DOI: 10.61782/fa.2023.0882

ROOM IMPULSE RESPONSE RECONSTRUCTION FROM DISTRIBUTED MICROPHONE ARRAYS USING KERNEL RIDGE REGRESSION

Antonio Figueroa-Duran[1*] **Efren Fernandez-Grande**[1]

[1] Acoustic Technology Group, Deparment of Electrical and Photonics Engineering
Technical University of Denmark, Kongens Lyngby, 2800 Denmark

ABSTRACT

The use of Bayesian Inference and probabilistic models is an increasingly important topic in the field of sound field analysis. Kernel functions, widely utilised in Gaussian Processes, enable us to describe a sound field in terms of its spatial covariance. In this study, we explore the use of kernel functions to reconstruct the late part of a room impulse response, based on measurements from a set of distributed spherical microphone arrays. As the density of reflections in a room increases quadratically with time, and the spatial statistics of reverberant fields are well described, we are able to express the spatial covariance of the field as a closed-form function, allowing to solve the problem algebraically, which is computationally very efficient. The experimental results of this study show a successful reconstruction of the room impulse response as well as a fair extrapolation of the sound field far from the measurement aperture. The results also indicate an improvement in the computational burden, and a good generalisability across different rooms.

Keywords: *room impulse response, kernel ridge regression, sound field reconstruction, representer theorem*

1. INTRODUCTION

Sound field reconstruction techniques have gathered significant attention in the recent years due to their extensive use in various applications, including near-field acoustic holography [1], active noise control [2, 3], sound field reproduction [4–6], and underwater acoustics [7]. In particular, this study addresses the reconstruction of reverberant sound fields in rooms, a highly relevant field given that much of technology and communication occurs in indoors environments.

Acoustic fields in a room are usually intricate due to the boundaries and elements in the room that introduce scattering and diffraction effects. In the literature, various approaches can be found that characterise the sound field in a room, including plane wave expansions [8], modal expansions [9], sparse reconstruction reconstruction [10, 11], Bayesian inference [12, 13], kernel regression [14], deep-learning techniques [15, 16], and others.

It is common to find methodologies that exploit the time structure of the sound field, using different treatments for the early and late parts of the room impulse response. Whereas the early part presents a sparse energy distribution, after a few milliseconds the reflections become indistinguishable from each other [17]. The late reverberation is characterised by an increasingly higher number of reflections of chaotic nature, which hinders an accurate deterministic reconstruction. Within this context, statistical methods, such as Bayesian inference or probabilistic models, are a suitable approach when reconstructing reverberant sound fields.

In this paper, we propose a sound field reconstruction technique for reverberant sound fields that can be applied over large domains. The proposed methodology primarily aims at preserving the statistical properties of the late reverberation, as well as maintaining a uniform energy density across the room.

**Corresponding author*: *anfig@dtu.dk.*

2. THEORETICAL BACKGROUND

A room impulse response presents a general time structure consisting of the direct sound, early reflections and late reverberation. Whilst the early part exhibits a sparse distribution of reflections, the late part is characterised by a larger reflection density that increases quadratically with time [17]. During the late part, the acoustic energy tends to decay exponentially with time, and exhibits a uniform energy density on average across space for a given time instance.

A reverberant sound field can be modelled using a set of infinite plane waves with random phase travelling in random directions. In this way, the sound field in the room is approximately diffuse, and the sound pressure p at a point $\mathbf{r}$ for a frequency $f = \omega/2\pi$ is given by [18]

$$p(\mathbf{r}) = \lim_{N\to\infty} \frac{1}{\sqrt{N}} \sum_{i=1}^{N} A_i e^{j(\omega t - \mathbf{k}_i \cdot \mathbf{r})}, \tag{1}$$

where $A_i = |A_i| e^{j\varphi_i} \in \mathbb{C}$ is the complex amplitude and $\mathbf{k}_i \in \mathbb{R}^3$ is the wave vector of the wave i. This model can be used to cast the sound field reconstruction problem as an optimisation task, where we consider the cost function

$$J(\mathbf{a}) = \|\mathbf{p} - \mathbf{H}\mathbf{a}\|_2^2 + \lambda \|\mathbf{a}\|_2^2. \tag{2}$$

Here, $\mathbf{H} \in \mathbb{C}^{M\times N}$ represents N plane waves measured at M different positions, $\mathbf{a} \in \mathbb{C}^N$ is the complex amplitude of each wave, $\mathbf{p} \in \mathbb{C}^M$ is the measured data for a specific frequency and λ is a regularisation parameter. Once the coefficients $\mathbf{a}$ are obtained, they can be used to extrapolate the measured data basing the reconstruction on a linear combination of the plane waves described in Eq. (1). Nevertheless, solving Eq. (2) for the case $N \to \infty$ would require an infinite number of coefficients. To overcome this challenge, we can resort to the kernel trick [19], a mathematical method that replace the dot products (as in $\mathbf{Ha}$) with a positive definite function, called the kernel function. This approach allows for estimating the sound pressure algebraically at a reconstruction point $\mathbf{p}_\bullet$ in a high-dimensional space without the need of the coefficients A_i. This high-dimensional space is defined by the kernel function, which corresponds to the spatial correlation of the sound field. The spatial correlation can be obtained from Eq. (1) considering two points of the domain, $\mathbf{r}_1$ and $\mathbf{r}_2$, and averaging over a whole sphere, and is explicitly given by [20]

$$\mathbb{E}\left[p(\mathbf{r}_1)p(\mathbf{r}_2)^*\right] = \sigma^2 \frac{\sin\left(k||\mathbf{r}_1 - \mathbf{r}_2||\right)}{k||\mathbf{r}_1 - \mathbf{r}_2||}, \tag{3}$$

where $\mathbb{E}[\cdot]$ is the expectation operator, $\sigma^2 = \mathbb{E}\left[|p|^2\right]$ is the variance of the acoustic pressure, and $||\cdot||$ is the 2-norm in an euclidean space. This kernel is suitable since captures the physical properties of a reverberant sound field based on the ensemble statistics, providing a closed-form expression that maximises the efficiency of the reconstruction. The reconstructed pressure using the kernel trick requires only of the measured data and the spatial correlation, and can be obtained via [19]

$$\mathbf{p}_\bullet = \mathbf{K}_{MN}^T (\mathbf{K}_{MM} + \lambda \mathbf{I})^{-1} \mathbf{p}, \tag{4}$$

where $\mathbf{K}_{MN} \in \mathbb{R}^{M\times N}$ and $\mathbf{K}_{MM} \in \mathbb{R}^{M\times M}$ are mapping matrices defined by Eq. (3).

2.1 Energy decay and Anchor Points

The reconstruction based on the Bessel kernel underestimates the energy of the sound field when extrapolating far from the measurement points. This is a natural consequence of the spatial correlation values, Eq. (3), for $kr >> 1$. Since the reconstruction relies on this correlation and it decreases far from the aperture, the reconstructed energy is also expected to be low far from the measurement points. This phenomenon presents a problem when measuring with a distributed set of microphone arrays, given that the energy density can be considered spatially uniform on average in a reverberant room.

In this paper, we propose using additional synthesised pressure values that anchor the correlation at locations where there are not enough measurements to infer the sound field with sufficient energy. These values are termed anchor points, and they are not directly measured but rather drawn (as samples) from the statistical distribution of the measured pressure. In particular, in a reverberant field the values of the mean squared pressure $|p_{RMS}|^2$ follow an exponential distribution when measured along a line [21]. The methodology is as follows: first, the measured data within a short time window is used to find the rate parameter of this exponential distribution by fitting the $|p_{RMS}|^2$ histogram. Then, we use this distribution to draw the anchor point values that are statistically similar to the data and preserve the energy density across the room. Since the spacing between the anchor points should be relative to the wavelength, a different number of anchor points will be used depending on the frequency. It must be noted that rather than a deterministic extrapolation of the data, the primary aim is to obtain a qualitative reconstruction that preserves the statistical properties of the sound field and specially the late part of the room impulse response.

3. RESULTS

Experimental measurements were conducted to evaluate the performance of the proposed method. Room impulse responses were measured in a fully-furnished classroom of dimensions $(6.2, 9.4, 3.0)$ m with an averaged reverberation time of $T_{30} = 0.6$ s. Two open spherical microphone arrays were used to reconstruct the sound pressure, where each array was comprised of 72 microphones (4192, HBK) of which 61 laid on a sphere of 25 cm and the remaining 11 were place in the interior for stability. The arrays were placed 3 m apart. As ground truth, a reference line of 456 positions was measured with a spacing of 1 cm between sensors, covering an aperture of 4.5 m. In order to reduce uncertainty in the microphone positioning, a robotic arm (UR5, Universal Robots) was used. The source utilised was a two-way loudspeaker (BM6, Dynaudio), driven with a logarithmic sweep covering the frequency range $20 - 20$k Hz. More details can be found in [22]. Figure 1 shows a simple sketch of the setup. Here, the domain covered by the anchor points is also included, although their specific position changes with the frequency.

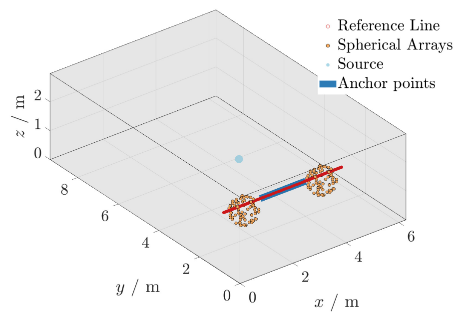

Figure 1: Sketch of the measurement setup, where the two open spherical microphone arrays, the reference line and the source are represented. Additionally, the coverage of the anchor points is included. Their specific position depends on the frequency.

We assess the reconstruction of the impulse response primarily based on the statistical properties and the energy density. That is, the correlation of the reconstruction should follow the Bessel kernel (c.f. Eq. (3)); the histogram of the mean squared pressure along a line should approximately follow an exponential distribution; and the energy density should be relatively uniform on average across the space for a given time window.

Figure 2 shows the mean squared pressure of the sound field for two different time windows. The different curves depict the ground truth and the reconstructed sound field using both the kernel trick ((4)) and the proposed method that includes the use of anchor points. The anchor points are displayed along the reconstruction domain, i.e. the reference line (see Fig. 1). For these results, the anchor points were set to be a wavelength apart from each other.

Comparing the reconstruction obtained using both methods (with and without anchor points) to the ground truth, we can see in Fig. 2 that the energy is underestimated over the entire measurement aperture. Considering that the method uses the Bessel kernel as the spatial correlation, there is a fraction of the true energy that does not follow Eq. (3) and is not being reconstructed. The so-called Representer Theorem in its semi-parametric form [23] offers a mathematical framework that can be exploited to refine the reconstruction, combining different sound field models. On the other hand, comparing the reconstruction at two time windows in Fig. 2 (top vs. bottom), it is clear that the reconstruction follows the energy decay as a function of time accurately.

As mentioned before, the energy of the reconstruction decays far from the array as consequence of the decay of the Bessel kernel. This is seen in Fig. 2 for the kernel trick method, where the energy is not uniform across the reconstruction aperture. In contrast, the proposed methodology includes the use of anchor points, and the results show a rather uniform energy distribution along the reconstruction domain, successfully covering an aperture of 4.5 m.

4. CONCLUSION

In this paper, we have presented a method for reconstructing the late part of the room impulse response using distributed arrays over a large spatial aperture. The method is based on the use of kernel methods, which exploits the prior knowledge on the spatial correlation of the sound field. However, since the spatial correlation is modelled as a Bessel function, classical kernel methods lead to underestimating the energy of the sound field far from the measurement points. The proposed methodology makes use of anchor points to maintain the energy density uniform at those points where measurements are not available, pre-

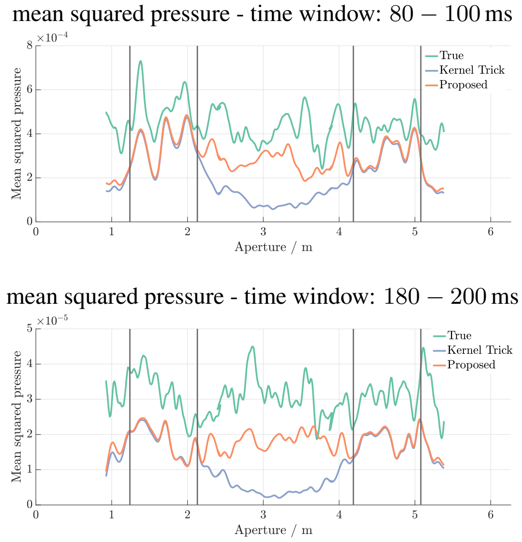

Figure 2: Mean squared pressure along the reconstruction line, averaged over the duration of each time window (20 ms). The three curves correspond to the reference true value, the kernel trick and the proposed method, where anchor points are included.

serving the statistical properties of the late reverberation. Therefore, the achieved reconstruction is spatially correlated to either the measurements or the anchor points, the mean squared pressure along the reconstruction domain follows an exponential distribution, and the energy is uniform on average across the room for a specific time window.

5. ACKNOWLEDGMENTS

This work has been partially supported by the VILLUM Foundation, under grant number 19179 for the project 'Large Scale Acoustic Holography'. The authors would also like to thank Xenofon Karakonstantis and Finn T. Agerkvist for the valuable discussions.

6. REFERENCES

[1] E. Fernandez-Grande, A. Xenaki, and P. Gerstoft, "A sparse equivalent source method for near-field acoustic holography," *The Journal of the Acoustical Society of America*, vol. 141, pp. 532–542, 1 2017.

[2] F. M. Heuchel, E. Fernandez-Grande, F. T. Agerkvist, and E. Shabalina, "Active room compensation for sound reinforcement using sound field separation techniques," *The Journal of the Acoustical Society of America*, vol. 143, pp. 1346–1354, 3 2018.

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

[4] S. Tervo, J. Pätynen, A. Kuusinen, and T. Lokki, "Spatial decomposition method for room impulse responses," *J. Audio Eng. Soc.*, vol. 61, 2013.

[5] J. G. Tylka and E. Y. Choueiri, "Domains of practical applicability for parametric interpolation methods for virtual sound field navigation," *J. Audio Eng. Soc*, vol. 67, pp. 882–893, 2019.

[6] A. Politis, S. Tervo, and V. Pulkki, "Compass: Coding and multidirectional parameterization of ambisonic sound scenes," vol. 2018-April, pp. 6802–6806, Institute of Electrical and Electronics Engineers Inc., 9 2018.

[7] P. Gerstoft, K. G. Sabra, P. Roux, W. A. Kuperman, and M. C. Fehler, "Green's functions extraction and surface-wave tomography from microseisms in southern california," *Geophysics*, vol. 71, 2006.

[8] D. N. Zotkin, R. Duraiswami, and N. A. Gumerov, "Plane-wave decomposition of acoustical scenes via spherical and cylindrical microphone arrays," *IEEE Transactions on Audio, Speech and Language Processing*, vol. 18, pp. 2–16, 1 2010.

[9] F. Asano and D. C. Swanson, "Sound equalization in enclosures using modal reconstruction," *The Journal of the Acoustical Society of America*, vol. 98, pp. 2062–2069, 10 1995.

[10] S. A. Verburg and E. Fernandez-Grande, "Reconstruction of the sound field in a room using compressive sensing," *The Journal of the Acoustical Society of America*, vol. 143, pp. 3770–3779, 6 2018.

[11] N. Antonello, E. D. Sena, M. Moonen, P. A. Naylor, and T. V. Waterschoot, "Room impulse response interpolation using a sparse spatio-temporal representation of the sound field," *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 25, pp. 1929–1941, 10 2017.

[12] D. Caviedes-Nozal, N. A. B. Riis, F. M. Heuchel, J. Brunskog, P. Gerstoft, and E. Fernandez-Grande, “Gaussian processes for sound field reconstruction,” *The Journal of the Acoustical Society of America*, vol. 149, pp. 1107–1119, 2 2021.

[13] J. M. Schmid, E. Fernandez-Grande, M. Hahmann, et al, C. Gurbuz, M. Eser, and S. Marburg, “Spatial reconstruction of the sound field in a room in the modal frequency range using bayesian inference,” *The Journal of the Acoustical Society of America*, vol. 150, pp. 4385–4394, 12 2021.

[14] N. Ueno, S. Koyama, and H. Saruwatari, “Kernel ridge regression with constraint of helmholtz equation for sound field interpolation,” *16th International Workshop on Acoustic Signal Enhancement, IWAENC 2018 - Proceedings*, pp. 436–440, 11 2018.

[15] E. Fernandez-Grande, X. Karakonstantis, D. Caviedes-Nozal, and P. Gerstoft, “Generative models for sound field reconstruction,” *The Journal of the Acoustical Society of America*, vol. 153, pp. 1179–1190, 2 2023.

[16] X. Karakonstantis and E. Fernandez-Grande, “Invertible neural networks for reconstructing acoustic fields,” *Journal of the Acoustical Society of America*, vol. 151, no. 4, pp. A231–A231, 2022.

[17] H. Kuttruff, *Room Acoustics*. Crc Press, 6th edition ed., 2019.

[18] F. Jacobsen and A. R. Molares, “The ensemble variance of pure-tone measurements in reverberation rooms,” *The Journal of the Acoustical Society of America*, vol. 127, pp. 233–237, 1 2010.

[19] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer New York, NY, 1st edition ed., 2006.

[20] D. Lubman, “Spatial averaging in a diffuse sound field,” *J Acoust Soc Am*, vol. 46, pp. 532–534, 1969.

[21] F. Jacobsen and P. M. Juhl, *Fundamentals of general linear acoustics*. Wiley, 2013.

[22] E. Fernandez-Grande, D. Caviedes-Nozal, M. Hahmann, X. Karakonstantis, and S. A. Verburg, “Reconstruction of room impulse responses over extended domains for navigable sound field reproduction,” in *International Conference on Immersive and 3D Audio*, pp. 1–8, Institute of Electrical and Electronics Engineers (IEEE), 11 2021.

[23] P. Bouboulis, K. Slavakis, and S. Theodoridis, “Adaptive kernel-based image denoising employing semi-parametric regularization,” *IEEE Transactions on Image Processing*, vol. 19, pp. 1465–1479, 6 2010.