KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Authors: Jesper Brunnström, Shoichi Koyama, Marc Moonen Title: Variable span trade-off filter for sound zone control with kernel interpolation weighting Journal/Conference/Book: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) Vol/Book Chapter: Pages: 1071-1075 ISBN: 978-1-6654-0540-9
Dates **Archived version**	**Conference paper/Abstract** • Date of conference: May 23-27, 2022 ☐ Submitted version ■ Accepted (Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher) ☐ Published ☐ Published (Open Access)
Acknowledgements	In this manuscript, the following funding is acknowledged: ■ EU H2020 ☐ EU Horizon Europe ☐ FWO ■ KULeuven Internal Funds ■ VLAIO ☐ AI ☐ Other
Published version	https://ieeexplore.ieee.org/document/9746550
Copyright statement	
Author contact	jesper.brunnstroem@kuleuven.be

VARIABLE SPAN TRADE-OFF FILTER FOR SOUND ZONE CONTROL WITH KERNEL INTERPOLATION WEIGHTING

Jesper Brunnström[⋆] *Shoichi Koyama*[†] *Marc Moonen*[⋆]

[⋆]STADIUS Center for Dynamical Systems, Department of Electrical Engineering (ESAT),
KU Leuven, Leuven 3001, Belgium
[†] Graduate School of Information Science and Technology,
the University of Tokyo, Tokyo 113-8656, Japan

ABSTRACT

A sound zone control method is proposed, based on the frequency domain variable span trade-off filter (VAST). Existing VAST methods optimizes the sound field at a set of discrete points, while the proposed method uses kernel interpolation to instead optimize the sound field over a continuous region. When the loudspeaker positions are known, the performance can be improved further by applying a directional weighting to the interpolation procedure. The proposed method is evaluated by simulating broadband sound in a reverberant environment, focusing on the case when microphone placement is restricted. The proposed method with directional weighting outperforms the pointwise VAST over the full bandwidth of the signal, and the proposed method without directional weighting outperforms the pointwise VAST at low frequencies.

Index Terms— Sound zone control, sound field control, spatial audio, kernel interpolation, variable span trade-off filter

1. INTRODUCTION

Sound zone control was introduced in [1], the idea of which is to reproduce individual sounds in multiple zones with a set of loudspeakers. The desired sound in each zone should be reproduced clearly, while creating little disturbance in the other zones. Different approaches have been considered, examples including beamforming [2] and multizone surround sound [3].

The optimization approach to sound zone control has largely been dominated by the two techniques acoustic contrast control (ACC) and pressure matching (PM). Proposed in [4], further developed in [5–9], in ACC the ratio between the sound power in the bright and dark zones is maximized. In PM, the mean square error between the desired and reproduced sound is minimized [10–12].

It was later shown in [13] that both PM and ACC can be obtained as extreme cases of a more general framework, dubbed variable span trade-off filter (VAST). Discovered through earlier work in speech enhancement [14, 15], VAST can be derived as a low-rank approximation of a weighted PM problem. The framework has been extended for perceptual optimization in [16, 17], more computationally efficient implementations in [18, 19], and frequency domain formulations in [20].

In conventional sound zone control, it is generally assumed that room impulse responses (RIRs) between the loudspeakers and a set of points distributed over the zones are measured prior to operation. In some cases this is difficult or not at all possible, such as when the loudspeaker system is unknown or moving, the acoustic environment changes periodically, or the listener is moving over too large spaces. Without pre-measured RIRs, microphones must be placed in the room during operation, considerably limiting the placement options, as the listener should be prioritized. In order to retain satisfactory performance, sound field interpolation can be applied to control the sound over a continuous region instead of only at the microphone positions.

Sound field interpolation based on kernel ridge regression [21, 22], referred to as kernel interpolation, has been shown to be an effective tool for spatial active noise control [23–25], binaural rendering [26], and sound field reproduction [27]. By interpolating the sound field from discrete measurement points, it becomes possible to control the sound field over a continuous region. Kernel interpolation can be applied to arbitrarily spaced arrays, requiring only the relative positions of the microphones to be known, making the method applicable in a wide variety of circumstances.

The main contribution of this paper is an extension of the VAST framework, using kernel interpolation to take the sound field over the continuous sound zones into account. The positions of the loudspeakers are generally known in sound zone control, which allows for improved interpolation performance by utilizing the directional weighting presented in [28]. The proposed approach is compatible with other works extending the VAST framework, making it possible to combine the proposed method with other methods in the literature.

2. BACKGROUND

2.1. Problem statement

Consider a multizone scenario where individual signals are to be reproduced in any number of zones. The problem can be solved with superposition, giving instead a scenario with a single bright zone where a desired sound is to be reproduced clearly, and an arbitrary number of dark zones where it should be quiet. The bright zone is denoted by Ω_b, and each of the dark zones are denoted by Ω_d for $d \in \mathcal{D}$, where $\mathcal{D}$ is an ordered index set containing the indices of the dark zones. There are M_k microphones associated with the zone k, each indexed by the set $\mathcal{M}_k$. These index sets are allowed to

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement No. 956369: "Service-Oriented Ubiquitous Network-Driven Sound — SOUNDS", KU Leuven Research Council project C14-21-0075 "A holistic approach to the design of integrated and distributed digital signal processing algorithms for audio and speech communication devices", and VLAIO O&O Project nr. HBC.2020.2197 "SPIC: Signal Processing and Integrated circuits for Communications".

overlap, meaning that the same microphone can be used for multiple zones. There are L loudspeakers, indexed by $\mathcal{L}$.

The signals in the following text are frequency domain signals except where mentioned, but the frequency dependence is implicit for notational simplicity. The input signal that should be reproduced in the bright zone is denoted by $x \in \mathbb{C}$, which is the value for a single frequency bin. The sound pressure $\boldsymbol{p}_k \in \mathbb{C}^{M_k}$ for zone k is measured at the associated M_k microphones, and can be expressed as the sum of the contributions from each loudspeaker as

$$\boldsymbol{p}_k = \sum_{l \in \mathcal{L}} \boldsymbol{h}_{kl} w_l x, \tag{1}$$

where $\boldsymbol{h}_{kl} \in \mathbb{C}^{M_k}$ is the transfer function from loudspeaker l to the microphones, and $w_l \in \mathbb{C}$ is the control filter which is to be optimized. The control filters for all loudspeakers are collected in a vector $\boldsymbol{w} = \text{col}\{w_l\}_{l \in \mathcal{L}} \in \mathbb{C}^L$, where $\text{col}\{\cdot\}$ concatenates its arguments along the column axis, viewing a scalar as a length-one vector. The transfer functions and relative positions of the microphones are assumed to be known.

The signal that should be reproduced at the bright zone microphones is the desired signal $\boldsymbol{d} \in \mathbb{C}^{M_b}$. To ensure the desired signal is physically feasible, it is specified as the input signal x reproduced by V virtual sources, the latter of which are indexed by $\mathcal{V}$. It can be expressed in terms of the contribution from each virtual source as

$$\boldsymbol{d} = \sum_{v \in \mathcal{V}} \tilde{\boldsymbol{h}}_v x, \tag{2}$$

where $\tilde{\boldsymbol{h}}_v \in \mathbb{C}^{M_b}$ is the transfer function from the virtual source v to the microphones associated with the bright zone. The position and associated RIRs of the virtual sources can be chosen freely. A practical choice is to select one of the physical loudspeakers, possibly with truncated RIRs to only keep the direct component.

2.2. Kernel interpolation

To estimate the sound pressure function $p(\boldsymbol{r})$ at an arbitrary point $\boldsymbol{r}$, kernel interpolation can be applied [21]. The estimate is obtained in closed form, from the reproducing kernel Hilbert space $\mathcal{H}$, which itself is completely determined by the choice of kernel function $\kappa(\cdot)$. It is important for sound field interpolation to not just choose any kernel function, but one that is likely to give a physically feasible solution. The space $\mathcal{H}$ can be restricted to functions satisfying Helmholtz equation, while also imposing a priori knowledge about the direction of the wave field, by using the kernel function [28]

$$\begin{aligned} \kappa(\boldsymbol{r}, \boldsymbol{r}', \boldsymbol{\theta}) &= j_0\Big(\sqrt{\boldsymbol{\xi}^\top \boldsymbol{\xi}}\Big), \\ \boldsymbol{\xi} &= \mathrm{j}\rho\boldsymbol{\theta} - \frac{\omega}{c}(\boldsymbol{r} - \boldsymbol{r}'). \end{aligned} \tag{3}$$

The vector $\boldsymbol{\theta} \in \mathbb{R}^3$ is a unit vector describing the assumed arrival direction of the wave field, $\rho \in \mathbb{R}_{\geq 0}$ is a parameter selecting the spread of the directional weighting, $\mathrm{j} = \sqrt{-1}$ is the imaginary unit, $j_0(\cdot)$ is the 0-th order spherical Bessel function of the first kind, ω is the angular frequency, and c is the speed of sound.

The sound pressure function estimate is obtained from the sound pressure at the microphone positions $\boldsymbol{p}_k$ by applying the interpolation filter $\boldsymbol{z}_k(\boldsymbol{r}, \boldsymbol{\theta}) \in \mathbb{C}^{M_k}$ as

$$p(\boldsymbol{r}) = \boldsymbol{z}_k^\top(\boldsymbol{r}, \boldsymbol{\theta}) \boldsymbol{p}_k \qquad \boldsymbol{r} \in \Omega_k. \tag{4}$$

The interpolation filter is defined as

$$\boldsymbol{z}_k(\boldsymbol{r}, \boldsymbol{\theta}) = (\boldsymbol{K}_k(\boldsymbol{\theta}) + \lambda \boldsymbol{I})^{-\top} \boldsymbol{\kappa}_k(\boldsymbol{r}, \boldsymbol{\theta}), \tag{5}$$

where $\lambda \in \mathbb{R}_{\geq 0}$ is a regularization parameter, and $\boldsymbol{I}$ is the identity matrix of appropriate size. The kernel vector is defined as $\boldsymbol{\kappa}_k(\boldsymbol{r}, \boldsymbol{\theta}) = \text{col}\{\kappa(\boldsymbol{r}, \boldsymbol{r}_m, \boldsymbol{\theta})\}_{m \in \mathcal{M}_k} \in \mathbb{C}^{M_k}$, and the Gram matrix is defined as $\boldsymbol{K}_k(\boldsymbol{\theta}) = \text{col}\{\boldsymbol{\kappa}_k^\top(\boldsymbol{r}_m, \boldsymbol{\theta})\}_{m \in \mathcal{M}_k} \in \mathbb{C}^{M_k \times M_k}$.

3. VARIABLE-SPAN TRADE-OFF FILTER WITH INTERPOLATION WEIGHTING

3.1. Cost function

In the bright zone the goal is to minimize the distortion, i.e. the difference between the reproduced and desired sound, giving a cost function of the form

$$\mathcal{J}_b = \frac{1}{|\Omega_b|} \int_{\Omega_b} |p(\boldsymbol{r}) - d(\boldsymbol{r})|^2 \, d\boldsymbol{r}, \tag{6}$$

where $|\Omega_b|$ is the volume of the bright zone. For the dark zones, the goal is the same, but the desired sound is zero, giving the cost function for all dark zones

$$\mathcal{J}_d = \frac{1}{\sum_{d \in \mathcal{D}} |\Omega_d|} \sum_{d \in \mathcal{D}} \int_{\Omega_d} |p(\boldsymbol{r})|^2 \, d\boldsymbol{r}. \tag{7}$$

The total cost function to be minimized is the weighted sum

$$\mathcal{J} = \mathcal{J}_b + \mu \mathcal{J}_d, \tag{8}$$

where the parameter $\mu \in \mathbb{R}_{\geq 0}$ controls the relative effort of minimizing distortion in the bright zone as opposed to minimizing the residual sound in the dark zones.

3.2. Estimate sound pressure and desired signal

To obtain $p(\boldsymbol{r})$ and $d(\boldsymbol{r})$ for all $\boldsymbol{r}$, the interpolation procedure presented in section 2.2 is applied. Under the assumption that the loudspeaker positions are known, the direction of the direct sound component emitted by each loudspeaker is known. The directionally weighted kernel in (3) can thus be used to define an interpolation filter (5) for each loudspeaker and zone combination. The suggested direction is from the center of the zone $\boldsymbol{r}_k$ to the loudspeaker position $\boldsymbol{r}_l$, defined as

$$\boldsymbol{\theta}_{lk} = \frac{\boldsymbol{r}_l - \boldsymbol{r}_k}{\|\boldsymbol{r}_l - \boldsymbol{r}_k\|}. \tag{9}$$

The ideal scenario for this choice of $\boldsymbol{\theta}_{lk}$ is that zone k is small and the associated microphones in $\mathcal{M}_k$ are placed close to the zone, as compared to the distance to loudspeaker l. If more knowledge about the application is available, the direction $\boldsymbol{\theta}_{lk}$ can be adjusted, and the set $\mathcal{M}_k$ can exclude or include any microphones such that the direction of the wave field at the microphones is similar enough to $\boldsymbol{\theta}_{lk}$.

The interpolation filter associated with each loudspeaker can be stacked into a vector $\boldsymbol{z}_k(\boldsymbol{r}) = \text{col}\{\boldsymbol{z}(\boldsymbol{r}, \boldsymbol{\theta}_{lk})\}_{l \in \mathcal{L}} \in \mathbb{C}^{LM_k}$. The transfer functions are collected into a matrix $\boldsymbol{H}_k = \text{blkdiag}\{\boldsymbol{h}_{kl}\}_{l \in \mathcal{L}} \in \mathbb{C}^{LM_k \times L}$, where $\text{blkdiag}\{\cdot\}$ forms a block diagonal matrix from its arguments. The estimated sound pressure can then be expressed

$$p(\boldsymbol{r}) = \boldsymbol{z}_k^\top(\boldsymbol{r}) \boldsymbol{H}_k \boldsymbol{w} x \quad \boldsymbol{r} \in \Omega_k. \tag{10}$$

For the bright zone, $d(\boldsymbol{r})$ can be estimated in a similar way. The interpolation filters and transfer functions associated with the virtual sources are stacked as $\tilde{\boldsymbol{z}}_b(\boldsymbol{r}) = \text{col}\{\boldsymbol{z}_b(\boldsymbol{r}, \boldsymbol{\theta}_{vb})\}_{v \in \mathcal{V}} \in \mathbb{C}^{VM_b}$ and $\tilde{\boldsymbol{h}} = \text{col}\{\tilde{\boldsymbol{h}}_v\}_{v \in \mathcal{V}} \in \mathbb{C}^{VM_b}$. The desired signal for a given position can then be estimated as

$$d(\boldsymbol{r}) = \tilde{\boldsymbol{z}}_b^\top(\boldsymbol{r}) \tilde{\boldsymbol{h}} x \quad \boldsymbol{r} \in \Omega_b. \tag{11}$$

3.3. Optimal control filter

The resulting cost function is obtained by inserting the expressions (10) and (11) into the cost function (8). To obtain the optimal control filter, the gradient with respect to the control filter $\boldsymbol{w}$ is computed, and set equal to zero. The obtained optimal control filter is

$$\boldsymbol{w} = (\boldsymbol{R}_b + \mu \boldsymbol{R}_d)^{-1} \boldsymbol{r}_b, \tag{12}$$

where the spatial correlation matrices are defined as

$$\begin{aligned} \boldsymbol{R}_d &= \frac{1}{\sum_{d\in\mathcal{D}} |\Omega_d|} \sum_{d\in\mathcal{D}} \boldsymbol{H}_d^{\mathsf{H}} \boldsymbol{A}_d \boldsymbol{H}_d, \\ \boldsymbol{R}_b &= \frac{1}{|\Omega_b|} \boldsymbol{H}_b^{\mathsf{H}} \boldsymbol{A}_b \boldsymbol{H}_b, \\ \boldsymbol{r}_b &= \frac{1}{|\Omega_b|} \boldsymbol{H}_b^{\mathsf{H}} \tilde{\boldsymbol{A}}_b \tilde{\boldsymbol{h}}, \end{aligned} \tag{13}$$

the interpolation weighting matrix for a zone k is defined as

$$\boldsymbol{A}_k = \int_{\Omega_k} \boldsymbol{z}_k^*(\boldsymbol{r}) \boldsymbol{z}_k^\top(\boldsymbol{r})\, d\boldsymbol{r}, \tag{14}$$

and the interpolation weighting matrix considering the virtual sources is defined as

$$\tilde{\boldsymbol{A}}_b = \int_{\Omega_b} \boldsymbol{z}_b^*(\boldsymbol{r}) \tilde{\boldsymbol{z}}_b^\top(\boldsymbol{r})\, d\boldsymbol{r}. \tag{15}$$

3.4. Computing weighting matrices

The integrals in (14) and (15) must in general be evaluated using a numerical integration procedure, although analytical expressions can exist for exceptional cases [29]. To facilitate a more computationally efficient numerical integration, the matrix inversion can be moved outside the integral expression by using the definition (5). Introducing the matrix $\boldsymbol{P}_k = \text{blkdiag}\{(\boldsymbol{K}_k(\boldsymbol{\theta}_{lk}) + \lambda \boldsymbol{I})^{-1}\}_{l\in\mathcal{L}} \in \mathbb{C}^{LM_k \times LM_k}$ and stacking the kernel vectors $\boldsymbol{\kappa}_k(\boldsymbol{r}) = \text{col}\{\boldsymbol{\kappa}_k(\boldsymbol{r}, \boldsymbol{\theta}_{lk})\}_{l\in\mathcal{L}} \in \mathbb{C}^{LM_k}$, the integral expression for zone k can be rewritten as

$$\boldsymbol{A}_k = \boldsymbol{P}_k^{\mathsf{H}} \int_{\Omega_k} \boldsymbol{\kappa}_k^*(\boldsymbol{r}) \boldsymbol{\kappa}_k^\top(\boldsymbol{r})\, d\boldsymbol{r} \boldsymbol{P}_k. \tag{16}$$

If the loudspeaker positions are unknown, or computational resources are scarce, the parameter ρ in (3) can be set to zero, simplifying the kernel function to $\kappa(\boldsymbol{r}, \boldsymbol{r}') = j_0(\frac{\omega}{c}\|\boldsymbol{r} - \boldsymbol{r}'\|_2)$. The simplification means that there are L^2 blocks of size $M_k \times M_k$ in the weighting matrix $\boldsymbol{A}_k$ that are now identical. The integral can thus be computed for only one block instead of the full matrix. In addition, $\boldsymbol{A}_b = \tilde{\boldsymbol{A}}_b$.

3.5. Low rank approximation

In the optimal control filter expression (12), a low-rank approximation can be used in exactly the same way as for the pointwise VAST [20]. The spatial correlation matrices are jointly diagonalized as

$$\begin{aligned} \boldsymbol{U}^{\mathsf{H}} \boldsymbol{R}_b \boldsymbol{U} &= \boldsymbol{\Lambda} \\ \boldsymbol{U}^{\mathsf{H}} \boldsymbol{R}_d \boldsymbol{U} &= \boldsymbol{I}, \end{aligned} \tag{17}$$

where the invertible matrix $\boldsymbol{U} \in \mathbb{C}^{L\times L}$ and diagonal matrix $\boldsymbol{\Lambda} \in \mathbb{C}^{L\times L}$ are the generalized eigenvector and eigenvalue matrices of $\boldsymbol{R}_d^{-1}\boldsymbol{R}_b$. Keeping only the R largest eigenvalues $\boldsymbol{\Lambda}_R \in \mathbb{C}^{R\times R}$, along with the corresponding eigenvectors $\boldsymbol{U}_R \in \mathbb{C}^{L\times R}$, the final expression for the control filter vector is

$$\boldsymbol{w} = \boldsymbol{U}_R (\boldsymbol{\Lambda}_R + \mu \boldsymbol{I})^{-1} \boldsymbol{U}_R^{\mathsf{H}} \boldsymbol{r}_b. \tag{18}$$

Analogously to the pointwise VAST, the PM and ACC solutions with interpolation weighting are obtained as special cases, setting $\mu = 1$, $R = L$ for PM and $\mu = 1$, $R = 1$ for ACC. As shown in [20], the eigenvalues can also be sorted jointly for all frequency bins.

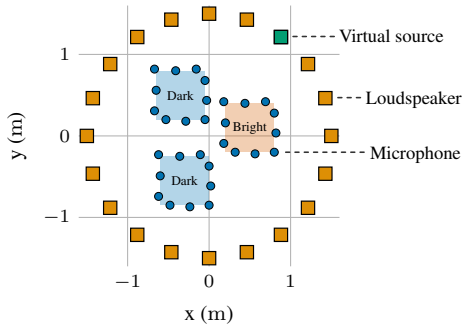


Fig. 1: Positions of microphones, sources, and zones in the experiments. The figure depicts the plane in a three-dimensional room on which all objects are placed.

4. EXPERIMENTS

To evaluate the proposed method, simulations of broadband sound in a reverberant environment were performed. The objective was to control sound in a horizontal plane of a three-dimensional room. A comparison was made between the proposed method with the directional weighting parameter defined in (3) set to $\rho = 3$, referred to as VAST-DKI, the proposed method without directional weighting, i.e., with $\rho = 0$, referred to as VAST-KI, and the pointwise frequency domain VAST method [20]. For all three methods, the cost weighting parameter was set to $\mu = 1$.

The placement of the microphones and loudspeakers can be seen in Fig. 1, all of which were placed on the same height. Each of the three microphone arrays consisted of 10 microphones, and 20 loudspeakers were used. For the two proposed methods, the microphone sets $\mathcal{M}_k$ were equal and contained all 30 microphones. For VAST, as the microphones cannot be associated with multiple zones, only the 10 microphones surrounding the zone k was used. The virtual source was selected to be one of the loudspeakers in the loudspeaker array, because its RIRs are known.

The RIRs were generated using the image-source method [30, 31], with an approximate reverberation time of $\text{RT}_{60} = 0.18\,\text{s}$. The optimal control filter with rank $R = 10$ was computed for $N_f = 2048$ frequencies, from which an finite impulse response (FIR) filter of length $N_f - 1$ was computed using the inverse discrete Fourier transform. The simulation was thereafter carried out in the time domain at a sample rate of 4000 Hz, using a white Gaussian input signal $x(n) \in \mathbb{R}$ of zero mean and unity variance, which is a function of the discrete time index n.

The regions defined for calculating the interpolation weighting matrices (14) and (15) were cuboids of size $0.6\,\text{m} \times 0.6\,\text{m} \times 0.05\,\text{m}$. The integration was performed with a naive Monte Carlo integration procedure with 1000 samples [32], and the regularization parameter set to $\lambda = 10^{-4}$.

	AC	SD	RE
VAST	2.88	−3.64	−2.02
VAST-KI	4.49	−4.23	−4.13
VAST-DKI	**6.74**	**−6.59**	**−7.25**

Table 1: The resulting acoustic contrast (AC), signal distortion (SD), and residual energy (RE) in dB for the three compared methods.

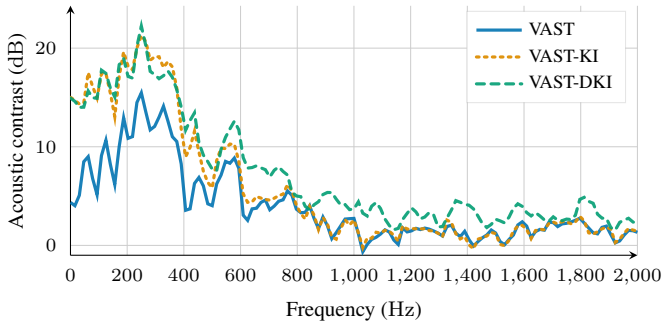


Fig. 2: Acoustic contrast in dB as a function of frequency for the three compared methods.

Evaluation points were equivalently spaced in all zones at a distance of 0.025 m, giving $B = 1058$ points in the bright zone and $D = 2116$ points in all dark zones. The B points in the bright zone can be indexed by the set $\mathcal{E}_b$, and the D points in all dark zones by $\mathcal{E}_d$. Three metrics were considered for evaluation, defined in terms of the time domain sound pressure $p(n, \boldsymbol{r}) \in \mathbb{R}$. The first metric, acoustic contrast (AC), is the ratio of sound power in the bright zone compared to the dark zones, defined as

$$\mathrm{AC} = \frac{\frac{1}{B}\sum_{b\in\mathcal{E}_b}\sum_{n=0}^{T-1} p(n,\boldsymbol{r}_b)^2}{\frac{1}{D}\sum_{d\in\mathcal{E}_d}\sum_{n=0}^{T-1} p(n,\boldsymbol{r}_d)^2}. \tag{19}$$

The signal distortion (SD) is the mean square error between the desired and reproduced signal in the bright zone, defined as

$$\mathrm{SD} = \frac{\sum_{b\in\mathcal{E}_b}\sum_{n=0}^{T-1}\big(p(n,\boldsymbol{r}_b) - d(n,\boldsymbol{r}_b)\big)^2}{\sum_{b\in\mathcal{E}_b}\sum_{n=0}^{T-1} d(n,\boldsymbol{r}_b)^2}. \tag{20}$$

Finally, the residual energy (RE) is the average sound power present in the dark zones, defined as

$$\mathrm{RE} = \frac{1}{D}\sum_{d\in\mathcal{E}_d}\sum_{n=0}^{T-1} p(n,\boldsymbol{r}_d)^2. \tag{21}$$

Ideally, AC should be high while SD and RE are low. The parameter T should be selected high enough to obtain accurate signal power estimates, and was set to $T = 30000$ samples.

Table 1 shows the resulting values for the three metrics for the three compared methods. Both proposed methods outperform VAST in all three metrics. In addition, the directionally weighted VAST-DKI outperforms VAST-KI with a significant margin in all three metrics.

The acoustic contrast as a function of frequency is shown in Fig. 2. In the figure it can be seen that VAST-KI has comparable performance to the superior VAST-DKI for low frequencies, but the performance gradually degrades, approaching the pointwise VAST as the frequency increase. The frequency domain acoustic contrast is defined as the ratio between the mean power spectrum at the bright zone evaluation points and the mean power spectrum at the dark zone evaluation points. The power spectrum for a single point is obtained as a 256-point Welch spectrum estimate with 50 % overlap, and using in total $T = 30000$ samples.

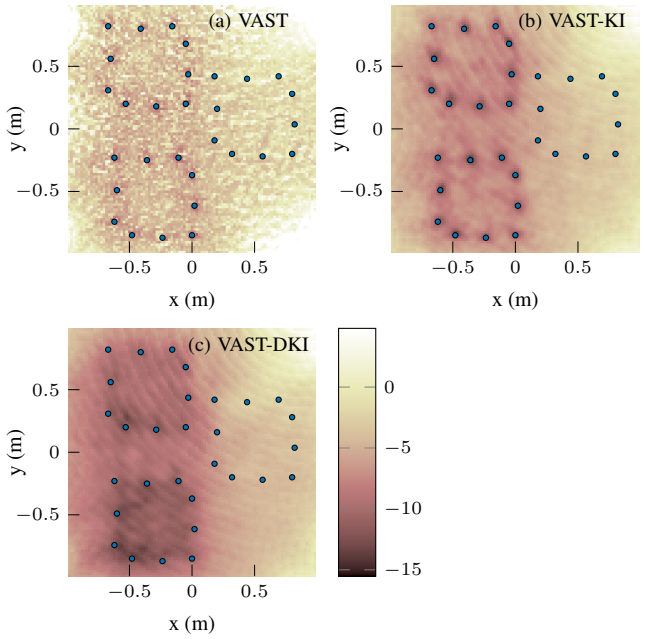


Fig. 3: Average power in dB of the sound field for the three compared methods. The colour range matches the values present within the sound zones, meaning that colours outside can be clipped.

Finally, the power distribution of the sound field is shown in Fig. 3. It is clear that VAST-DKI, and to a lesser extent VAST-KI, manages to take the entire bright and dark zones into account, even though only RIR measurements for a few discrete points are available.

5. CONCLUSION

A method for sound zone control has been proposed, addressing the situation when physical microphones must be used instead of pre-measured RIRs. Applying kernel interpolation, the sound pressure over the continuous zone can be taken into account, improving performance over the existing pointwise VAST, especially at lower frequencies. The performance is improved further by utilizing prior information about the relative positions of the loudspeakers and zones. It was shown through simulations in a reverberant space that the proposed method could outperform the pointwise VAST when RIRs are available only at the zone boundaries.

6. REFERENCES

[1] W. F. Druyvesteyn and J. Garas, "Personal Sound," *JAES*, vol. 45, no. 9, pp. 685–701, Sept. 1997.

[2] Y. Hur, S. W. Kim, Y.-c. Park, and D. H. Youn, "Highly focused sound beamforming algorithm using loudspeaker array system," in *Proc. AES Conv.* Oct. 2008, pp. 1–6.

[3] Y. J. Wu and T. D. Abhayapala, "Spatial multizone soundfield reproduction: Theory and design," *IEEE Trans. Audio Speech Lang. Process.*, vol. 19, no. 6, pp. 1711–1720, Aug. 2011.

[4] J.-W. Choi and Y.-H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," *J. Acoust. Soc. Am.*, vol. 111, no. 4, pp. 1695–1700, Apr. 2002.

[5] J.-H. Chang, C.-H. Lee, J.-Y. Park, and Y.-H. Kim, "A realization of sound focused personal audio system using acoustic contrast control," *J. Acoust. Soc. Am.*, vol. 125, no. 4, pp. 2091–2097, Apr. 2009.

[6] S. J. Elliott, J. Cheer, J. Choi, and Y. Kim, "Robustness and regularization of personal audio systems," *IEEE Trans. Audio Speech Lang. Process.*, vol. 20, no. 7, pp. 2123–2133, Sept. 2012.

[7] P. Coleman, P. Jackson, M. Olik, and J. A. Pedersen, "Optimizing the planarity of sound zones," in *Proc. AES Int. Conf.* Sept. 2013, pp. 1–10.

[8] Y. Cai, M. Wu, L. Liu, and J. Yang, "Time-domain acoustic contrast control design with response differential constraint in personal audio systems," *J. Acoust. Soc. Am.*, vol. 135, no. 6, pp. EL252–EL257, May 2014.

[9] D. H. M. Schellekens, M. B. Møller, and M. Olsen, "Time domain acoustic contrast control implementation of sound zones for low-frequency input signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Mar. 2016, pp. 365–369.

[10] M. Poletti, "An investigation of 2-D multizone surround sound systems," in *Proc. AES Conv.* Oct. 2008, pp. 1–9.

[11] T. Betlehem and P. D. Teal, "A constrained optimization approach for multi-zone surround sound," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2011, pp. 437–440.

[12] J.-H. Chang and F. Jacobsen, "Sound field control with a circular double-layer array of loudspeakers," *J. Acoust. Soc. Am.*, vol. 131, no. 6, pp. 4518–4525, June 2012.

[13] 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 *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Apr. 2018, pp. 491–495.

[14] R. Serizel, M. Moonen, B. V. Dijk, and J. Wouters, "Low-rank approximation based multichannel wiener filter algorithms for noise reduction with application in cochlear implants," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 4, pp. 785–799, Apr. 2014.

[15] J. R. Jensen, J. Benesty, and M. G. Christensen, "Noise reduction with optimal variable span linear filters," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 4, pp. 631–644, Apr. 2016.

[16] T. Lee, J. K. Nielsen, and M. G. Christensen, "Towards perceptually optimized sound zones: A proof-of-concept study," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2019, pp. 136–140.

[17] T. Lee, J. K. Nielsen, and M. G. Christensen, "Signal-adaptive and perceptually optimized sound zones with variable span trade-off filters," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 28, pp. 2412–2426, July 2020.

[18] L. Shi, T. Lee, L. Zhang, J. K. Nielsen, and M. G. Christensen, "A fast reduced-rank sound zone control algorithm using the conjugate gradient method," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2020, pp. 436–440.

[19] L. Shi, T. Lee, L. Zhang, J. K. Nielsen, and M. G. Christensen, "Generation of personal sound zones with physical meaningful constraints and conjugate gradient method," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 823–837, Jan. 2021.

[20] T. Lee, L. Shi, J. K. Nielsen, and M. G. Christensen, "Fast generation of sound zones using variable span trade-off filters in the DFT-domain," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 363–378, Dec. 2020.

[21] N. Ueno, S. Koyama, and H. Saruwatari, "Kernel ridge regression with constraint of Helmholtz equation for sound field interpolation," in *Proc. Int. Workshop Acoust. Signal Enhancement (IWAENC)*. Sept. 2018, pp. 436–440.

[22] N. Ueno, S. Koyama, and H. Saruwatari, "Sound field recording using distributed microphones based on harmonic analysis of infinite order," *IEEE Signal Process. Lett.*, vol. 25, no. 1, pp. 135–139, Jan. 2018.

[23] H. Ito, S. Koyama, N. Ueno, and H. Saruwatari, "Feedforward spatial active noise control based on kernel interpolation of sound field," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*. May 2019, pp. 511–515.

[24] J. Brunnström and S. Koyama, "Kernel-interpolation-based filtered-x least mean square for spatial active noise control in time domain," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, June 2021, pp. 161–165.

[25] S. Koyama, J. Brunnström, H. Ito, N. Ueno, and H. Saruwatari, "Spatial active noise control based on kernel interpolation of sound field," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 29, pp. 3052–3063, Aug. 2021.

[26] N. Iijima, S. Koyama, and H. Saruwatari, "Binaural rendering from microphone array signals of arbitrary geometry," *J. Acoust. Soc. Am.*, vol. 150, no. 4, pp. 2479–2491, Oct. 2021.

[27] S. Koyama, K. Kimura, and N. Ueno, "Sound field reproduction with weighted mode matching and infinite-dimensional harmonic analysis: An experimental evaluation," in *Proc. Int. Conf. Immersive 3D Audio*, Sept. 2021.

[28] N. Ueno, S. Koyama, and H. Saruwatari, "Directionally weighted wave field estimation exploiting prior information on source direction," *IEEE Trans. Signal Process.*, vol. 69, pp. 2383–2395, Apr. 2021.

[29] H. Ito, S. Koyama, N. Ueno, and H. Saruwatari, "Three-dimensional spatial active noise control based on kernel-induced sound field interpolation," in *Proc. Int. Congr. Acoust. (ICA)*, Sept. 2019, pp. 1101–1108.

[30] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[31] R. Scheibler, E. Bezzam, and I. Dokmanić, "Pyroomacoustics: A Python package for audio room simulations and array processing algorithms," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Apr. 2018, pp. 351–355.

[32] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, third edition, 2007.