Sound zone control for arbitrary sound field reproduction methods

Jesper Brunnström, Toon van Waterschoot and Marc Moonen
Dept. of Electrical Engineering (ESAT), STADIUS, KU Leuven, Leuven, Belgium

***Abstract*—Existing sound zone control methods can in principle be used for spatial audio, but faithful reproduction of a spatial sound scene is difficult, and current sound zone control methods do not allow for easy combination with state-of-the-art sound field reproduction methods. An alternative sound zone control problem statement is considered, where sound field reproduction is considered a separate process which generates loudspeaker signals representing the desired signal. The task is then to modify those signals such that the sound field in the dark zones is suppressed while the bright zone is preserved. A general problem statement is constructed, and solved using common assumptions of linearity and a cost function based on expected square Euclidian distance. The result is a simple algorithm that decouples the sound zone control and sound field reproduction steps. Under certain conditions the proposed method is equivalent to the conventional sound zone control method pressure matching. However, the proposed method can be easily applied in a wider range of situations, for example when the sound field reproduction is considerably complex, or when the desired sound field is unknown. The effectiveness of the method is demonstrated in a simulated reverberant environment applied to channel-based surround sound.**

***Index Terms*—sound zone control, spatial audio, sound field reproduction**

I. Introduction

Sound zone control [1] refers to the task of reproducing individual audio content for multiple listeners in the same acoustic environment with a loudspeaker array. Well established methods include variations on acoustic contrast control (ACC) [2]–[5], and variations on pressure matching, which minimizes the pressure error between the resulting and a desired sound field [6]–[9]. Both ACC and pressure matching can be considered special cases of the variable span trade-off filter (VAST) [10]–[13]. The problem is commonly considered under superposition, a problem statement where there is only a single bright zone where sound should be reproduced, while the other zones are dark zones where it should be quiet. By solving the problem separately with each zone as the bright zone and subsequently overlaying the solutions, the full sound zone control problem can be solved.

This research work was carried out at the ESAT Laboratory of KU Leuven, in the frame of KU Leuven internal funding, C14-21-0075 "A holistic approach to the design of integrated and distributed digital signal processing algorithms for audio and speech communication devices". This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 956369: 'Service-Oriented Ubiquitous Network-Driven Sound — SOUNDS'. The scientific responsibility is assumed by its authors. The research leading to these results has received funding from the European Research Council under the European Union's Horizon 2020 research and innovation program / ERC Consolidator Grant: SONORA (no. 773268). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

Sound zone control is generally considered as the reproduction of a single or a number of monophonic audio signals in the bright zone. This means that the spatial distribution of the sound field inside the bright zone is not determined a priori. Some methods such as ACC do not control the spatial distribution of the sound field at all, which can lead to a subpar listening experience, although this can be mitigated by enforcing spatial uniformity or planarity of the sound field [14], [15]. Another approach is to explicitly construct a desired sound field and then minimize the error between the reproduced and desired sound field, which is done in pressure matching, VAST, and their variations. This requires the sound zone control method to incorporate a way of constructing this desired sound field.

A common way to compute the desired sound field is by considering a number of virtual sources, and then the desired sound field is the one created by the virtual sources reproducing a number of monophonic audio signals. How to place the virtual sources, or what to choose as virtual source transfer functions, is not obvious. These choices can be viewed as degrees of freedom that may be used to improve the audio separation and audio quality, which has been done in [9], [16], [17]. The choice of desired sound field can also be reformulated into a simpler problem, as in [18], [19], where only the desired amplitude has to be chosen.

Instead of monophonic audio, spatial audio by sound field reproduction is considered in this paper. State-of-the-art sound field reproduction methods can be considerably complex [20], and different perceptual and spatial aspects must be taken into account for a good result [21]–[26]. There exists only for a small minority of methods a corresponding multizone method, one of them being ambisonics [27]–[29]. The way sound zone control is currently being developed, the sound field reproduction and sound zone control problems become intertwined, which means that each sound field reproduction method must be redeveloped individually for the sound zone control setting. Clearly it is not practical to do this for all methods, instead a more general sound zone control problem should be considered, that can be applied in combination with a wide range of sound field reproduction methods.

In this paper, the sound zone control problem is considered from a different point of view, assuming that the loudspeaker signals are already created by a separate sound field reproduction process. These loudspeaker signals are transformed such

ISBN: 978-9-4645-9360-0

that the sound in the dark zones is suppressed while the sound in the bright zone is left intact. The primary benefit of the simple resulting algorithm is a decoupling, where the sound zone control is only dependent on the room transfer functions between loudspeakers and control points, and the sound field reproduction does not need to take the sound zone control into account. This allows the use of any sound field reproduction method without modification. It is also shown that for a range of situations the proposed method is equivalent to sound zone control via pressure matching, meaning that the decoupling does not necessarily come at a cost in performance.

II. Problem statement

A sound reproduction system with L loudspeakers is considered. Following the superposition approach, the system should reproduce audio content in the bright zone, while producing as little sound as possible in the other dark zones. The loudspeaker signals are assumed to be processed in blocks which are transformed into the frequency domain. In practice this can be implemented in the short-time Fourier transform domain using a method such as weighted overlap-add (WOLA) or overlap-save [30]–[32]. All signals in the following derivations will be functions of frequency, but this will be left out of the notation to avoid clutter.

There is a sound field reproduction process generating a vector of loudspeaker signals $\boldsymbol{y} \in \mathbb{C}^L$. It is assumed that the generated loudspeaker signals are such that the resulting sound field in the bright zone optimally approximates the desired sound field, while the dark zones are not taken into account. The task of the sound zone control method is then to reduce the sound pressure level in the dark zones, while affecting the sound field in the bright zone as little as possible.

The acoustic transfer function from the loudspeakers to the bright zone control points is $\boldsymbol{H}_b \in \mathbb{C}^{M_b \times L}$, and to the control points in all the dark zones is $\boldsymbol{H}_d \in \mathbb{C}^{M_d \times L}$, where M_b and M_d is the number of control points in the bright zone and dark zones respectively. The sound pressure at the control points in the bright zone is then $\boldsymbol{p}_b = \boldsymbol{H}_b \boldsymbol{y} \in \mathbb{C}^{M_b}$, and at the control points in all the dark zones is $\boldsymbol{p}_d = \boldsymbol{H}_d \boldsymbol{y} \in \mathbb{C}^{M_d}$.

The problem is to find a mapping $f : \mathbb{C}^L \to \mathbb{C}^L$ that will return modified loudspeaker signals $\tilde{\boldsymbol{y}} = f(\boldsymbol{y})$ which retains the sound field in the bright zone, while reducing sound power in the dark zones. The modified loudspeaker signals give rise to a modified sound pressure, according to $\tilde{\boldsymbol{p}}_b = \boldsymbol{H}_b \tilde{\boldsymbol{y}}$ and $\tilde{\boldsymbol{p}}_d = \boldsymbol{H}_d \tilde{\boldsymbol{y}}$. The sound zone control objective can be written in terms of the modified sound pressures as

$$\underset{f}{\text{minimize}} \quad d(\tilde{\boldsymbol{p}}_b, \boldsymbol{p}_b) + \mu d(\tilde{\boldsymbol{p}}_d, \boldsymbol{0}) \tag{1}$$

where $d : \mathbb{C}^L \times \mathbb{C}^L \to \mathbb{R}_{\geq 0}$ is an appropriate distance measure, and $\mu \in \mathbb{R}_{\geq 0}$ is a parameter to trade off the effort in reducing the error in the bright zone versus dark zones.

The sound pressure is only defined at the control points, which will be sufficient if the control points are spaced evenly and densely enough. If not, sound field interpolation can be applied according to [33].

III. Optimization for linear transform

A. Optimization problem

If the loudspeaker signal mapping f is restricted to be linear, the modified loudspeaker signals can be defined as

$$\tilde{\boldsymbol{y}} = f(\boldsymbol{y}) = \boldsymbol{W}\boldsymbol{y} \tag{2}$$

for a square matrix $\boldsymbol{W} \in \mathbb{C}^{L \times L}$, referred to as the control filter matrix. If the expected square Euclidian distance is used in (1) as distance measure, the optimization problem can be expanded to

$$\underset{\boldsymbol{W}}{\text{minimize}} \quad \mathbb{E}\Big[\|\boldsymbol{H}_b\boldsymbol{W}\boldsymbol{y} - \boldsymbol{H}_b\boldsymbol{y}\|_2^2 + \mu\|\boldsymbol{H}_d\boldsymbol{W}\boldsymbol{y}\|_2^2\Big] \tag{3}$$

The optimum can be obtained by differentiating and setting the gradient to zero. The control filter matrix is optimal if

$$(\boldsymbol{W} - \mathcal{R})\boldsymbol{R}_y = 0 \tag{4}$$

where $\mathcal{R} = (\boldsymbol{R}_b + \mu\boldsymbol{R}_d)^{-1}\boldsymbol{R}_b$, and the constituent matrices are defined as

$$\begin{aligned} \boldsymbol{R}_b &= \boldsymbol{H}_b^{\mathsf{H}}\boldsymbol{H}_b \\ \boldsymbol{R}_d &= \boldsymbol{H}_d^{\mathsf{H}}\boldsymbol{H}_d \\ \boldsymbol{R}_y &= \mathbb{E}[\boldsymbol{y}\boldsymbol{y}^{\mathsf{H}}] \end{aligned} \tag{5}$$

It is assumed that there are more control points than loudspeakers, i.e. $M_b \geq L$ and $M_d \geq L$, which means that both $\boldsymbol{R}_b$ and $\boldsymbol{R}_d$ will be invertible except in degenerate cases.

If $\boldsymbol{R}_y$ is full rank, the optimal solution is unique and given by

$$\boldsymbol{W} = (\boldsymbol{R}_b + \mu\boldsymbol{R}_d)^{-1}\boldsymbol{R}_b \tag{6}$$

If $\boldsymbol{R}_y$ is rank-deficient, the solution is not unique. However, for a rank-deficient $\boldsymbol{R}_y$, (6) will still be optimal, so is therefore the only solution optimal for all possible $\boldsymbol{R}_y$.

B. Minimum-norm solution

If $\boldsymbol{R}_y$ is not full rank, the optimal $\boldsymbol{W}$ is not unique. To distinguish between optimal solutions, a natural choice is to prefer solutions with a smaller norm. The optimality criteria $(\boldsymbol{W} - \mathcal{R})\boldsymbol{R}_y = 0$ is equivalent to solving the following optimization problem

$$\begin{aligned} \underset{\boldsymbol{W}}{\text{minimize}} \quad & \|\boldsymbol{W}\|_F^2 \\ \text{subject to} \quad & (\boldsymbol{W} - \mathcal{R})\boldsymbol{R}_y = 0 \end{aligned} \tag{7}$$

The minimum-norm solution is given by $\boldsymbol{W} = \mathcal{R}\boldsymbol{R}_y\boldsymbol{R}_y^{\dagger}$, where $\boldsymbol{R}_y^{\dagger}$ is the Moore-Penrose inverse of $\boldsymbol{R}_y$. The matrix $\boldsymbol{R}_y\boldsymbol{R}_y^{\dagger}$ is an orthogonal projection onto the column space of $\boldsymbol{R}_y$, meaning that $\boldsymbol{R}_y\boldsymbol{R}_y^{\dagger}\boldsymbol{y} = \boldsymbol{y}$. Note that since $\boldsymbol{y}$ is used to estimate $\boldsymbol{R}_y$ the relationship holds even for sample covariance estimates of $\boldsymbol{R}_y$. The modified loudspeaker signals using the solution of (7) is therefore identical to (6). However, the solution of (7) is dependent on $\boldsymbol{y}$, which is undesirable.

C. Low-rank approximation

It is possible to achieve a higher acoustic contrast by performing a low-rank approximation via the generalized eigenvalue decomposition (GEVD) of the pencil $(\boldsymbol{R}_b, \boldsymbol{R}_d)$ [11]. The simultaneous diagonalization is defined as

$$\begin{aligned} \boldsymbol{U}^{\mathsf{H}}\boldsymbol{R}_b\boldsymbol{U} &= \boldsymbol{\Lambda} \\ \boldsymbol{U}^{\mathsf{H}}\boldsymbol{R}_d\boldsymbol{U} &= \boldsymbol{I} \end{aligned} \tag{8}$$

where $\boldsymbol{U} \in \mathbb{C}^{L\times L}$ is a matrix where the columns are the generalized eigenvectors and $\boldsymbol{\Lambda}$ is a diagonal matrix with the generalized eigenvalues $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_r\}_{r=1}^{R}$, with λ_1 being the largest and λ_R the smallest. With the definition $\boldsymbol{X} = \boldsymbol{U}^{-\mathsf{H}}$, and the vectors $\boldsymbol{u}_i$ and $\boldsymbol{x}_i$ being the ith columns of $\boldsymbol{U}$ and $\boldsymbol{X}$ respectively, the rank-R solution is

$$\boldsymbol{W} = \sum_{r=1}^{R} \frac{\lambda_r}{\lambda_r + \mu} \boldsymbol{u}_r \boldsymbol{x}_r^{\mathsf{H}} \tag{9}$$

The solution of (6) can be obtained as a special case of (9) when $R = L$.

D. Loudspeaker transmit power regularization

It is desirable to keep the loudspeaker transmit power low. Therefore, a regularization term can be added to the cost function in the optimization problem (3), i.e.

$$\mathbb{E}\Big[\|\boldsymbol{H}_b\boldsymbol{W}\boldsymbol{y} - \boldsymbol{H}_b\boldsymbol{y}\|_2^2 + \mu\|\boldsymbol{H}_d\boldsymbol{W}\boldsymbol{y}\|_2^2 + \gamma\|\boldsymbol{W}\boldsymbol{y}\|_2^2\Big] \tag{10}$$

for some $\gamma \in \mathbb{R}_{\geq 0}$. The optimal solution assuming a full rank $\boldsymbol{R}_y$ is

$$\boldsymbol{W} = (\boldsymbol{R}_b + \mu\boldsymbol{R}_d + \gamma\boldsymbol{I})^{-1}\boldsymbol{R}_b \tag{11}$$

For a low-rank solution as given by (9), the GEVD of the pencil $(\boldsymbol{R}_b, \boldsymbol{R}_d + \gamma\boldsymbol{I})$ should be used. The regularization both helps in lowering the transmit power, but also with numerical problems in the case of an ill-conditioned $\boldsymbol{R}_d$ [4].

IV. Equivalence to conventional methods

It is possible to show an equivalence between the proposed method and the conventional sound zone control method pressure matching under certain conditions. A low-rank approximation can be applied to the pressure matching solution according to [11], which means that the equivalence extends to VAST as well as ACC.

A. Pressure matching

The pressure matching method aims to reproduce the audio signal $x \in \mathbb{C}$ in the bright zone by transmitting the loudspeaker signals $\boldsymbol{w}x$, where $\boldsymbol{w} \in \mathbb{C}^{L}$ is the control filter. The desired signal in the bright zone is $\boldsymbol{h}_v x$ where $\boldsymbol{h}_v \in \mathbb{C}^{M_b}$ is the transfer function from the virtual source to the M_b control points in the bright zone. The control filter minimizes the following pressure matching cost function

$$\underset{\boldsymbol{w}}{\text{minimize}} \quad \mathbb{E}\Big[\|\boldsymbol{H}_b\boldsymbol{w}x - \boldsymbol{h}_v x\|_2^2 + \mu\|\boldsymbol{H}_d\boldsymbol{w}x\|_2^2\Big] \tag{12}$$

The optimal control filter is unique and is given by

$$\boldsymbol{w} = (\boldsymbol{R}_b + \mu\boldsymbol{R}_d)^{-1}\boldsymbol{H}_b^{\mathsf{H}}\boldsymbol{h}_v \tag{13}$$

The GEVD can be applied to the pencil $(\boldsymbol{R}_b, \boldsymbol{R}_d)$, and by performing a low-rank approximation a higher acoustic contrast can be obtained at the cost of distortion in the bright zone. For a regularized solution, the pencil $(\boldsymbol{R}_b, \boldsymbol{R}_d + \gamma\boldsymbol{I})$ should be used instead.

B. Equivalence when virtual sources are same as loudspeakers

Considering the loudspeakers as virtual sources, pressure matching is equivalent to the proposed method. L control filters $\begin{bmatrix}\boldsymbol{w}_1 & \ldots & \boldsymbol{w}_L\end{bmatrix}$ are required, each computed according to (13). The virtual source transfer function $\boldsymbol{h}_{v_l}$ of the lth control filter will be equal to the transfer function associated with the lth loudspeaker, i.e. $\boldsymbol{H}_b = \begin{bmatrix}\boldsymbol{h}_{v_1} & \ldots & \boldsymbol{h}_{v_L}\end{bmatrix}$. The audio signal x_l of the lth control filter will be equal to the lth component of the vector of loudspeaker signals $\boldsymbol{y}$, i.e. $\boldsymbol{y} = \begin{bmatrix}x_1, \ldots, x_L\end{bmatrix}^{\top}$. Given this construction, the pressure matching control filters will be equal to the proposed control filter matrix from (6) or (9), i.e. $\boldsymbol{W} = \begin{bmatrix}\boldsymbol{w}_1 & \ldots & \boldsymbol{w}_L\end{bmatrix}$, and the resulting modified loudspeaker signals generated by pressure matching equals that of the proposed method, i.e. $\tilde{\boldsymbol{y}} = \boldsymbol{W}\boldsymbol{y} = \sum_{l=1}^{L}\boldsymbol{w}_l x_l$.

C. Equivalence when loudspeaker signals are rendered by pressure matching

If the sound field reproduction method used together with the proposed method is pressure matching, the equivalence holds for a wider range of situations. To generate loudspeaker signals for sound field reproduction using pressure matching, the following optimization problem is solved.

$$\underset{\boldsymbol{y}}{\text{minimize}} \quad \mathbb{E}[\|\boldsymbol{H}_b\boldsymbol{y} - \boldsymbol{h}_v x\|_2^2] \tag{14}$$

which has an optimal solution in closed form,

$$\boldsymbol{y} = \boldsymbol{R}_b^{-1}\boldsymbol{H}_b^{\mathsf{H}}\boldsymbol{h}_v x \tag{15}$$

Applying the proposed linear mapping (6) to the loudspeaker signals generated by pressure matching (15), the modified loudspeaker signals are $\tilde{\boldsymbol{y}} = \boldsymbol{W}\boldsymbol{y} = (\boldsymbol{R}_b + \mu\boldsymbol{R}_d)^{-1}\boldsymbol{H}_b^{\mathsf{H}}\boldsymbol{h}_v x$. This is equal to $\boldsymbol{w}x$ where $\boldsymbol{w}$ is the control filter obtained from (13). The equivalence holds regardless of the virtual source path $\boldsymbol{h}_v$ and audio signal x.

V. Simulations

An experiment is presented here as a demonstration of the usefulness of the proposed method. Reproduction of channel-based surround sound is considered, in which case the desired sound field is not easily specified through virtual sources. Two zones are considered, each associated with a separate 5.1 surround sound reproduction system. One such system has six loudspeakers in total, three in front of the listener (left, center, right), two behind the listener (side left, side right), and one for low-frequency content (low-frequency effects). Six components of the unmodified loudspeaker signals $\boldsymbol{y}$ are the unchanged audio for the reproduction system associated with the bright zone, while the other six components are zero. Then, because both reproduction systems are considered jointly when

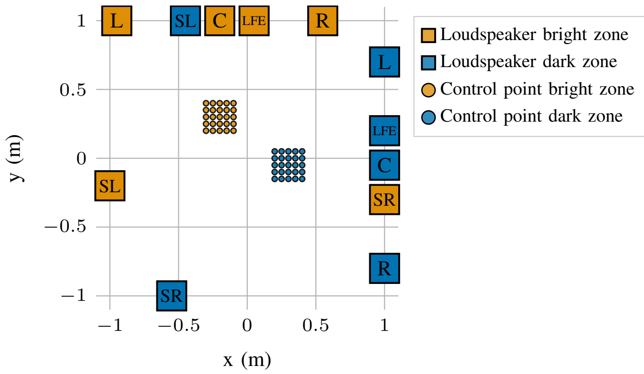

Fig. 1. Positions of the loudspeakers and control points associated with the bright zone (orange), and dark zone (blue). The loudspeakers are marked with their designated channel in the 5.1 surround sound reproduction system, left (L), center (C), right (R), side left (SL), side right (SR), and low frequency effects (LFE).

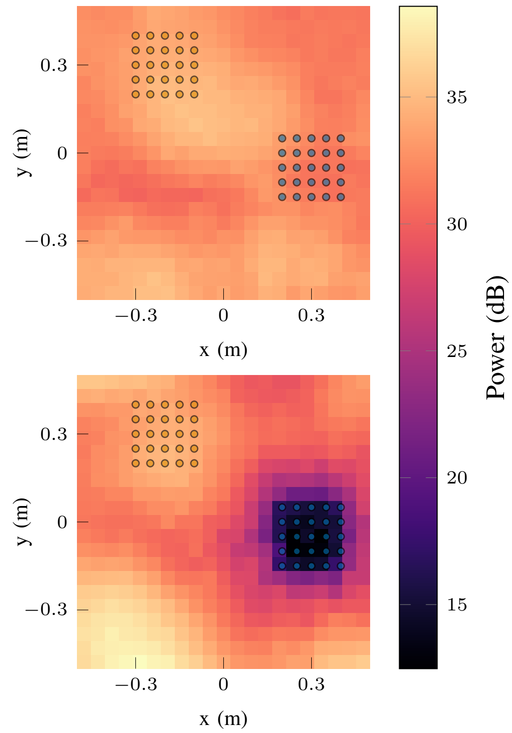

Fig. 2. Mean power in decibels of the sound field for the unmodified sound (top) and proposed method (bottom). Circles denote the control points associated with the bright zone (orange) and dark zone (blue).

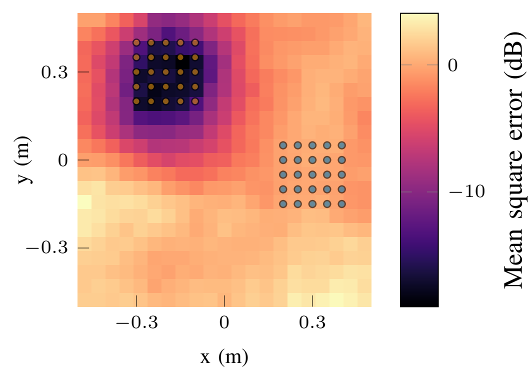

Fig. 3. Normalized mean square error of the proposed method in decibels. The unmodified sound field is considered the correct sound pressure values.

applying the proposed method, all 12 loudspeakers can be utilized. The system layout is shown in Fig. 1, where the listener in the bright zone is intended to face in the positive y-direction, and the listener in the dark zone in the positive x-direction.

To simulate a reverberant environment, recorded impulse responses are used [34]. For each of the bright and dark zones 25 control points are set on an equally spaced 5×5 grid at 5 cm distance. The position of the control points are shown in Fig. 1. The simulation is carried out in the time domain at a sampling frequency of 4 kHz, and the proposed algorithm is implemented using WOLA, with a block size of 4096 samples, overlap of 50% and a square root Hann window. The proposed method uses the low-rank approximation (9) with $R = 6$, weighting parameter set to $\mu = 1$, and a regularization parameter of $\gamma = 10^{-4}$ is used. The audio used is a 5.1 surround mix of the song *Home* by the artist *Please Keep Going*, available at [35]. For all following experiments 10 s of audio is used, starting 80 s into the song.

The mean power of the sound field is shown in Fig. 2. It can be seen that the sound power is considerably decreased in the dark zone while being essentially the same in the bright zone. The acoustic contrast at the control points, defined as the ratio of sound power in the bright zone to the power in the dark zone is 18.7 dB. In Fig. 3 the mean square error between the proposed method and the unmodified sound is shown. The figure shows that the proposed method produces a sound field similar to the unmodified sound field near the bright zone. The balance between suppressing sound in the dark zone and maintaining a low error in the bright zone can be controlled by the parameter μ and the rank R.

The loudspeaker transmit power as a function of frequency is shown in Fig. 4. The spectrum is calculated as the mean of the spectrum for all 12 loudspeakers, where each one is obtained as a 512-point Welch spectrum estimate using a Hann window. The figure shows that the loudspeaker transmit power required for the modified loudspeaker signals $\tilde{\boldsymbol{y}}$ is not considerably higher than the unmodified loudspeaker signals $\boldsymbol{y}$. For some frequencies the power required is slightly increased, and for some it is slightly decreased.

VI. Conclusion

A method for sound zone control has been presented, based on an alternative problem statement, where loudspeaker signals generated by a separate sound field reproduction process are modified to suppress noise in the dark zones, without affecting the bright zone. A general problem statement for this sound zone control approach has been provided, as well as a complete solution for one instance of the problem. The resulting algorithm can effectively perform the desired task without knowledge of how the loudspeaker signals are generated. This

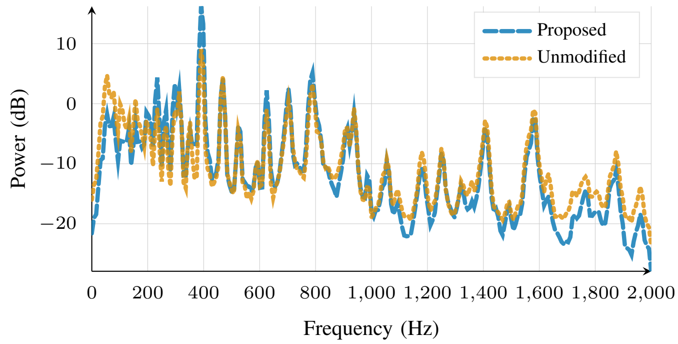

Fig. 4. Loudspeaker transmit power as a function of frequency for the proposed method as well as the unmodified sound.

approach offers considerable advantages when considering complex sound scenes, as it can be used in combination with essentially any sound field reproduction method without modification.

REFERENCES

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

[2] 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.

[3] 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.

[4] 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, Sep. 2012.

[5] 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.

[6] M. Poletti, "An investigation of 2-D multizone surround sound systems," in *Preprints AES Conv.*, Oct. 2008, paper 7551.

[7] 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.

[8] 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, Jun. 2012.

[9] V. Molés-Cases, S. J. Elliott, J. Cheer, G. Piñero, and A. Gonzalez, "Weighted pressure matching with windowed targets for personal sound zones," *J. Acoust. Soc. Am.*, vol. 151, no. 1, pp. 334–345, Jan. 2022.

[10] 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.

[11] 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.

[12] 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, Jul. 2020.

[13] 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.

[14] P. Coleman, P. Jackson, M. Olik, and J. A. Pedersen, "Optimizing the planarity of sound zones," in *Proc. AES Int. Conf.*, Sep. 2013, paper 5-1.

[15] S. Zhao and I. S. Burnett, "Time-domain acoustic contrast control with a spatial uniformity constraint for personal audio systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2022, pp. 1061–1065.

[16] M. Hu, H. Zou, J. Lu, and M. G. Christensen, "Maximizing the acoustic contrast with constrained reconstruction error under a generalized pressure matching framework in sound zone control," *J. Acoust. Soc. Am.*, vol. 151, no. 4, pp. 2751–2759, Apr. 2022.

[17] A. Borroni, M. Martalò, A. Costalunga, C. Tripodi, and R. Raheli, "Pressure matching with optimized target phase for personal sound zone systems," in *Proc. Int. Conf. Electr. Comput. Commun. Mechatron. Eng. (ICECCME)*, Nov. 2022.

[18] S. Koyama, T. Amakasu, N. Ueno, and H. Saruwatari, "Amplitude matching: Majorization–minimization algorithm for sound field control only with amplitude constraint," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Jun. 2021, pp. 411–415.

[19] T. Abe, S. Koyama, N. Ueno, and H. Saruwatari, "Amplitude matching for multizone sound field control," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 31, pp. 656–669, Dec. 2022.

[20] W. Zhang, P. N. Samarasinghe, H. Chen, and T. D. Abhayapala, "Surround by sound: A review of spatial audio recording and reproduction," *Appl. Sci.*, vol. 7, no. 5, May 2017, Art. no. 532.

[21] J. Merimaa and V. Pulkki, "Spatial impulse response rendering I: Analysis and synthesis," *J. Audio Eng. Soc.*, vol. 53, no. 12, pp. 1115–1127, Dec. 2005.

[22] V. Pulkki, "Spatial sound reproduction with directional audio coding," *J. Audio Eng. Soc.*, vol. 55, no. 6, pp. 503–516, Jun. 2007.

[23] J. Ahrens and S. Spors, "Reproduction of moving virtual sound sources with special attention to the doppler effect," in *Preprints AES Conv.*, May 2008, paper 7363.

[24] S. Tervo, J. P. Tynen, A. Kuusinen, and T. Lokki, "Spatial decomposition method for room impulse responses," *J. Audio Eng. Soc.*, vol. 61, no. 1, pp. 17–28, Mar. 2013.

[25] N. Ueno, S. Koyama, and H. Saruwatari, "Three-dimensional sound field reproduction based on weighted mode-matching method," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 27, no. 12, pp. 1852–1867, Dec. 2019.

[26] E. Erdem, E. De Sena, H. Hacıhabiboğlu, and Z. Cvetković, "Perceptual soundfield reconstruction in three dimensions via sound field extrapolation," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2019, pp. 8023–8027.

[27] J. Daniel, S. Moreau, and R. Nicol, "Further investigations of high-order ambisonics and wavefield synthesis for holophonic sound imaging," in *Preprints AES Conv.*, Mar. 2003, paper 5788.

[28] 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.

[29] H. Zuo, T. D. Abhayapala, and P. N. Samarasinghe, "3D multizone soundfield reproduction in a reverberant environment using intensity matching method," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Jun. 2021, pp. 416–420.

[30] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Process. Mag.*, vol. 9, no. 1, pp. 14–37, Jan. 1992.

[31] R. Crochiere, "A weighted overlap-add method of short-time Fourier analysis/synthesis," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 28, no. 1, pp. 99–102, Feb. 1980.

[32] S. Ruiz, T. Dietzen, T. van Waterschoot, and M. Moonen, "A comparison between overlap-save and weighted overlap-add filter banks for multi-channel Wiener filter based noise reduction," in *Proc. European Signal Process. Conf. (EUSIPCO)*, Aug. 2021, pp. 336–340.

[33] J. Brunnström, S. Koyama, and M. Moonen, "Variable span trade-off filter for sound zone control with kernel interpolation weighting," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2022, pp. 1071–1075.

[34] S. Koyama, T. Nishida, K. Kimura, T. Abe, N. Ueno, and J. Brunnström, "MeshRIR: A dataset of room impulse responses on meshed grid points for evaluating sound field analysis and synthesis methods," in *Proc. IEEE Int. Workshop Appl. Signal Process. Audio Acoust. (WASPAA)*, Oct. 2021, pp. 151–155.

[35] J. Brunnström, T. van Waterschoot, and M. Moonen, "Supplementary material for 'sound zone control for arbitrary sound field reproduction methods'," 2023. [Online]. Available: https://github.com/SOUNDS-RESEARCH/eusipco2023-szc-arbitrary-sf-reproduction