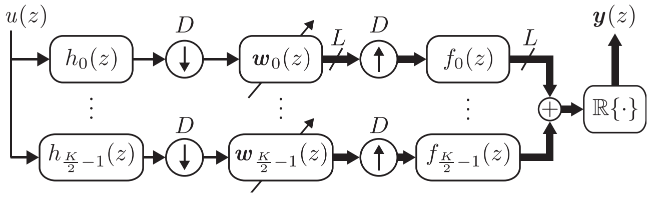


FIG. 4. Schematic overview of the subband adaptive filtering system going from the single input audio signal $u(z)$ to the L loudspeaker signals $\boldsymbol{y}(z) = [y^{(1)}(z),\dots,y^{(L)}]^T$. Each of the adaptive filter blocks $\boldsymbol{w}_0(z)$ to $\boldsymbol{w}_{K/2-1}(z)$ represents a system of the type depicted in Fig. 2.

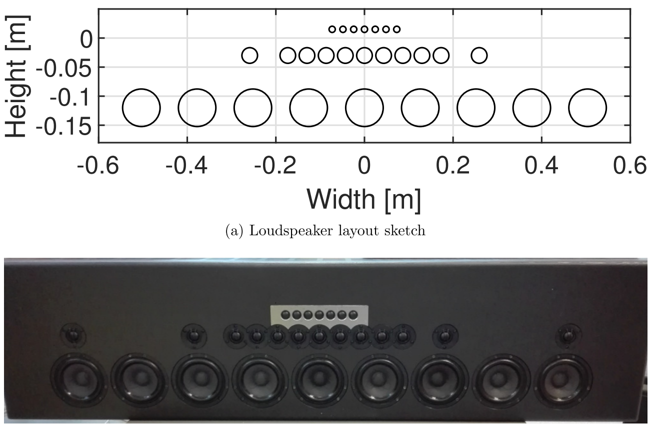
D. Relation to the control problem

In order to illustrate how the adaptive sound zone filter algorithm of Sec. III fits into the subband processing framework, the relationship is sketched in Fig. 4. In this figure, it is illustrated that the adaptation of the loudspeaker control filters $\boldsymbol{w}$ are performed independently in each subband channel. Thus, in each subband the control filters are updated according to the update Eqs. (20)–(22). The loudspeaker control filters are implemented directly in the subbands and the loudspeaker input signals are recovered through synthesis of the subband components.

V. RESULTS

A. Loudspeaker array

For the results, we used the loudspeaker array depicted in Fig. 5. This array is designed for flexible reproduction in most of the audible frequency range. Such a design poses the challenges that the array should be both comparable to the wavelength at the lowest frequency of interest, and that the interelement distance should be less than half of the wavelength at the highest frequency of interest [to avoid spatial aliasing (Ahrens and Spors, 2010)]. To reduce the required number of loudspeaker drivers, such a linear array can be implemented as a combination of harmonically nested linear arrays (Radmanesh *et al.*, 2016). To ensure a uniform control performance, it is desired that the chosen



(a) Loudspeaker layout sketch

(b) Picture of implemented array

FIG. 5. (Color online) Sketch and picture of the 27 driver linear loudspeaker array utilized for the evaluation of results.

loudspeaker drivers radiate sound omnidirectionally in their frequency range of concern.

The outcome design is an array consisting of three different types of loudspeaker drivers, a 12 cm midwoofer, a 19 mm subtweeter, and a 14 mm tweeter. The layout of the drivers is based on harmonically nesting line arrays of 7 drivers. This leads to one line of tweeters, two lines of subtweeters (of single and double spacing), and two lines of midwoofers (of single and double spacing) that has been truncated to the given width of the array. The minimum center-to-center distances between the drivers are 12.6, 4.3, and 2.4 cm, corresponding to spatial aliasing at 1.36, 3.99, and 7.15 kHz. The crossover frequency between midwoofer and subtweeter is chosen at 1.5 kHz to slightly increase the frequency overlap where both midwoofer and subtweeter are capable of emitting sound. The crossover frequency between subtweeter and tweeter is chosen as 4 kHz. The tweeters are furthermore lowpass filtered at 10 kHz as a compromise between increasing the bandwidth and limiting spatial aliasing artifacts. This array has been used for field tests, although those tests did not utilize the presented adaptive algorithm (Jacobsen *et al.*, 2023).

With three different types of loudspeaker drivers which cover the approximate frequency ranges 100 Hz–1.5 kHz, 1.5–4 kHz, and 4–10 kHz, respectively, it is clear that no loudspeaker driver requires information about the full audio bandwidth. As such, it is possible to restrict the subband channels, which are processed for each loudspeaker driver, to those that overlap with the frequency range reproduced by the loudspeaker driver. This observation can significantly reduce the computational complexity as described next.

B. Complexity analysis

The computational complexity in this section is considered as the number of floating-point operations (flops) performed per full-rate input sample. For the full-rate solution this means that the complexity is given as

$$C_{\mathrm{FR}} = C_{\mathrm{AF}}/B, \tag{27}$$

where C_{AF} is the flops required to update the adaptive filters and output B samples to each loudspeaker (given B new input samples).

For the subband solution, the length of the subband channel room impulse response is (Reilly *et al.*, 2002)

$$N_{R_k} = \left\lceil \frac{N_P + N_R - 1}{D} \right\rceil - \left\lceil \frac{N_P}{D} \right\rceil + 1, \tag{28}$$

where N_P is the length of the prototype filter used for the filter bank, N_R is the length of the full-rate room impulse response, and $\lceil \cdot \rceil$ denotes the ceiling operator. From the adaptive overlap-save framework, the block size B is chosen larger than the linear convolution between the room impulse response and the loudspeaker control filter. If the control filter in the subband channel has the length $J_{w\downarrow D} = \lceil J/D \rceil$ the subband block size $B_{\downarrow D}$ is chosen as the next radix 2 number larger than $N_{R_k} + J_{w\downarrow D} - 1$.

The subband number of computations per sample of the full-rate input signal can be expressed as

$$C_{\mathrm{SB}} = (B_{\downarrow D}(C_{\mathrm{AN}} + LC_{\mathrm{SYN}}) + K_A C_{\mathrm{AF,SB}})/(B_{\downarrow D} D), \tag{29}$$

where C_{AN} and C_{SYN} are the complexities of the analysis and synthesis filter banks, respectively. The complexity for updating the adaptive filters in a subband channel is $C_{\mathrm{AF,SB}}$ and $K_A \leq K/2$ denotes the number of active subbands. Note that L times as many synthesis steps are required, compared to analysis, due to the L loudspeakers.

1. Complexity example

We now look at an example for computing the complexity assuming a set of filters for the three-layer nested line array as specified in Sec. V A. The assumed anechoic impulse responses have a length of $N_R = 700$ samples and the desired control filters are of length $J = 300$ samples, both at a sampling frequency of 48 kHz. The step-size is determined according to Algorithm 1 in the Appendix with maxItr = 2. For this example, and in the rest of the paper, the bright and dark zones each consist of three microphone positions, covering the ranges $[-25°, -15°]$ and $[15°, 25°]$ of the horizontal directivity, respectively (see Fig. 11 for the resulting directivity pattern). In the complexity analysis, it is assumed that the number of channels in the filter bank is a power of 2, hence, the complexity is evaluated for 4 to 128 subband channels. The prototype filters are iteratively designed following the procedure in Weiss *et al.* (1998a) and the filter banks are implemented according to Weiss (2002). Computational savings can be achieved by taking into account that each of the loudspeaker drivers will only reproduce audio in a subset of the audible frequency range. This is usually ensured by a crossover network bandpass filtering the input signals to each of the loudspeaker drivers. The outcome is that particular subband channels will have close to no input signal due to this crossover network. Therefore, some of the subband channels can be assumed to be zero and do not require computations. The chosen threshold for ignoring subband channels is when the passband edge of the subband analysis filter is 1/3rd octave outside the cutoff frequency of the loudspeaker crossover network. This threshold is chosen relative to the 8th order Linkwitz-Riley crossover filters applied in this work.

The results in Fig. 6 show the computational complexity for oversampling ratios O of 2, 3/2, and 5/4. The filter lengths are chosen according to the oversampling ratios as $4K$, $8K$, and $16K$, respectively. From the results it is observed that there is a benefit in using the subband processing scheme, over the full-rate scheme, when only the relevant subband channels are computed.

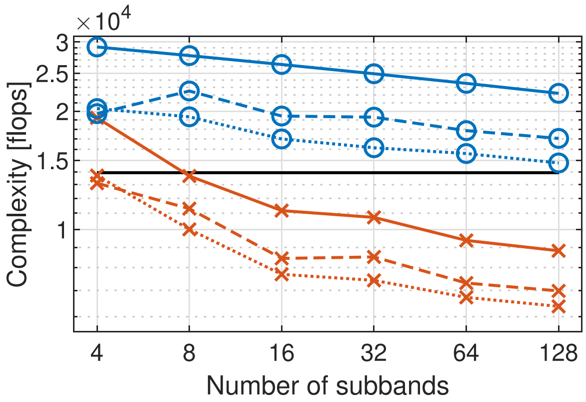


FIG. 6. (Color online) Comparison of the computational complexity in flops per input sample at 48 kHz. (—): Full-rate adaptive filtering. (○): Subband adaptive filtering, all subbands active. (×): Only relevant subbands active. (Full): $O=2$, $N_p=4K$. (Dashed): $O=3/2$, $N_p=8K$. (Dotted): $O=5/4$, $N_p=16K$.

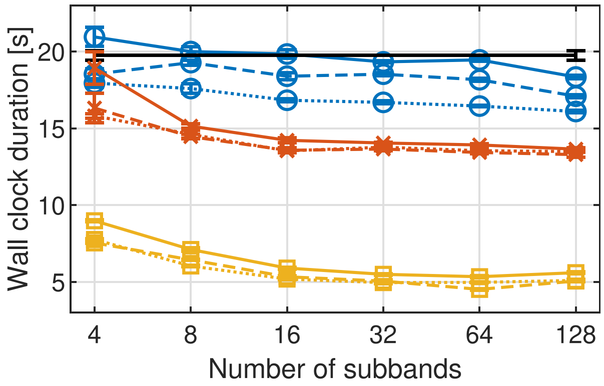


FIG. 7. (Color online) Comparison of the wall clock duration to process 30 s white noise for various subband configurations. Each configuration was repeated 10 times. Error bars indicate 1 standard deviation. (—): Single-threaded, full-rate adaptive filtering. (○): Single-threaded, subband adaptive filtering, all subbands active. (×): Single-threaded, only relevant subbands active. (□): Multithreaded, only relevant subbands active. (Full): $O=2$, $N_p=4K$. (Dashed): $O=3/2$, $N_p=8K$. (Dotted): $O=5/4$, $N_p=16K$.

2. Wall-clock results

To provide further insights in the problem for the theoretical complexity results in Fig. 6, the algorithms have been implemented in C++, without specialized libraries for numerical computations, and compiled to the MATLAB.mex format using GCC 13.1.0. The FFT was implemented as a standard conjugate-pair split-radix FFT as described in, e.g., Johnson and Frigo (2007). For this experiment, 30 s of white noise was processed for each of the subband configurations. The wall clock duration measurements were repeated 10 times and the results include both the analysis and synthesis duration of the subband system along with the processing time in each subband. The results were measured on an HP ZBook Studio G8 laptop with an Intel i9-11950H processor running MATLAB 2022b and WINDOWS 10. The results were measured using both single threaded and multithreaded executions. For the multithreaded results[3] the execution is parallelized across the adaptive filtering updates in individual subbands as well as synthesizing the signals for the individual loudspeaker drivers.

From the results shown in Fig. 7, it is observed that the full-rate implementation is comparable to the subband filtering with all subbands being active and an oversampling ratio of 2. The full-rate solution is slower than the subband implementation when the adaptive filtering is only performed in the relevant subbands. Last, the multithreaded implementation of the subband solution with only active subbands is significantly faster than the other implementations. Only slight improvements are observed for increasing the number of subbands beyond 8. The remainder of the results section will use the example filter bank with $K=8$ subband channels, an oversampling ratio of $O=5/4$, and a prototype filter length of $N_p=128$ samples.

C. Simulated response

1. Tracking performance towards a sudden shift in focus direction

The intended purpose of the adaptive structure is to adapt to changes in the desired room impulse responses, e.g., due to the desired location of the zones moving. As a proxy for this scenario, a simplified scenario is established. After reproducing audio for 15 s, the bright and dark zones exchange location instantaneously. This provides an indication of the speed at which the system can adapt to sudden changes, as would be needed to track moving listeners.

The evaluation metric used for this investigation is the normalized mean square error of the filters relative to the Wiener solution $\mathbf{w}_o$ [as specified in Eq. (A1)],

$$\text{NMSE} = \frac{||\mathbf{w}_{2BL,i} - \mathbf{G}^{10}_{2BL\times JL}\mathbf{w}_o||^2}{||\mathbf{G}^{10}_{2BL\times JL}\mathbf{w}_o||^2}. \quad (30)$$

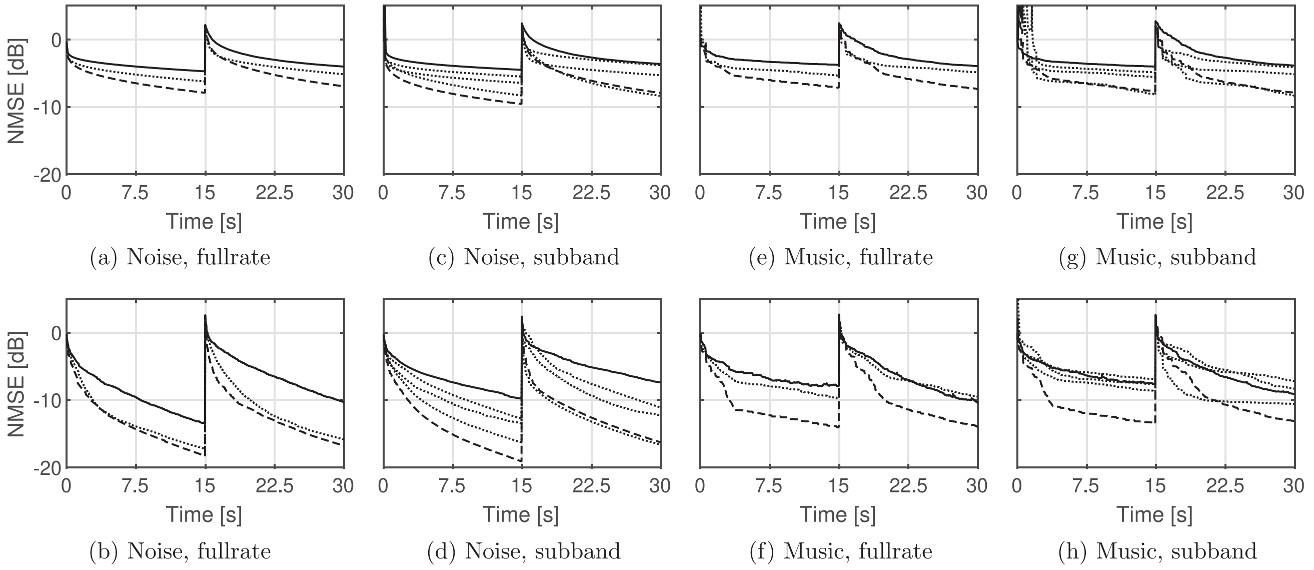
For the case of a varying step-size, the maximal step-size is chosen according to $2/\lambda_{\max}(\mathbb{E}\{\boldsymbol{X}_i\})$ (as defined in Appendix A) where the expectation, denoted by $\mathbb{E}\{\cdot\}$, is calculated from the full audio signal used for the test.[4] The initial step-size is chosen as 1/100 of the maximal step-size. For the scenario without updating the step-size, the initial step-size is kept for the duration of the experiment.

The tracking performance when the input signal is white noise is shown in Figs. 8(a)–8(d). From the white noise results, it is seen that the variable step-size results exhibit faster tracking than the static step-size. It is also observed that the NMSE of the subband filters decrease at a similar rate to the full-rate adaptive filters.

The tracking performance in the case of music[5] is presented in Figs. 8(e)–8(h). The results are similar to the situation with white noise, although the convergence is less smooth. One difference to note is that the changes in the music can cause the variable step-size NMSE to locally increase, due to the short-term average of the spectrum being different from the long-term average used to determine the Wiener solution.

D. Anechoic response

The evaluation of the algorithm is performed using measurements from a large room, which are truncated in



(a) Noise, fullrate (c) Noise, subband (e) Music, fullrate (g) Music, subband

(b) Noise, fullrate (d) Noise, subband (f) Music, fullrate (h) Music, subband

FIG. 8. Normalized mean square deviation from the Wiener filter solution for the full-rate and subband channel filters for 30 s white noise and music, with the bright and dark zones being swapped after 15 s. Top-row: Fixed step-size. Bottom row: Variable step-size. (Solid): Midwoofer channels, (dashed): subtweeter channels, (dotted): tweeter channels.

time to provide anechoic responses. The room is 12 m by 12 m by 12 m, and the loudspeaker array is mounted on a movable platform which makes it possible to measure individual impulse responses from the loudspeaker drivers to microphone positions on a half-circle of radius 3 m centered at the midpoint of the loudspeaker array. The half-circle is sampled at 5° resolution and can be used to evaluate the directivity response of the combined loudspeaker array. The impulse response measurements are performed as the average of two exponentially swept sine measurements (Farina, 2007) from 5 Hz to 24 kHz over a duration of 3 s. The estimated impulse responses are then truncated in time to remove any reflections from the boundaries of the room.

The correspondence between the convergence rate of the adaptive filters and the acoustic separation between the bright and dark zones is evaluated through the time-domain contrast. The time-domain contrast is here defined as the ratio of mean square sound pressure levels for a block of audio samples

$$\text{Contrast}[i] = \frac{M_B^{-1}\sum_{m=1}^{M_B}\sum_{n=1}^{N}||\mathrm{p}_B^{(m)}[Ni+n]||_2^2}{M_D^{-1}\sum_{m=1}^{M_D}\sum_{n=1}^{N}||\mathrm{p}_D^{(m)}[Ni+n]||_2^2}. \qquad (31)$$

In the above, $\mathrm{p}_B^{(m)}$ and $\mathrm{p}_D^{(m)}$ refers to the sound pressures observed at points in the bright and dark zones, respectively. To avoid overfitting in the test, the adaptive filters and loudspeaker input signals are calculated based on point-source simulations in free-field, while the sound pressure is evaluated by convolving the loudspeaker signals with the measured anechoic impulse responses. The scenario where the bright and dark zones suddenly change position, as used in Sec. V C, is repeated here using the same music signal. Due to the temporal variations in the music, the block size is chosen as 1024 samples, corresponding to 21.3 ms at 48 kHz sampling frequency. The results in Fig. 9 depict the contrast between the directivity ranges [−25°,−15°] and [15°,25°], hence, the contrast becomes negative after switching the roles of the bright and dark zones. The results reaffirm that the variable step-size leads to faster convergence than the fixed step-size as observed from the steeper transition between positive and negative contrast. Another observation is that the subband adaptive filtering does not lead to a different contrast than the full-rate solution when an adaptive step-size is used.

Given this similarity between the full rate and subband solution, it is of interest to determine whether there are any spectral differences between the solutions due to some subbands converging to different NMSEs. For this purpose, the algorithms were run with 30 s of white noise as input without changing the location of the zones. The power spectra in

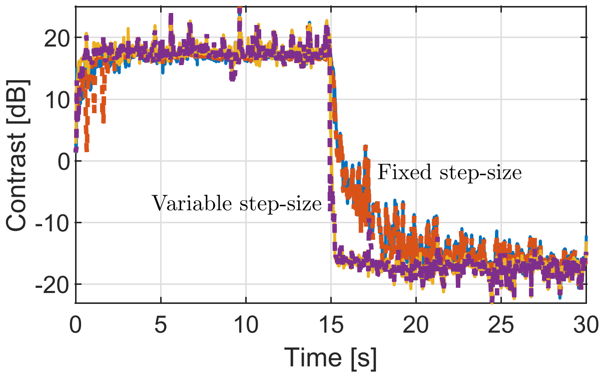


FIG. 9. (Color online) Time-domain contrast. Input signal = music. (—): Full-rate, fixed step-size, (—): Full-rate, variable step-size, (– – –): Subband $O = 5/4$, fixed step-size. (– – –): Subband $O = 5/4$, variable step-size. Note that the bright and dark zone switches position at time = 15 s.

the center of the bright and dark zones were then estimated using Welch averaging with block sizes of 4096 samples, Hanning windows, and no overlap. The resulting sound pressure power spectra were smoothed using a 1/24th octave moving average filter and normalized relative to the average bright zone level in the frequency range between 500 Hz and 8 kHz. The results in Fig. 10, show negligible differences between the full-rate and subband solutions.

E. Correspondence to point source response

To evaluate the spatial response of the loudspeaker array when the adaptive filtering is being used, the directivity response of each loudspeaker driver was measured as described in Sec. V D. The loudspeaker input signals are generated assuming point-source transfer functions and evaluated using either point-source simulations or measured loudspeaker responses. The directivity response is evaluated as the pressure power spectra at 3 m in 5° angular resolution, across 30 s of adaptive filtering given a white noise input signal. In Fig. 11, the results are seen for a bright zone covering the range [−25°,−15°] and a dark zone covering [15°,25°]. As seen from the plots, the point source simulation captures the majority of the response of the loudspeaker array. This is due to the choice and arrangement of loudspeaker drivers combined with the chosen crossover network, i.e., the loudspeaker drivers are almost omnidirectional in the frequency ranges they are used. The outcome of this is that it is sufficient to use the point source responses for controlling the loudspeaker array. This prevents overfitting to the production variations in the individual loudspeaker drivers, when basing the control on measured directivity responses. Note that this sensitivity can be further reduced by increasing the λ parameter in Eq. (21), although that increases the misalignment error. This trade-off is discussed in Fraanje *et al.* (2007) for the case of FXLMS.

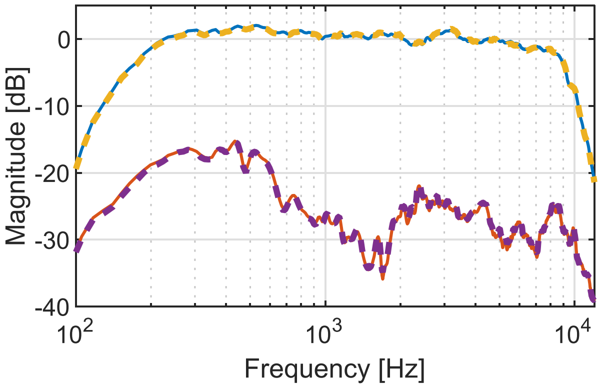


FIG. 10. (Color online) Free field pressure power spectra averaged across 30 s white noise and normalized to the bright zone level in the 500 Hz - 8 kHz range. (—): Full-rate, variable step-size, bright zone, (—): Full-rate, variable step-size, dark zone, (– – –): Subband $O = 5/4$, variable step-size, bright zone. (– – –): Subband $O = 5/4$, variable step-size, dark zone.

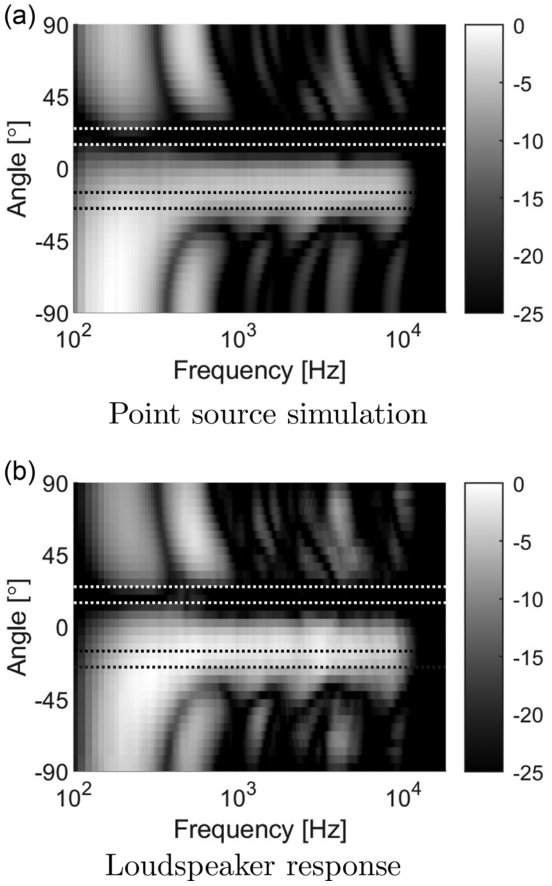


FIG. 11. Directivity response of simulated and implemented loudspeaker arrays, using the adaptive algorithms with white noise input signal while the bright and dark zones cover the ranges $[-15°, -25°]$ and $[15°, 25°]$, respectively (highlighted with dotted lines). The response in dB is normalized relative to maximal value across both angle and frequency. $K = 8$, $O = 5/4$, and $N_p = 16K$.

F. Room response

To validate the behavior of the proposed adaptive procedure in a reflective environment, a series of room impulse response measurements were conducted in a 143 m^3 room with raised ceiling and reverberation time of $T_{20} = 0.53$ s. The measurements were made from each individual loudspeaker driver to two microphone array positions at positions (0.75 m; 2.92 m) and (–0.91 m; 2.87 m) relative to the center of the loudspeaker array and 1.10 m above the floor. The microphone array consisted of a 4 × 3 rectangular array with 10 cm spacing between adjacent microphones. A sketch of the setup is shown in Fig. 12.

The measurements were conducted using exponential sweeps of 8 s duration. For the midwoofers and subtweeters, the sweeps were from 10 Hz to 24 kHz, and for the tweeters the sweeps were from 100 Hz to 24 kHz.

In a scenario where little information is known about the reflective environment, it might be suitable to rely on free field assumptions for controlling the sound field, rather than slightly inaccurate *in situ* measurements (Møller and Olsen, 2019). As seen from the anechoic results, the simulated point source responses are a reasonable approximation

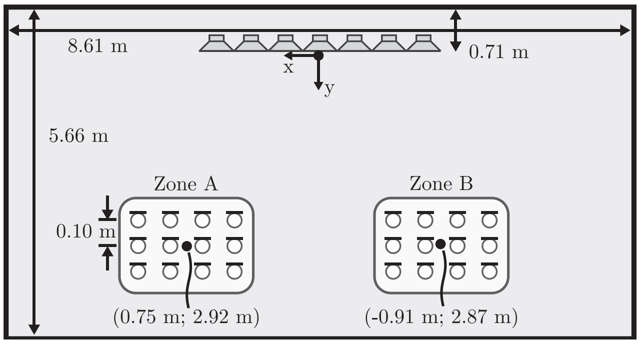


FIG. 12. Sketch of the layout used for the room evaluation measurements.

for the loudspeaker drivers with the given array. As such, the investigated scenario utilizes point source simulations for calculating the loudspeaker signals. The resulting sound field in the room is determined by convolving the loudspeaker signals from the adaptive algorithm with the measured RIRs to the two microphone array locations. For the investigations in this section, the loudspeaker signals are based on point source simulations to three points in the free field representing each zone in the angle ranges [−25°,−15°] and [15°,25°].

To provide an indication of the tracking performance of the system, 30 s of white noise is reproduced by the array and the RIRs for the bright and dark zones are swapped after 15 s. Initially, the bright zone covers the range [−25°,−15°] and the dark zone covers [15°,25°]. From the contrast evaluated across time, plotted in Fig. 13, it is seen that approximately 10 dB of contrast is generated between the zones in the room. This reduction in separation is expected due to the reflective environment (Simón-Gálvez *et al.*, 2014).

To provide insights into the frequency dependence of the separation, the averaged pressure power spectra in the two zones are shown in Fig. 14. Here, the average is taken across all microphones in a zone and 30 s of white noise when the sound is focused towards [−25°,−15°] and suppressed towards [15°,25°]. As seen from the free field directivity in Fig. 11, the given array is hardly directive below 600 Hz. This result is also observed from the small level differences between the zones at low frequencies in Fig. 14. For improved separation between the zones, it would therefore be advantageous to combine the presented solution with subwoofers distributed throughout the room is discussed in, e.g., Druyvesteyn and Garas (1997) and Møller *et al.* (2019).

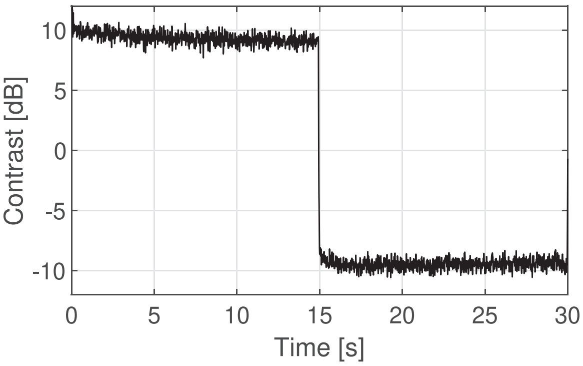


FIG. 13. Contrast between zone A and B in the room sketched in Fig. 12 due to white noise input signal. For the initial 15 s, zone A is the bright zone and zone B is the dark. For the last 15 s, the roles of the zones are swapped.

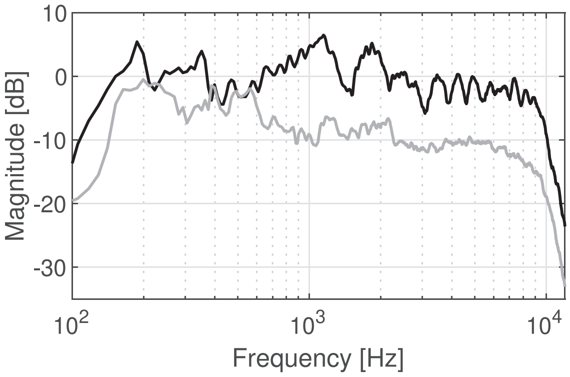


FIG. 14. Sound pressure reproduced in the room sketched in Fig. 12. Power spectra averaged across 12 microphones in each zone and 30 s of white noise. The results are normalized to the bright zone level in the 500 Hz–8 kHz range. (Black line): Bright zone. (Gray line): Dark zone.

VI. DISCUSSION

A. Quality vs prototype filter length

One concern with block-based processing is whether audible artifacts are introduced due to the processing. To give a brief insight into the potential challenges with quality, a simple experiment is set up. Here, the sound pressure reproduced in the bright zone (using the free field point source model), is compared to a reference signal. The reference is the music signal, filtered by the crossover network and recombined to a single signal. This is done to remove spectral differences caused by the crossover network from the quality comparison.

The quality is predicted by the Perceptual Evaluation of Audio Quality (PEAQ) model (ITU-R, 2001), which compares a degraded audio excerpt against a reference and predicts an objective difference grade (ODG) from 0 (imperceptible) to −4 (very annoying) using the MATLAB implementation (Kabal, 2004). The evaluation was performed using $K = 8$ subbands and oversampling ratios of 2, 3/2, and 5/4. The performance is an interplay between the aliasing suppression of the prototype filter design and the accuracy of the decomposition of the RIRs into subband approximations. From Fig. 15, it is observed that the quality increases with the length of the prototype filter, which improves both the suppression of aliasing within the subbands as well as the accuracy of the subband approximation of the RIRs.

B. Latency vs computational complexity

In the present work, the focus has been on improving the computational complexity. As such, the adaptive filtering has been implemented in the frequency domain.

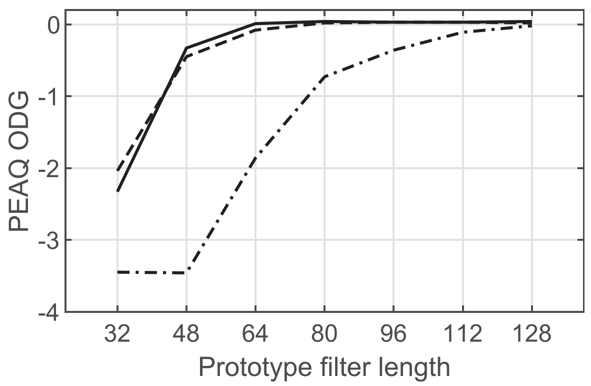


FIG. 15. Comparison of the PEAQ ODG score vs prototype filter length for evaluating 30 s of music, with processing of the relevant subbands out of $K = 8$. (Full): $O = 2$. (dashed): $O = 3/2$. (dashed-dotted): $O = 5/4$.

This incurs an increased latency in the processing chain, proportional to the block sizes (which are given by the linear convolution between the decomposed RIRs and the subband control filters). This latency could be reduced by introducing multi-delay filtering (Moulines *et al.*, 1995; Soo and Pang, 1990) although that comes at the cost of an increased computational complexity. At 48 kHz, the anechoic RIRs are of length 700 and the control filters are 300 taps long. Given the choices for the block size and the 50% overlap between blocks, the block processing introduces 21.3 ms of latency.

Another factor introducing latency is the analysis and synthesis filter banks. The latency of analysis and synthesis step in samples is equal to the length of the prototype filter minus 1 (Weiss *et al.*, 1998b). Thereby, increasing the number of subbands or decreasing the oversampling factor, will increase the latency due to the required steeper slopes of the prototype filter. With the choice of eight subbands and an oversampling factor of 5/4, the filter bank used to generate Figs. 9–14 introduces 127 samples of latency, corresponding to 2.6 ms at 48 kHz sampling frequency.

VII. CONCLUSION

In this work, sound zones were formulated as frequency domain block adaptive filtering. This was evaluated for both full-rate and subband processing (without cross-adaptive filters). It was seen that the benefit of the subband approach relies on recognizing that specific subbands do not require processing due to the inherent frequency range limitations imposed by both loudspeaker array design and the properties of loudspeaker drivers.

The reduced computational complexity was seen to be attained without reducing the accuracy of the reproduced sound field. Furthermore, it was seen that the subband formulation naturally lends itself well to parallel computations leading to faster execution times.

Future work should consider potential computational complexity reductions in algorithms for choosing and updating the step-size of the adaptive algorithm.

ACKNOWLEDGMENTS

This work was partly funded by the Innovation Fund Denmark (IFD) under File No 9069-00038B in the project Interactive Sound Zones for Better Living (ISOBEL).

AUTHOR DECLARATIONS

Conflict of Interest

The authors declare that they have no conflict of interest to disclose.

DATA AVAILABILITY

The data that support the findings of this study are available on github (Møller, 2024).

APPENDIX A: STEP SIZE CONDITIONS FOR STABILITY

The conditions for mean square stability can be determined following the approach of energy conservation as described in Sayed (2008) and Sayed and Al-Naffouri (2001). Here, the approach is modified for the sound zone application with given room impulse responses.

Start by defining the Wiener solution as

$$\mathbf{w}_o : = \underset{\mathbf{w}}{\operatorname{argmin}}\ \mathbb{E}\left\{ \left\| \begin{bmatrix} (\mathbf{I}_L \otimes \mathbf{G}^{01}_{B\times 2B} \boldsymbol{U}_i)\mathbf{r}_t \\ \mathbf{0} \end{bmatrix} \right.\right. \\ \left.\left. - \begin{bmatrix} (\mathbf{I}_M \otimes \mathbf{G}^{01}_{B\times 2B} \boldsymbol{U}_i)\mathbf{R}(\mathbf{I}_L \otimes \mathbf{G}^{10}_{2B\times J}) \\ \sqrt{\lambda}(\mathbf{I}_L \otimes \mathbf{G}^{10}_{2B\times J}) \end{bmatrix} \mathbf{w} \right\|^2 \right\}, \tag{A1}$$

where

$$\boldsymbol{R} := \begin{bmatrix} \boldsymbol{R}_{1,1} & \cdots & \boldsymbol{R}_{1,L} \\ \vdots & \ddots & \vdots \\ \boldsymbol{R}_{M,1} & \cdots & \boldsymbol{R}_{M,L} \end{bmatrix}, \quad \boldsymbol{r}_t := \begin{bmatrix} \boldsymbol{R}_{1,\ell_t} \boldsymbol{w}^{(1,\ell_t)} \\ \vdots \\ \boldsymbol{R}_{M,\ell_t} \boldsymbol{w}^{(M,\ell_t)} \end{bmatrix}. \tag{A2}$$

Introducing $\mathbf{G} := (\mathbf{I}_L \otimes \mathbf{G}^{10}_{2B\times J}), \boldsymbol{S}_i := (\mathbf{I}_M \otimes \boldsymbol{\Lambda}^{-1} \boldsymbol{U}_i^H \mathbf{G}^{01}_{2B\times 2B} \boldsymbol{U}_i)$, the weight update equation is written as

$$\boldsymbol{w}_{i+1} = (1 - \mu\lambda)\boldsymbol{w}_i + \mu \mathbf{G}^{\mathsf{H}} \boldsymbol{R}^{\mathsf{H}} \boldsymbol{S}_i (\boldsymbol{r}_t - \boldsymbol{R}\mathbf{G}\boldsymbol{w}_i). \tag{A3}$$

The weight error is introduced as $\tilde{\boldsymbol{w}}_i := \boldsymbol{w}_o - \boldsymbol{w}_i$. Both sides of Eq. (A3) can be subtracted from the Wiener solution, which makes it possible to write

$$\tilde{\boldsymbol{w}}_{i+1} = (\mathbf{I}_{JL} - \mu \boldsymbol{X}_i)\tilde{\boldsymbol{w}}_i + \mu \boldsymbol{\xi}_i \tag{A4}$$

with $\boldsymbol{X}_i := \lambda \mathbf{I}_{JL} + \mathbf{G}^{\mathsf{H}} \boldsymbol{R}^{\mathsf{H}} \boldsymbol{S}_i \boldsymbol{R}\mathbf{G}$ and $\boldsymbol{\xi}_i := \lambda \boldsymbol{w}_o - \mathbf{G}^{\mathsf{H}} \boldsymbol{R}^{\mathsf{H}} \boldsymbol{S}_i (\boldsymbol{r}_t - \boldsymbol{R}\mathbf{G}\boldsymbol{w}_o)$.

Following the steps of Sayed and Al-Naffouri (2001) and Chap. 24 of Sayed (2008), the conditions for mean square stability is that the step size is chosen as

$$\mu < \min\left\{\frac{2}{\lambda_{\max}(\mathbb{E}\{\boldsymbol{X}_i\})}, \frac{1}{\lambda_{\max}(\boldsymbol{A}^{-1}\boldsymbol{B})}, \frac{1}{\lambda_{\mathbb{R}}(\boldsymbol{H})}\right\}, \tag{A5}$$

where $\lambda_{\max}(\cdot)$ denotes the maximal eigenvalue of a matrix and $\lambda_{\mathbb{R}}(\cdot)$ denotes the maximal positive real eigenvalue of a matrix. The matrices $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{H}$ are given as

$$\begin{aligned}\boldsymbol{A} &:= \mathbb{E}\{\boldsymbol{X}_i^{\mathsf{T}} \otimes \mathbf{I}_{JL} + \mathbf{I}_{JL} \otimes \boldsymbol{X}_i^{\mathsf{H}}\},\\ \boldsymbol{B} &:= \mathbb{E}\{\boldsymbol{X}_i^{\mathsf{T}} \otimes \boldsymbol{X}_i^{\mathsf{H}}\},\\ \boldsymbol{H} &:= \begin{bmatrix}\boldsymbol{A}/2 & -\boldsymbol{B}/2\\ \mathbf{I}_{(JL)^2} & \mathbf{0}_{(JL)^2}\end{bmatrix}.\end{aligned} \tag{A6}$$

This result relies on the independence assumption, i.e., that the input sequence vector $\{[\mathrm{u}[iB-B],\ldots,\mathrm{u}[iB+B-1]]^{\mathsf{T}}\}$ is independent and identically distributed, without necessarily following a Gaussian distribution.

APPENDIX B: BACKTRACKING LINE SEARCH

One way of determining the step size α is by performing a backtracking line search. For this purpose, we describe the current filter and the updated filter as

$$\boldsymbol{w}_{2B,i+1} = \boldsymbol{w}_{2B,i} - \alpha_i \nabla J_{2B}(\boldsymbol{w}_i). \tag{B1}$$

We can then perform a backtracking line search to determine $\alpha_i \in \mathbb{R}_+$ to provide sufficient reduction of the cost-function by satisfying the Armijo condition (Nocedal and Wright, 2006)

$$J(\boldsymbol{w}_i - \alpha_i \nabla J_{2B}(\boldsymbol{w}_i)) \leq J(\boldsymbol{w}_i) - c_1 \alpha_i ||\nabla J_{2B}(\boldsymbol{w}_i)||_2^2, \tag{B2}$$

where $c_1 \in (0,1)$. One option to speed-up the convergence is to initialize the line search with the previous step size $\alpha_i = \alpha$. However, this approach introduces the risk of getting stuck at a very small step size. Therefore, if the Armijo condition is already satisfied with the initial step size, we can modify the line search to search for the largest step size satisfying the sufficient decrease condition. To ensure the convergence in the mean, the maximal step size can be determined as suggested in Appendix A. Such a line search is described in Algorithm 1. Note that other, more computationally efficient, algorithms for adjusting the step-size might exist.

ALGORITHM 1. Backtracking line search dual direction.

```
$\rho \in (0,1), c_1 \in (0,1)$; Set $\alpha \leftarrow \alpha_{i-1}$, itr $\leftarrow 0$
if $J(\boldsymbol{w}_i - \alpha\nabla J_{2B}(\boldsymbol{w}_i)) \leq J(\boldsymbol{w}_i) - c_1\alpha||\nabla J_{2B}(\boldsymbol{w}_i)||_2^2$ then
    while $J(\boldsymbol{w}_i - \alpha\nabla J_{2B}(\boldsymbol{w}_i)) \leq J(\boldsymbol{w}_i) - c_1\alpha||\nabla J_{2B}(\boldsymbol{w}_i)||_2^2$ and itr $<$ maxItr
    do
        itr $++$;
        $\alpha \leftarrow \frac{\alpha}{\rho}$;
    end while
    $\alpha \leftarrow \rho\alpha_i$;
    if $\alpha > \frac{2}{\lambda_{\max}}$ then
        $\alpha \leftarrow \frac{2}{\lambda_{\max}}$
    end if
else
    while $J(\boldsymbol{w}_i - \alpha\nabla J(\boldsymbol{w}_i)) > J_{2B}(\boldsymbol{w}_i) - c_1\alpha||\nabla J_{2B}(\boldsymbol{w}_i)||_2^2$ and itr $<$ maxItr
    do
        itr $++$;
        $\alpha \leftarrow \rho\alpha$;
    end while
end if
Terminate with $\alpha_i \leftarrow \alpha$.
```

[1] Although Sec. V only considers anechoic conditions for the calculation of the filters, we will use the terms room transfer functions and room impulse responses to highlight the relation to controlling sound fields in reflective environments.

[2] In the present work, the target loudspeaker is chosen as the center loudspeaker of the line array.

[3] The multithreading was implemented using OPENMP.

[4] The reason for not choosing the maximal step-size according to Eq. (A5) is the impracticality of evaluating eigenvalues of matrices of the size $(JL)^2 \times (JL)^2$. In the given example $J = 300$ and L is between 7 and 11, i.e., the matrices would be at least 4.41×10^6 by 4.41×10^6.

[5] Daft Punk—Give life back to music, from 00:31 to 01:01.

Ahrens, J., and Spors, S. (**2010**). "Sound field reproduction using planar and linear arrays of loudspeakers," IEEE Trans. Audio Speech Lang. Process. **18**(8), 2038–2050.

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

Brandwood, D. H. (**1983**). "A complex gradient operator and its application in adaptive array theory," IEE Proc. H Microwave Opt. Antennas. **130**(1), 11–16.

Buchner, H., Benesty, J., and Kellermann, W. (**2005**). "Generalized multichannel frequency-domain adaptive filtering: Efficient realization and application to hands-free speech communication," Signal Process. **85**, 549–570.

Caviedes-Nozal, D., Riis, N. A. B., Heuchel, F. M., Brunskog, J., Gerstoft, P., and Fernandez-Grande, E. (**2021**). "Gaussian processes for sound field reconstruction," J. Acoust. Soc. Am. **149**(2), 1107–1119.

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

Cheer, J., Elliott, S., and Gálvez, M. F. S. (**2013**). "Design and implementation of a car cabin personal audio system," J. Audio Eng. Soc. **61**(6), 412–424, available at https://www.aes.org/e-lib/browse.cfm?elib=16832.

Druyvesteyn, W. F., and Garas, J. (**1997**). "Personal sound," J. Audio. Eng. Soc. **45**(9), 685–701, available at https://www.aes.org/e-lib/browse.cfm?elib=7843.

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

Farina, A. (**2007**). "Advancements in impulse response measurements by sine sweeps," in *Audio Engineering Society Convention 122*, Paper No. 7121.

Fraanje, R., Elliott, S. J., and Verhaegen, M. (**2007**). "Robustness of the filtered-X LMS algorithm—Part II: Robustness enhancement by minimal regularization for norm bounded uncertainty," IEEE Trans. Signal Process. **55**(8), 4038–4047.

Gálvez, M. F. S., Elliott, S. J., and Cheer, J. (**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.

Gilloire, A., and Vetterli, M. (**1992**). "Adaptive filtering in subbands with critical sampling: Anaylsis, experiments, and application to acoustic echo cancellation," IEEE Trans. Signal Process. **40**(8), 1862–1875.

Harteneck, M., Paez-Borrallo, J. M., and Stewart, R. W. (**1998**). "An oversampled subband adaptive filter without cross adaptive filters," Signal Process. **64**, 93–101.

Hu, M., Shi, L., Zou, H., Christensen, M. G., and Lu, J. (**2023**). "Sound zone control with fixed acoustic contrast and simultaneous tracking of acoustic transfer functions," J. Acoust. Soc. Am. **153**(5), 2538–2544.

ITU-R (**2001**). Rec. ITU-R BS.1387-1: "Method for objective measurement of perceived audio quality," Standard.

Jacobsen, R. M., Fangel Skov, K., Johansen, S. S., Skov, M. B., and Kjeldskov, J. (**2023**). "Living with sound zones: A long-term field study of dynamic sound zones in a domestic context," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23.

Jacobsen, R. M., Johansen, S. S., van Berkel, N., Skov, M. B., and Kjeldskov, J. (**2022**). "In the zone!—Controlling and visualising sound zones," in *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI EA '22 (Association for Computing Machinery, New York).

Jin, W., and Kleijn, W. B. (**2015**). "Theory and design of multizone soundfield reproduction using sparse methods," IEEE/ACM Trans. Audio Speech Lang. Process. **23**(12), 2343–2355.

Johnson, S. G., and Frigo, M. (**2007**). "A modified split-radix FFT with fewer arithmetic operations," IEEE Trans. Signal Process. **55**(1), 111–119.

Kabal, P. (**2004**). "Pqevalaudio 1.0 matlab toolbox," https://www.mmsp.ece.mcgill.ca/Documents/Software/index.html (Last viewed March 20, 2024).

Kellermann, W. (**1988**). "Analysis and design of multirate systems for cancellation of acoustical echoes," in *ICASSP-88, International Conference on Acoustics, Speech, and Signal Processing*, Vol. 5, pp. 2570–2573.

Lluís, F., Martínez-Nuevo, P., Bo Møller, M., and Ewan Shepstone, S. (**2020**). "Sound field reconstruction in rooms: Inpainting meets super-resolution," J. Acoust. Soc. Am. **148**(2), 649–659.

Moles-Cases, V., Pinero, G., de Diego, M., and Gonzalez, A. (**2020**). "Personal sound zones by subband filtering and time domain optimization," IEEE/ACM Trans. Audio. Speech. Lang. Process. **28**, 2684–2696.

Møller, M. B. (**2024**). "sz-subband-block-adaptive-filtering," github.com/macoustics/sz-subband-block-adaptive-filtering (Last viewed March 20, 2024).

Møller, M. B., Nielsen, J. K., Fernandez-Grande, E., and Olesen, S. K. (**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.

Møller, M. B., and Olsen, M. (**2019**). "On *in situ* beamforming in an automotive cabin using a planar loudspeaker array," in *Proceedings of the 23rd International Congress on Acoustics*, Aachen, Germany, pp. 1109–1116.

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

Moulines, E., Amrane, O. A., and Grenier, Y. (**1995**). "The generalized multidelay adaptive filter: Structure and convergence analysis," IEEE Trans. Signal Process. **43**(1), 14–28.

Nelson, P. A., Hamada, H., and Elliott, S. J. (**1992**). "Adaptive inverse filters for stereophonic sound reproduction," IEEE Trans. Signal Process. **40**(7), 1621–1632.

Nocedal, J., and Wright, S. J. (**2006**). *Numerical Optimization*, 2nd ed. (Springer, Berlin).

Olsen, M., and Møller, M. B. (**2017**). "Sound zones: On the effect of ambient temperature variations in feed-forward systems," in *Audio Engineering Society Convention 142*.

Pham Vu, T., and Lissek, H. (**2020**). "Low frequency sound field reconstruction in a non-rectangular room using a small number of microphones," Acta Acust. **4**(2), 5.

Poletti, M. (**2008**). "An investigation of 2-D multizone surround sound systems," in *Audio Engineering Society Convention 125*.

Radmanesh, N., Burnett, I. S., and Rao, B. D. (**2016**). "A lasso-LS optimization with a frequency variable dictionary in a multizone sound system," IEEE/ACM Trans. Audio. Speech. Lang. Process. **24**(3), 583–593.

Reilly, J. P., Wilbur, M., Seibert, M., and Ahmadvand, N. (**2002**). "The complex subband decomposition and its application to the decimation of large adaptive filtering problems," IEEE Trans. Signal Process. **50**(11), 2730–2743.

Sayed, A., and Al-Naffouri, T. (**2001**). "Mean-square analysis of normalized leaky adaptive filters," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*, Vol. 6, pp. 3873–3876.

Sayed, A. H. (**2008**). *Adaptive Filters* (Wiley, New York).

Shin, M., Lee, S. Q., Fazi, F. M., Nelson, P. A., Kim, D., Wang, S., Park, K. H., and Seo, J. (**2010**). "Maximization of acoustic energy difference between two spaces," J. Acoust. Soc. Am. **128**(1), 121–131.

Simón-Gálvez, M. F., Elliott, S. J., and Cheer, J. (**2014**). "The effect of reverberation on personal audio devices," J. Acoust. Soc. Am. **135**(5), 2654–2663.

Soo, J.-S., and Pang, K. K. (**1990**). "Multidelay block frequency domain adaptive filter," IEEE Trans. Acoust. Speech Signal Process. **38**(2), 373–376.

Vindrola, L., Melon, M., Chamard, J.-C., and Gazengel, B. (**2021**). "Use of the filtered-x least-mean-squares algorithm to adapt personal sound zones in a car cabin," J. Acoust. Soc. Am. **150**(3), 1779–1793.

Weiss, S. (**2002**). "Analysis and fast implementation of oversampled modulated filter banks," in *Mathematics in Signal Processing V*, edited by J. G. McWhirter and I. K. Proudler (Oxford University Press, Oxford), Chap. 23, pp. 263–274.

Weiss, S., Harteneck, M., and Stewart, R. W. (**1998a**). "On implementation and design of filter banks for subband adaptive systems," in *1998 IEEE Workshop on Signal Processing Systems. SIPS 98. Design and Implementation*, pp. 172–181.

Weiss, S., Lampe, L., and Stewart, R. W. (**1998b**). "Efficient implementations of complex and real valued filter banks for comparative subband processing with an application to adaptive filtering," in *1st International Symposium on Communication Systems and Digital Signal Processing*.

Yang, F., Enzner, G., and Yang, J. (**2019**). "A unified approach to statistical convergence analysis of frequency-domain adaptive filters," IEEE Trans. Signal Process. **67**(7), 1785–1796.

Zhao, S., and Burnett, I. S. (**2022**). "Adaptive personal sound zones systems with online plant modelling," in *Proceedings of the 24th International Congress on Acoustics*.