



Reconstruction of the sound field in a room using compressive sensing

Verburg, Samuel A.; Fernandez Grande, Efrén

Published in:
Journal of the Acoustical Society of America

Link to article, DOI:
[10.1121/1.5042247](https://doi.org/10.1121/1.5042247)

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Verburg, S. A., & Fernandez Grande, E. (2018). Reconstruction of the sound field in a room using compressive sensing. *Journal of the Acoustical Society of America*, 143(6), 3770-3779. <https://doi.org/10.1121/1.5042247>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Reconstruction of the sound field in a room using compressive sensing

Samuel A. Verburg, and Efren Fernandez-Grande

Citation: [The Journal of the Acoustical Society of America](#) **143**, 3770 (2018); doi: 10.1121/1.5042247

View online: <https://doi.org/10.1121/1.5042247>

View Table of Contents: <http://asa.scitation.org/toc/jas/143/6>

Published by the [Acoustical Society of America](#)

Articles you may be interested in

[Introduction to special issue on compressive sensing in acoustics](#)

[The Journal of the Acoustical Society of America](#) **143**, 3731 (2018); 10.1121/1.5043089

Reconstruction of the sound field in a room using compressive sensing

Samuel A. Verburg^{a)} and Efren Fernandez-Grande

Acoustic Technology, Department of Electrical Engineering, Technical University of Denmark, Ørstedes Plads, building 352, 2800 Kongens Lyngby, Denmark

(Received 21 December 2017; revised 12 April 2018; accepted 16 April 2018; published online 29 June 2018)

Capturing the impulse or frequency response functions within extended regions of a room requires an unfeasible number of measurements. In this study, a method to reconstruct the response at arbitrary points based on compressive sensing (CS) is examined. The sound field is expanded into plane waves and their amplitudes are estimated via CS, obtaining a spatially sparse representation of the sound field. The validity of the CS assumptions are discussed, namely, the assumption of the wave field spatial sparsity (which depends strongly on the properties of the specific room), and the coherence of the sensing matrix due to different spatial sampling schemes. An experimental study is presented in order to analyze the accuracy of the reconstruction. Measurements with a scanning robotic arm make it possible to circumvent uncertainty due to positioning and transducer mismatch, and examine the accuracy of the reconstruction over extended regions of space. The results indicate that near perfect reconstructions are possible at low frequencies, even from a limited set of measurements. In addition, the study shows that it is possible to reconstruct damped room responses with reasonable accuracy well into the mid-frequency range. © 2018 Acoustical Society of America. <https://doi.org/10.1121/1.5042247>

[WS]

Pages: 3770–3779

I. INTRODUCTION

The frequency response function (or analogously the impulse response function) between two points in a room describes the acoustic wave propagation between them. The experimental estimation of frequency responses in rooms is typically limited to the set of measured points.¹ Nonetheless, knowledge of spatially extended room responses is valuable in many applications related to the analysis, reproduction and control of sound fields, e.g., in room compensation for audio reproduction systems.^{2–6} Consequently, it is of interest to acquire the responses over an extended space of the room.

The spatial sampling of room responses has been addressed in several studies. The acquisition of room impulse responses can be conceived as a standard sampling problem,⁷ for which the Shannon–Nyquist theorem provides the uniform spatial sampling rate necessary for the reconstruction up to certain frequency f_u . In practice, the uniform sampling leads to an extremely large number of measurements that grows proportionally to f_u^3 in the three dimensional (3D) case. The Green's function provides an analytic expression for the acoustic transfer between any two points, and inside an enclosure it can be described as a modal expansion.⁸ The method presented in Ref. 9 makes use of this to reduce the number of measurements needed, and it shows results for the interpolation and extrapolation of responses along a line. More recently, Mignot *et al.*¹⁰ further reduced the number of measurements by exploiting some knowledge on the propagation of sound waves, showing that the response of rectangular rooms can be characterized with a limited, *spatially sparse*

set of waves. This sparsity makes compressive sensing (CS) well-suited to the problem. The temporal sparsity of early reflections has also been exploited in order to interpolate the early part of impulse responses.¹¹ Antonello *et al.*¹² compared different methods for the interpolation of impulse responses, promoting sparsity in different domains. In Ref. 13 the reverberant sound field is modelled as a sum of plane waves assuming a two dimensional sound field, and in Ref. 14 the transfer function between a source and receiver regions inside a room is characterized by a set of coefficients that are independent of the precise location of the sensors.

CS makes it possible to reconstruct apparently *under-sampled* signals, below the Shannon–Nyquist limit. CS is based on two main assumptions: first, that the underlying signal is sparse, i.e., in some basis, only a few of the signal coefficients are non-zero; and second, that the columns of the sensing matrix (transfer matrix between measurements and signal coefficients) are mutually incoherent. The column incoherence serves to ensure uniqueness of the solution,¹⁵ although in the case of spatial sampling of waves, the sensing matrices depend on the spatial sampling scheme and on a dictionary of wave directions. Consequently, strict incoherence is unlikely, and a perfect reconstruction cannot be guaranteed.

In this study, we propose a CS reconstruction method to extrapolate the sound field outside the measurement area (or to interpolate it inside), and estimate the frequency and impulse responses over extended regions of space. Furthermore, we examine the suitability of CS to model the sound field in a room, and address some of the questions that seem lacking in the existing literature, namely, the assumption of spatial sparsity of the sound field in a room, the influence of damping in the room, and the influence of the

^{a)}Electronic mail: saveri@elektro.dtu.dk

coherence of the sensing matrix. An experimental validation is presented, based on measurements with a scanning robotic arm to prevent uncertainty due to position and transducer mismatch. The study makes it possible to evaluate quantitatively and qualitatively the accuracy of the reconstruction not only at single positions (as typically found in the literature), but over extended regions of space ($\approx 1.43 \text{ m}^3$).

The paper is organized as follows: a method for reconstructing an arbitrary sound field from a limited set of measurements is presented in Sec. II. The method is based on the decomposition of the sound field into a limited number of plane waves, and their amplitude is estimated via CS. A discussion on the sparsity of the problem is presented in Sec. II A. In Sec. II B considerations on the coherence of the sampling process are discussed. Section III presents the experimental results of the spatial response reconstruction in two acoustically different rooms. Finally, Sec. IV describes the conclusions that can be withdrawn from this study.

II. THEORY

A reverberant sound field in a convex source-free domain $\Omega \subset \mathbb{R}^3$ is a solution of the homogeneous Helmholtz equation with unknown boundary conditions. The pressure distribution at a given frequency f can be approximated as¹⁶

$$p(k, \mathbf{r}) \simeq \sum_{j=1}^N X_j e^{i\mathbf{k}_j \mathbf{r}} + n(k, \mathbf{r}), \quad (1)$$

where each of the terms in the sum corresponds to one of the N plane waves that compose the sound field. The coefficients X_j represent the complex amplitude of the j th wave. The direction of propagation of such wave is given by the wave-number vector $\mathbf{k}_j = k_x \hat{x} + k_y \hat{y} + k_z \hat{z}$, while the point in which the pressure is observed is represented by the vector $\mathbf{r} = x\hat{x} + y\hat{y} + z\hat{z}$ with $\mathbf{r} \in \Omega$. In this paper only propagating plane waves are treated, i.e., \mathbf{k}_j lies on the surface of a sphere of radius $k = 2\pi f/c$,¹⁷ where c is the speed of sound in air. The term $n(k, \mathbf{r})$ represents additive Gaussian noise. Note that the plane wave expansion has been defined in a Cartesian coordinate system; however, the decomposition can also be expressed as a sum of spherical harmonics.¹⁷

The Vekua theory¹⁶ serves to show the convergence of the approximation stated in Eq. (1) in terms of N and the size of Ω . Reference 18 provides an estimate of the minimum number of plane waves necessary to approximate a solution of the homogeneous Helmholtz equation in the 3D domain Ω with characteristic radius a_Ω

$$N \simeq (\lceil ka_\Omega \rceil + 1)^2, \quad (2)$$

where $\lceil \cdot \rceil$ is the ceiling function. It is useful to express Eq. (1) as a matrix equation

$$\mathbf{p} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad (3)$$

where $\mathbf{H} \in \mathbb{C}^{M \times N}$ is the sensing matrix, with $\mathbf{H}_{ij} = e^{i\mathbf{k}_j \mathbf{r}_i}$, M the number of points in which the pressure is measured, and N the number of plane waves that constitute the sound field

at one frequency. The elements of vector $\mathbf{x} \in \mathbb{C}^N$ are the unknown complex amplitudes of each wave with direction \mathbf{k}_j , denoted by X_j in Eq. (1); while the vector $\mathbf{p} \in \mathbb{C}^M$ is composed by the pressure p observed at each point of the space \mathbf{r}_i . The vector $\mathbf{n} \in \mathbb{C}^M$ consists of the noise at each position.

The estimation of wave amplitudes $\tilde{\mathbf{x}}$ requires the discretization of the continuum of space directions resulting in a dictionary with N possible wave directions. Assuming that the sound field in the room is composed by a limited number of waves,⁸ only the corresponding coefficients of \mathbf{x} will be non-zero. In most of the cases $N > M$, thus, estimating $\tilde{\mathbf{x}}$ poses an underdetermined system. The process of finding the best set of all possible solutions can be expressed as an optimization problem, which can be solved via CS assuming \mathbf{x} to be sparse¹⁹

$$\tilde{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{C}^N} \|\mathbf{x}\|_1 \text{ subject to } \|\mathbf{H}\mathbf{x} - \mathbf{p}\|_2 \leq \epsilon, \quad (4)$$

where the operator $\|\cdot\|_n$ indicates the l_n -norm. The parameter ϵ is an estimation of the upper bound for the noise present in the sensing process. Note that the choice of ϵ will influence the number of non-zero elements of $\tilde{\mathbf{x}}$: a too low ϵ estimation will induce noise components in the solution; while a too high estimate of ϵ will lead to discarding waves with smaller amplitude from $\tilde{\mathbf{x}}$. In this study it is assumed that a good estimation for ϵ is $\|\mathbf{n}\|_2$.

It is possible to use the wave basis to predict the sound field at a set of positions $\mathbf{r}_R \in \Omega$ that have not been measured, enabling to estimate the sound field over an extended region of space

$$\tilde{\mathbf{p}}(k, \mathbf{r}_R) = \mathbf{H}_R \tilde{\mathbf{x}}. \quad (5)$$

Note that this formulation allows not only the interpolation of the sound field within the volume in which the measurements are distributed, but also the extrapolation at points outside this volume, as long as the extrapolated point remains far from the source.

A. Sparsity of reverberant sound fields

CS is based on the assumption that the retrieved signal is sparse in some basis. The vector of complex amplitudes \mathbf{x} is *s-sparse* if just a small number S of its coefficients are non-zero, with $S \ll N$. Furthermore, a vector can be *approximately* sparse if a few of its entries are much larger than the rest, thus, not much information is lost when the small value coefficients are discarded. Reference 19 shows the number of measurements needed for an exact CS reconstruction with high probability

$$M \geq C\mu^2 S \log N, \quad (6)$$

where C is a positive constant and μ is a measure of the coherence of the sensing matrix \mathbf{H} (see Sec. II B). The use of CS in the response reconstruction task and the number of measurements needed is then conditioned by the number and amplitude of the waves that constitute the room response at each frequency. In rectangular rooms, assuming frequencies

in which the oblique modes dominate, the number of waves can be approximated by the average number of active modes (*modal overlap*) times the number of plane waves per mode (eight in the case of oblique modes).⁸ For example, at the Schroeder frequency f_{Schro} , in which the modal overlap is three, the resulting sound field is composed by approximately 24 waves of comparable magnitudes. The modal overlap depends on the size and reverberation time (T_{30}) of the room, and it increases with the square of f .⁸

However, the modal overlap is not necessarily sufficient to assess the sparsity of the sound field. In highly damped rooms, the modal overlap is high, but the sound field is more sparse, as the reflections from walls tend to be attenuated. In this case, the *image source method*²⁰ is more informative on the number of waves that constitute the response, since it models each reflection on the room boundaries with an image source. Inside highly damped rooms, the energy of late reflections drops fast, and not much information is lost when those weak reflections are discarded. Figures 1(a) and 1(b) shows the frequency response of a room with different absorption coefficients $\alpha=0.85$ and $\alpha=0.5$, simulated with the image source method. When $\alpha=0.85$, the frequency response obtained when only the 30 strongest reflections are considered is similar to the frequency response calculated from the 300 strongest reflections; while, when $\alpha=0.5$, reducing the number of reflections results in a completely different frequency response.

Figure 1(c) shows the energy of the 300 strongest reflections sorted from maximum to minimum for the two absorption cases. Let us assume that reflections with energy of less than -30 dB can be discarded and still obtain a good approximation of the frequency response. For the case of $\alpha=0.85$, 30 reflections would be enough to approximate the response. On the other hand, with $\alpha=0.5$, more than 300 reflections would be needed to model the room frequency response. Equation (6) shows that the number of measurements M is proportional to S , therefore, approximately 10 times more measurements are needed in the $\alpha=0.5$ case than in the more damped (and more sparse) $\alpha=0.85$ case. The simulation evidences that the wave field in lightly damped rooms is not sparse ($S \ll N$), thus, the validity of the CS sparsity assumption depends strongly on the acoustic properties of the specific room.

The proposed method relies on the spatial sparsity of the frequency response, and not on the temporal sparsity of reflections in the impulse response. The damping in the room is critical in order to assess the spatial sparsity. The

size of the room has an influence too on the modal count, but it is eventually the damping of the modes what determines the spatial sparsity of the solution. It is worth to note that for other methods relying on the sparsity of reflections over time (e.g., Refs. 11 and 12), the size of the room will play an important role in addition to absorption.

B. Coherence and sampling

For a precise recovery of \mathbf{x} , CS relies on a low coherence of the sensing matrix \mathbf{H} (Refs. 19, 21, and 22) [Eq. (6)]. The mutual coherence $\mu(\mathbf{H})$ is defined as the maximum off-diagonal element of the Gram matrix Γ , in which each element is the normalized inner product of two columns of $\mathbf{H} = [\mathbf{h}_1, \dots, \mathbf{h}_N]$.

$$\mu(\mathbf{H}) = \max_{i \neq j} \Gamma_{ij} = \max_{i \neq j} \frac{\|\mathbf{h}_i^H \mathbf{h}_j\|_2}{\|\mathbf{h}_i\|_2 \cdot \|\mathbf{h}_j\|_2}, \quad (7)$$

where the superindex H indicates the Hermitian transpose.

Equation (1) shows that \mathbf{H} is a function of the frequency, the wave directions, and the microphone positions, with each column \mathbf{h}_i corresponding to a wave direction. If two columns, \mathbf{h}_i and \mathbf{h}_j , are correlated, it will not be possible to determine whether the wave is coming from one direction or the other. In other words, $\mu(\mathbf{H}) = 0$ indicates that \mathbf{H} constitutes an orthonormal basis.¹⁵

Figure 2 shows one column of the Gram matrix Γ_{i1} for three different arrays: a cubic array with equally spaced microphones, an open spherical array with microphones randomly placed on its surface, and a random array with microphones distributed randomly inside the volume of a sphere. Γ_{i1} indicates the correlation of direction $(\phi, \theta)_1 = 0, 0$ with the rest of $N = 1024$ directions result of the uniform sampling of a sphere, where ϕ and θ are the azimuth and elevation angles, respectively. Figure 2 shows high correlation of all directions at 100 Hz. In this case, the wavelength is much larger than the arrays dimension, thus, the pressure at all microphones will be similar. The coherence is slightly lower for the spherical array since its microphones are more distant to one another. At 600 Hz the sampling distance of the cubic array is similar to half of the wavelength, and plane waves with directions perpendicular the cube faces will lead to a maximum correlation, as Fig. 2 shows for directions $(\phi, \theta)_1 = 0, 0$ and $(\phi, \theta)_2 = 0, \pm\pi$. To improve the performance of CS algorithms, it is common to choose random sampling schemes.²³

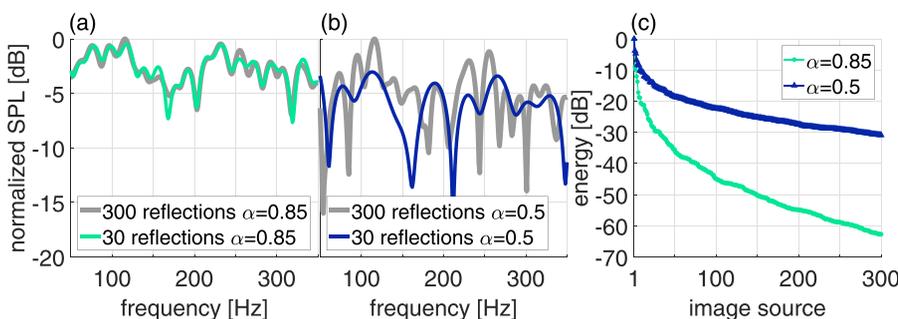


FIG. 1. (Color online) Frequency response of a $7.5 \times 4.7 \times 2.8$ m rectangular room with a source placed at 1.5, 1, 0.7 m and a receiver at 6, 3, 1.8 m simulated via the image source method considering the 30 and the 300 strongest reflections. (a) All the surfaces with $\alpha=0.85$. (b) All the surfaces with $\alpha=0.5$. (c) Energy of the 300 strongest reflections sorted from maximum to minimum energy.

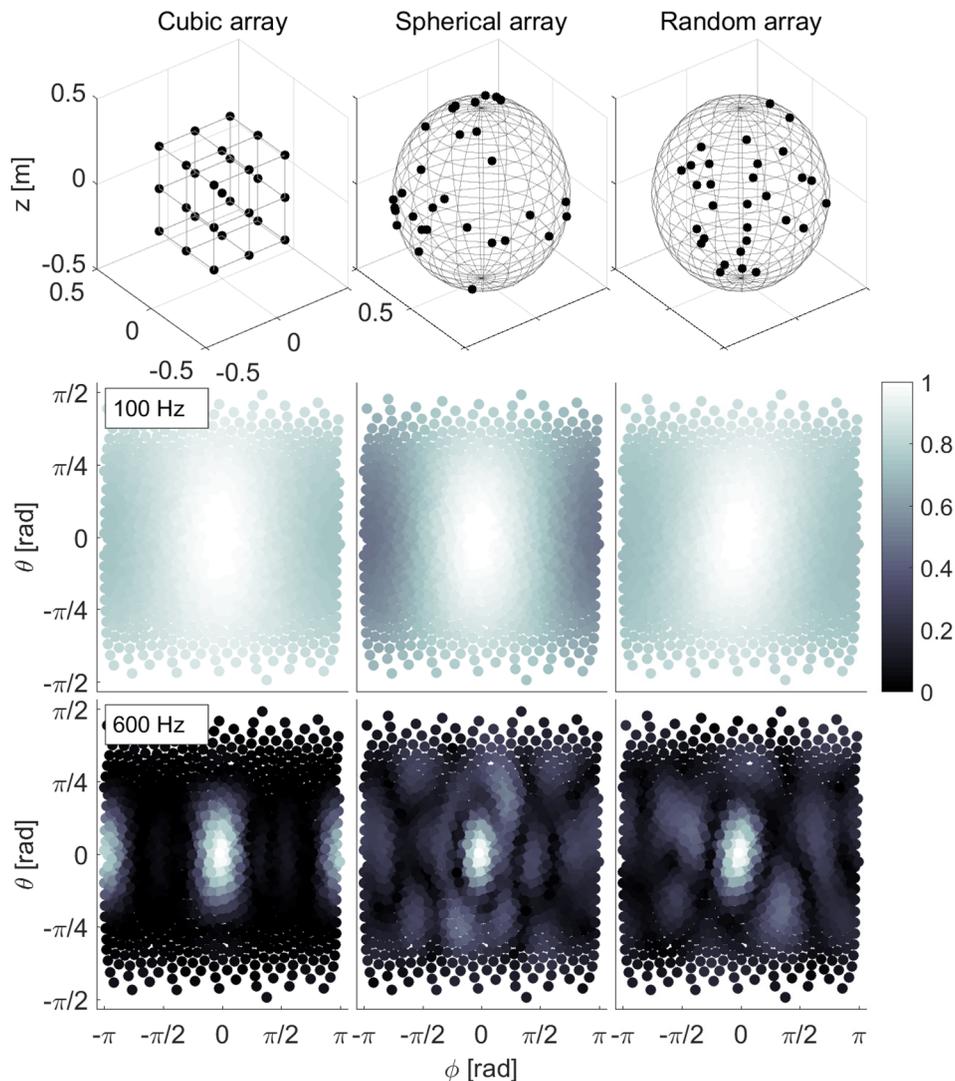


FIG. 2. (Color online) One column of the Gram matrix showing the coherence of one direction $[\phi_1, \theta_1 = (0^\circ, 0^\circ)]$ with other 1023 different directions at 100 Hz (second row) and 600 Hz (third row). Left: cubic array with microphones equally spaced 0.25 m. Center: 27 microphones distributed randomly over the surface of a sphere with radius 0.5 m. Right: 27 microphones taken randomly inside the volume of a sphere with radius 0.5 m.

III. EXPERIMENTAL RESULTS

An experimental study is conducted in two different rooms (a lightly damped and a damped room) to examine the accuracy of the sound field reconstruction. The considerations on coherence discussed in Sec. II B are used as guidelines to design a suitable spatial sampling scheme. A Universal Robots (Odense, Denmark) UR5 robotic arm was used to move a microphone and measure the frequency response at 97 positions on the surface of a sphere with radius $a = 0.5$ m and with its center placed 1.6 m above the floor, as shown in Fig. 3. These points correspond to the uniform sampling of the sphere at 128 points, discarding 31 points near the base of the UR5 that were not accessible due to the presence of the base. The 97 measurements on the sphere were combined with five measurements inside its volume to create an array robust to the zero-crossings of the Bessel function,²⁴ with a total number of $M = 102$ measurements. This array, and a decimated version of it randomly taking $M = 35$ from the complete set, were used to reconstruct the response over a spherical volume of $a_\Omega = 0.7$ m radius. Finally, for the sake of benchmarking, all the response functions were measured in a dense grid of 709 equally spaced points in a circle of radius 0.7 m at the equator of the sphere. This extensive sampling served as a true

reference sound field, so that the reconstruction can be qualitatively and quantitatively compared to it.

Table I shows the dimensions and acoustic properties of the two rooms where the measurements were conducted, with the reverberation time T_{30} given in octave bands. The

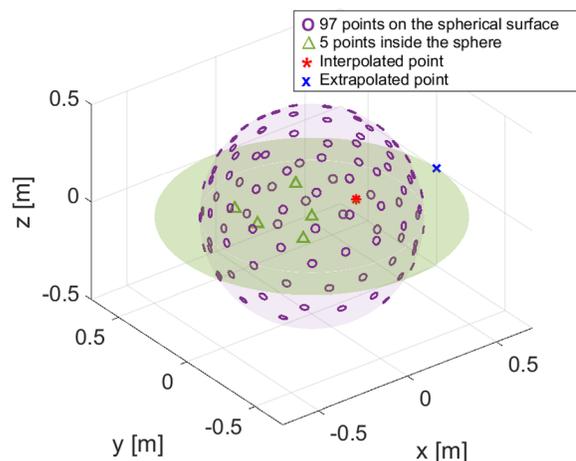


FIG. 3. (Color online) Measurement positions. The shadowed plane represents the reference plane, where the 709 measured points are not shown for the sake of clarity.

TABLE I. Characteristics of the studied rooms.

Room	Dimensions [m]	T_{30} [s] (125, 250, 500, 1k, 2k, 4k Hz)	Absorption	f_{Schro} [Hz]
Lightly damped room	$3.29 \times 2.97 \times 4.38$	3.0, 2.9, 3.1, 2.8, 2.1, 1.6	No: concrete walls, floor, and ceiling	485
Damped room	$7.5 \times 4.74 \times 2.8$	0.5, 0.4, 0.4, 0.4, 0.4, 0.4	Yes: on walls and ceiling, carpet on the floor	133

use of a robotic arm made possible to measure with a single microphone achieving high accuracy on its positioning. During the measurement, the room conditions (such as air temperature) were kept constant to reduce bias. The rooms were excited using a logarithmic sine sweep from 20 to 20 000 Hz with a duration of 10 s. The background noise was also recorded in order to estimate ϵ . The measurement process took a total time of approximately 5 h for the 806 measurements in each room. The equipment used was a Dynaudio (Skanderborg, Denmark) BM6 loudspeaker, a Universal Robots (Odense, Denmark) UR5 robotic arm, and a Brüel&Kjær (Nærum, Denmark) free field microphone.

The number of plane waves of the dictionary N was $N = 700$, making it possible to approximate the sound field up to 2000 Hz [Eq. (2)]. The directions were uniformly sampled on the surface of a sphere. The sampling was based on the Thomson problem for determining the position of electrons on the surface of a sphere by minimizing the system energy.²⁵ The optimization problem [Eq. (4)] was solved making use of an interior point method, with the CVX²⁶ toolbox for convex optimization.

Two measures of the error are used: the normalized mean square error (NMSE):

$$\text{NMSE}_{\text{dB}} = 10 \log_{10} \left[\frac{1}{M_R} \sum_{m=1}^{M_R} \frac{\|\mathbf{P}_{m,:} - \tilde{\mathbf{P}}_{m,:}\|_2^2}{\|\mathbf{P}_{m,:}\|_2^2} \right], \quad (8)$$

where $\mathbf{P}, \tilde{\mathbf{P}} \in \mathbb{C}^{M_R \times N_f}$ are the matrices gathering the measured and estimated pressure, respectively; with M_R the number of positions $\mathbf{r}_R \in \Omega$ in which the sound field is reconstructed, and N_f the number of reconstructed frequencies. The other error measure is the modal assurance criterion (MAC):²⁷

$$\text{MAC}(k_i) = \frac{|\mathbf{P}_{:,i}^H \tilde{\mathbf{P}}_{:,i}|^2}{\left(\mathbf{P}_{:,i}^H \mathbf{P}_{:,i} \right) \left(\tilde{\mathbf{P}}_{:,i}^H \tilde{\mathbf{P}}_{:,i} \right)}. \quad (9)$$

The MAC is a measure of the spatial correspondence of two vectors, ranging from 0 to 1. This is a valuable measure to evaluate the spatial similarity between the estimated reconstruction and the true sound field on the reference plane. High values of MAC will indicate that the reconstructed pressure distribution over the space is consistent with the true one. The NMSE gives an indication of the average error over frequency. These two measures are complementary and serve to evaluate the error on the reference plane, i.e., on a surface of approximately 1.54 m^2 .

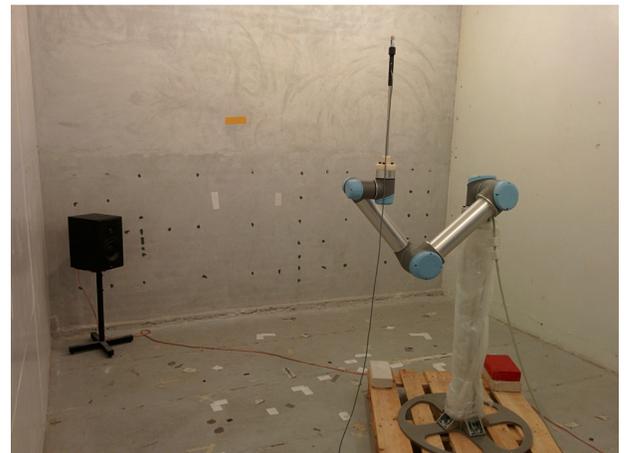
A. Lightly damped room

The lightly damped room, shown in Fig. 4(a), is a rectangular room with very reflective surfaces and no scattering

elements. Although these idealized conditions do not reflect the acoustics of “common” rooms, they constitute a controlled scenario in which the performance of the reconstruction in a highly reverberant environment can be studied. Figure 5 shows the actual and reconstructed frequency responses for one point interpolated inside the array volume $(r, \phi, \theta)_{\text{interp}} = 0.25, 0, 0$, and one point extrapolated outside $(r, \phi, \theta)_{\text{extrap}} = 0.7, 0, 0$ (positions in Fig. 3) using the complete set of measurements ($M = 102$) and the decimated version ($M = 35$). The reconstruction is nearly perfect below 400 Hz, while it degrades towards high frequencies.

Strong modes dominate the response at low frequencies, and the (promoted) sparse solution will contain the largest coefficients, which corresponds to the waves that conform the mode-shapes, as discussed in Sec. II A. Conversely, at higher frequencies, the sound field is the result of many

(a)



(b)

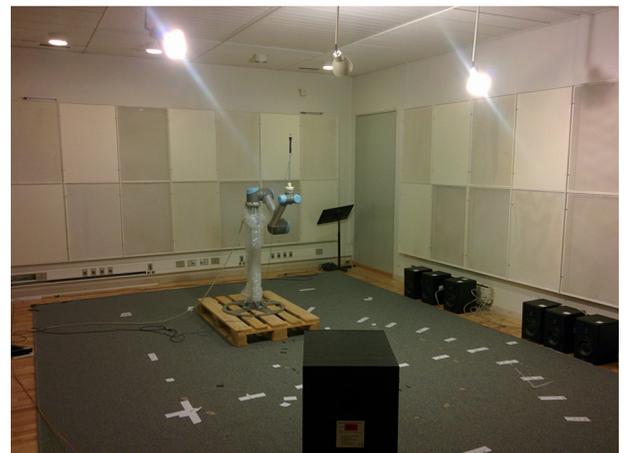


FIG. 4. (Color online) Experimental setup in (a) the lightly damped room and (b) the damped room.

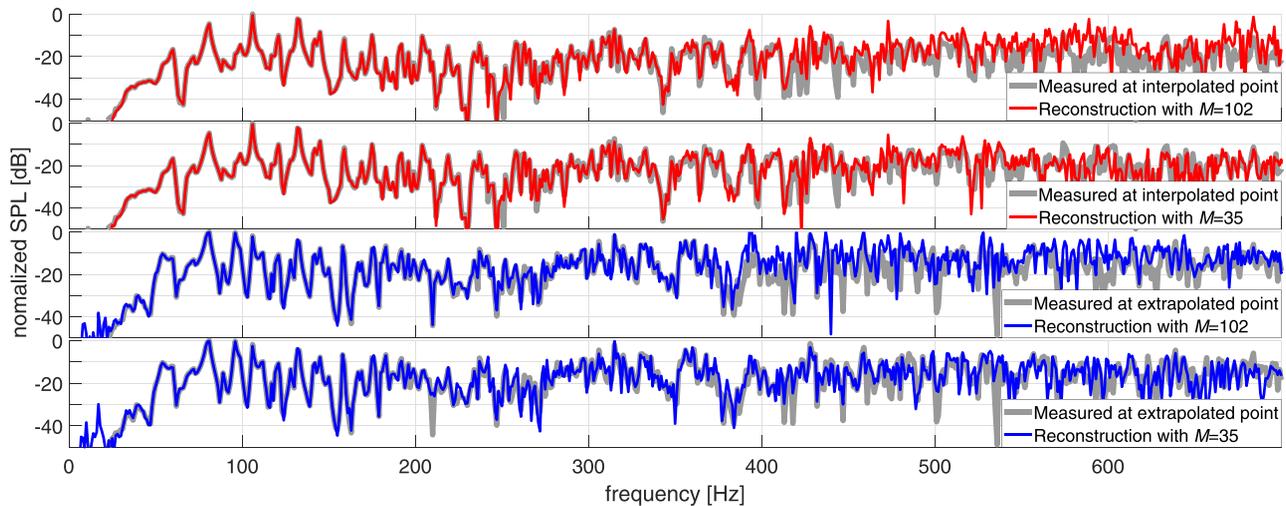


FIG. 5. (Color online) Measured and reconstructed frequency responses for two points (see Fig. 3) in the lightly damped room using $M = 102$ and $M = 35$ measurements.

waves interfering, with no clear dominant components. In this case, since sparsity is still promoted, some of the waves that contribute significantly to the sound field are not found, resulting in a wrong estimation. The frequency response reconstruction on those two single points does not benefit from increasing the number of measurements from $M = 35$ to $M = 102$. The similar reconstruction performance using the complete measurement set and the decimated version suggest that the problem lies on the lack of sparsity of the underlying signal, rather than on its undersampling.

Figure 6 shows the measured and reconstructed impulse response for the interpolation point. The first 200 ms of the impulse response are characterized by a large number of strong reflections. The slow decay of the reflections energy evidences the lack of sparsity of the sound field.

Figure 7 shows the MAC over frequency for the reconstruction on the reference plane. The frequency limit for a nearly perfect spatial agreement of measurement and reconstruction is approximately 400 Hz for $M = 102$, and 200 Hz for $M = 35$. Figure 8 shows the measured and reconstructed spatial pressure distribution on the reference plane at 100, 250, 500, and 800 Hz with $M = 35$. At very low frequencies (< 200 Hz), the sound field is composed by a few waves, and it is rather even on the plane. Therefore, even when \mathbf{H} is coherent, this simple sound field is easy to reconstruct,

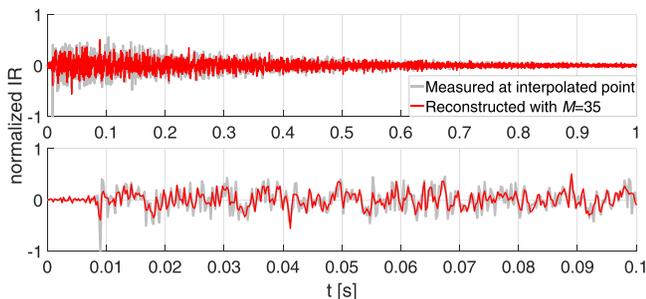


FIG. 6. (Color online) Measured and reconstructed impulse response for one point (see Fig. 3) in the lightly damped room using $M = 35$ measurements.

resulting in high MAC values. In the range of 200–350 Hz, the MAC with $M = 35$ shows a clear degradation. In this frequency range, the wavelength is still larger than the size of the array, resulting in a high coherence. Equation (6) indicates that increasing the coherence μ requires an increase of the number of measurements M . Accordingly, Fig. 7 shows that at this frequency range, the reconstruction using $M = 102$ is improved and the spatial correspondence is nearly perfect. For frequencies above 350 Hz, although the coherence decreases, the sound field is composed by a larger number of waves. The lack of sparsity of \mathbf{x} makes the CS estimation fail. It is worth noting that using $M = 102$, even at 800 Hz (well above the Schroeder frequency of the room $f_{\text{Schro}} = 485$ Hz), the reconstruction leads to MAC values in the range of 0.6–0.8, indicating that the reconstruction is degraded, but not far from correct.

B. Damped room

The damped room, shown in Fig. 4(b), complies with the IEC standard 268-13 and it is typically used for the listening evaluation of loudspeakers. Its walls and ceiling are covered with acoustic absorbers that help to reduce the reverberation time ($V \approx 100 \text{ m}^3$, $T_{60} = 0.4 \text{ s}$), and the acoustic conditions are somewhat closer to the ones that can be commonly found in rooms such as living rooms, classrooms, etc. Figure 9 shows the measured and reconstructed frequency responses at the interpolation and extrapolation points. There is a near

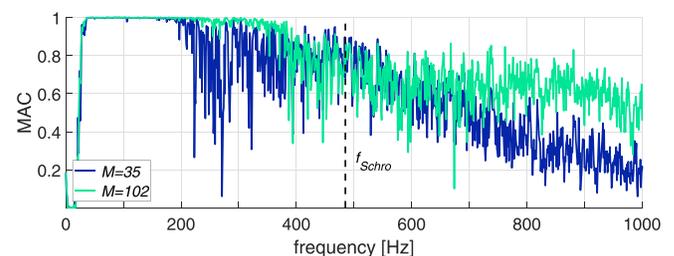


FIG. 7. (Color online) MAC on the reference plane for the lightly damped room using $M = 35$ and $M = 102$ measurements.

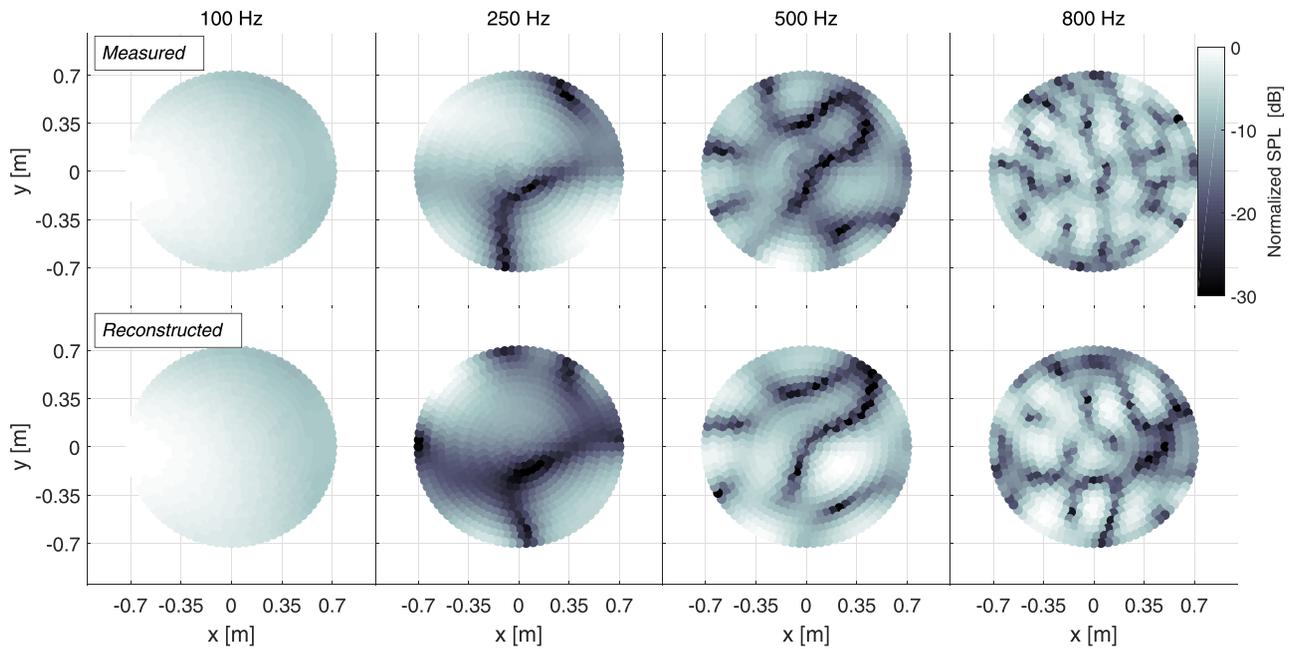


FIG. 8. (Color online) Normalized pressure distribution on the reference plane inside the lightly damped room for different frequencies. Top: measured. Bottom: reconstructed with $M = 35$.

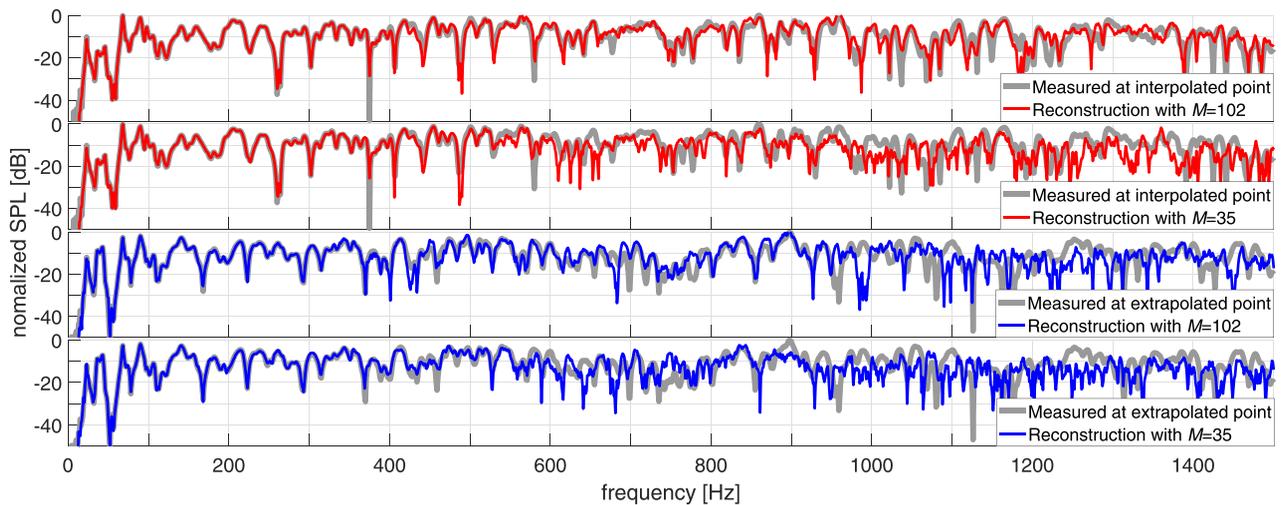


FIG. 9. (Color online) Measured and reconstructed frequency responses for two points (see Fig. 3) in the damped room using $M = 102$ and $M = 35$ measurements.

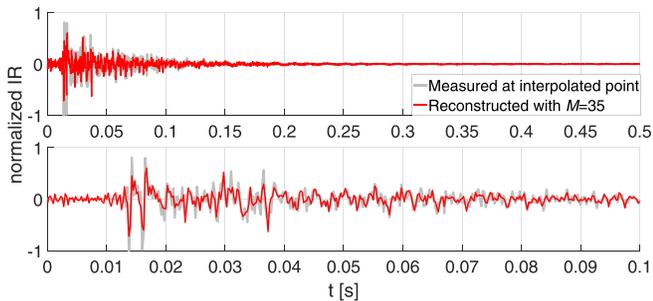


FIG. 10. (Color online) Measured and reconstructed impulse response for one point (see Fig. 3) in the damped room using $M = 35$ measurements.

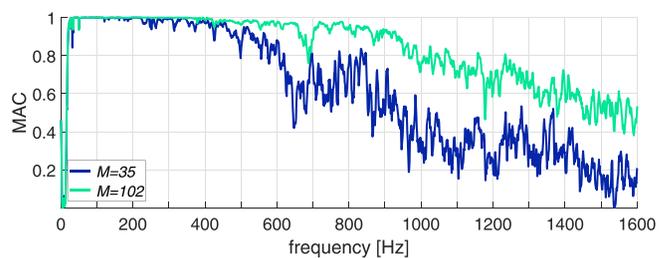


FIG. 11. (Color online) MAC on the reference plane for the damped room using $M = 35$ and $M = 102$ measurements.

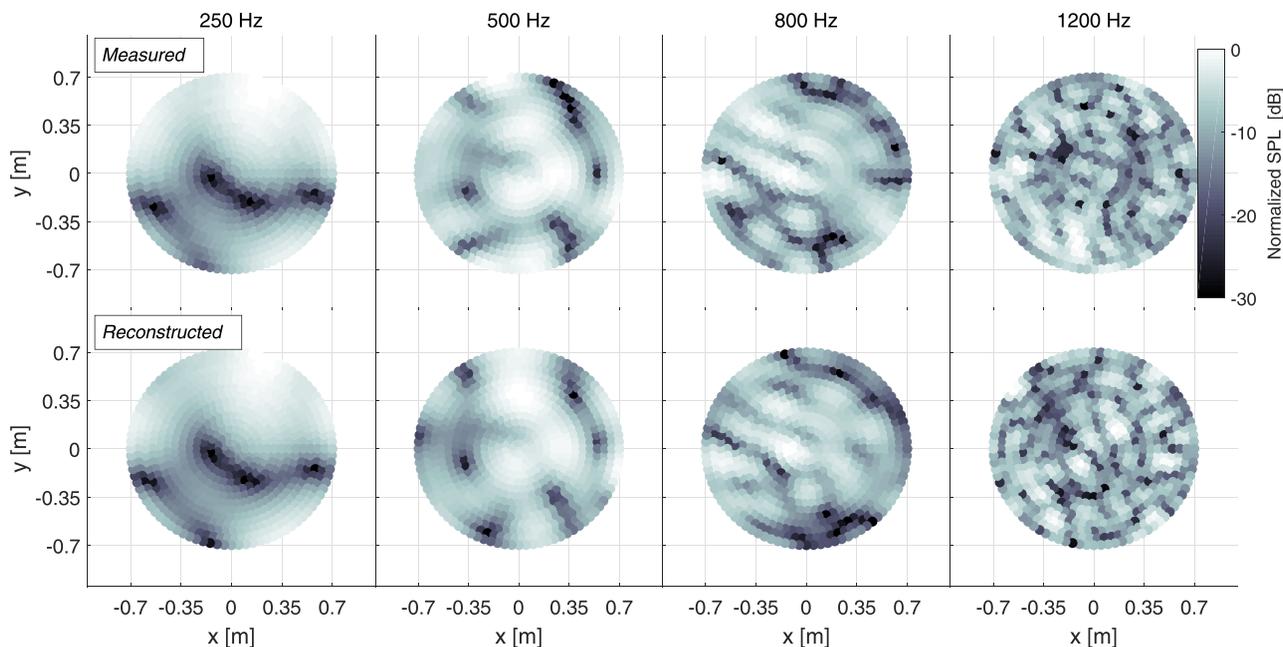


FIG. 12. (Color online) Normalized pressure distribution on the reference plane inside the damped room for different frequencies. Top: measured. Bottom: reconstructed with $M = 35$.

perfect agreement up to approximately 1000 Hz for $M = 102$ and 600 Hz for $M = 35$. For the interpolation point, a good reconstruction is achieved with $M = 102$ even up to 1500 Hz.

Figure 10 shows the measured and reconstructed impulse response for the interpolation point. The direct sound and a single strong reflection can be seen in the first 20 ms, and the following reflections are highly attenuated. Therefore, the response is primarily composed by a few waves, and consequently, the assumption of spatial sparsity of the wave field inside the room holds. These results indicate that in the case of damped rooms, the spatial sparsity assumption holds even at high frequencies (as illustrated in Sec. II A), thus, the frequency response can be accurately reconstructed if enough measurements are performed.

Figure 11 shows the MAC over frequency in the damped room when reconstructing on the reference plane. High values of MAC are achieved up to 1000 Hz with $M = 102$, and 600 Hz with $M = 35$. Figure 12 shows the pressure distribution on the reference plane at 250, 500, 800, and 1200 Hz using $M = 35$ measurements. At 250 Hz, the correspondence between the reconstructed sound field and the true reference field is nearly perfect. Although the coherence of the sensing matrix \mathbf{H} is high at low frequencies, the underlying sound field is approximately sparse, leading to a perfect spatial reconstruction. At 500 and 800 Hz the sound field is more complex, being composed by a larger number of waves. At those frequencies, the reconstruction is more challenging due to the rapid spatial changes near nodal planes. Nonetheless, the general pressure distribution on the reference plane is still well recovered. The sound field at high frequencies (1200 Hz) presents large pressure variations within small regions of space, resulting in a wrong spatial reconstructions and lower spatial correspondence.

C. Room comparison

A direct comparison of the rooms studied is valuable to examine the use of compressive sensing for reconstructing the sound field in a room, and to assess its applicability to different rooms. Figure 13 shows the mean over frequency of the estimated coefficients $\tilde{\mathbf{x}}_{\mathbf{f}}(\phi, \theta) = (1/N_f) \sum_{i=1}^{N_f} |\tilde{\mathbf{x}}(f_i, \phi, \theta)|$ inside the two rooms for the $N = 700$ wave directions in the dictionary. The average is done from 0 to 2000 Hz, the estimation $\tilde{\mathbf{x}}$ is based on $M = 35$ measurements, and $\tilde{\mathbf{x}}_{\mathbf{f}}$ is normalized to evaluate values between 0 and 1 ($\tilde{\mathbf{x}}_{\mathbf{f}\text{norm}} = \tilde{\mathbf{x}}_{\mathbf{f}}/\max|\tilde{\mathbf{x}}_{\mathbf{f}}|$). In the damped room, coefficients corresponding to the direction $(\phi, \theta)_1 \approx -3\pi/4, -\pi/8$ show high amplitude $\tilde{\mathbf{x}}_{\mathbf{f}\text{norm}}(\phi_1, \theta_1) \approx 1$, while for the rest of directions the amplitude is close to zero. The direction $(\phi, \theta)_1$ corresponds to the direct sound, and illustrates the marked overall sparsity of the sound field. On the other hand, in the lightly damped room, many of the coefficients (averaged through frequency) have an amplitude

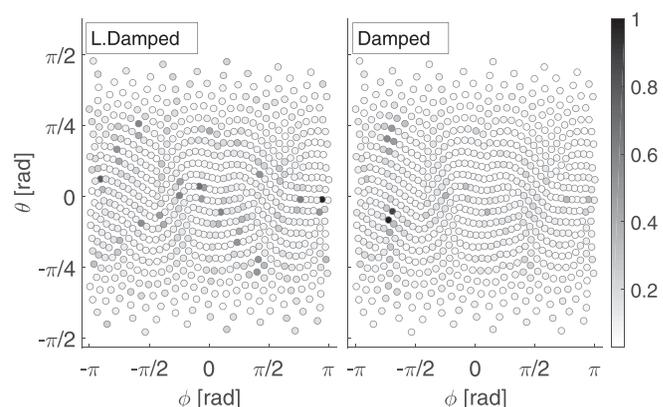


FIG. 13. Normalized mean over frequency of the estimated wave amplitudes $\tilde{\mathbf{x}}$ from the reconstruction with $M = 35$ in the lightly damped and damped rooms.

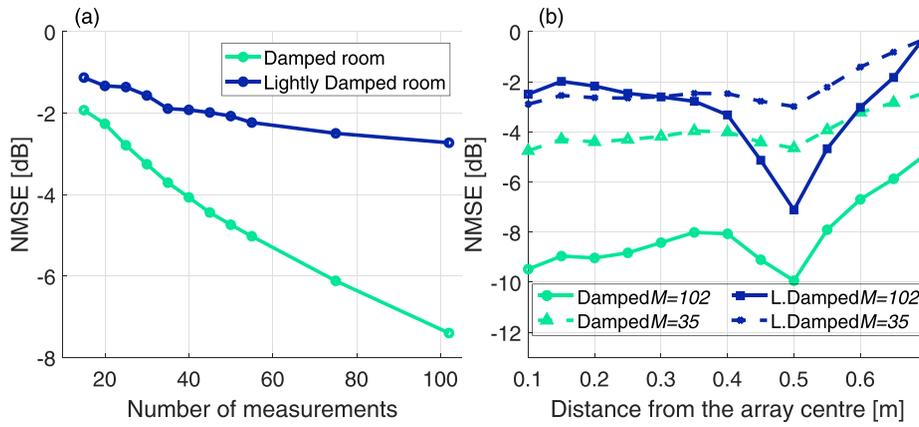


FIG. 14. (Color online) (a) NMSE with increasing number of measurements in the lightly damped and damped rooms. (b) NMSE over distance from the array centre using $M = 35$ and $M = 102$ measurements in the lightly damped and damped rooms.

different from zero, indicating that the overall sound field consists of waves arriving from many different directions caused by strong reflections present throughout the frequency range.

Figure 14(a) shows the NMSE calculated for different number of measurements M used in the reconstruction. While the performance increases linearly when increasing the number of measurements in the damped room, the error is steady for the lightly damped room. In highly reverberant environments, there are very rapid spatial changes due to the strong wave interference. Those large interference depths are difficult to reconstruct even if the number of measurements is increased. In damped rooms spatial changes are not as pronounced, the responses are more even over frequency and space, and increasing the number of microphones improves the spatial reconstruction, as indicated in Fig. 14(a).

Figure 14(b) shows the NMSE calculated from 0 to 2000 Hz at points equally distanced from the array centre. The error inside the array volume (distance < 0.5 m) is rather constant. There is a drop on the array surface ($a = 0.5$ m) since most of the measurements are concentrated in that region. For extrapolated points (distance > 0.5 m) the error increases linearly. The improvement achieved by increasing the number of measurements in the damped room does not occur in the lightly damped room.

IV. CONCLUSIONS

In this study a method based on compressive sensing to reconstruct the frequency response at arbitrary points inside a room has been examined. The accuracy of the method has been studied qualitatively and quantitatively, based on measurements with a scanning robotic arm. Two central aspects of CS have been investigated: the assumption of spatial sparsity of the underlying sound field (which depends on the damping of the room), and the assumption of a sensing matrix with low (ideally zero) column coherence. The results show that the assumption of spatial sparsity holds in damped rooms, and a perfect reconstruction can be achieved from a limited number of spatial samples. In this study, 102 measurements in a volume of $\approx 0.52 \text{ m}^3$ were sufficient to perfectly reconstruct the sound field over a volume of $\approx 1.43 \text{ m}^3$ up to 1 kHz. In strongly reverberant spaces, the lack of spatial sparsity leads to a poor reconstruction of the sound field, except at low frequencies, where the modal density is low. In fact, the study shows that at low frequencies, in spite

of the high column coherence, the reconstruction of sound fields is successful due to the lower modal overlap, i.e., the greater sensing matrix coherence is compensated by the fact that the underlying sound field is spatially sparse. The proposed volumetric reconstruction has promising prospects in room compensation for audio reproduction systems in non-anechoic conditions. This is particularly relevant at low frequencies, where the proposed method is especially successful. The method can make it possible to reconstruct the spatial sound field at low frequencies with many less measurements than currently used. All in all, the results show the validity and potential of CS-based approaches for reconstructing sound fields over space inside enclosed spaces.

ACKNOWLEDGMENTS

The authors would like to thank to Morten Birkmose and Christoffer Klærke from Goertek EU, as well as Dr. Finn T. Agerkvist for their valuable suggestions and helpful advice.

- ¹J. Mourjopoulos, "On the variation and invertibility of room impulse response functions," *J. Sound Vib.* **102**(2), 217–228 (1985).
- ²B. D. Radlović, R. C. Williamson, and R. A. Kennedy, "Equalization in an acoustic reverberant environment: Robustness results," *IEEE Trans. Audio, Speech, Lang. Process.* **8**(3), 311–319 (2000).
- ³S. Spors, H. Buchner, R. Rabenstein, and W. Herboldt, "Active listening room compensation for massive multichannel sound reproduction systems using wave-domain adaptive filtering," *J. Acoust. Soc. Am.* **122**(354), 354–369 (2007).
- ⁴T. Betlehem and T. D. Abhayapala, "Theory and design of sound field reproduction in reverberant rooms," *J. Acoust. Soc. Am.* **117**(4), 2100–2111 (2005).
- ⁵F. M. Heuchel, E. Fernandez-Grande, F. T. Agerkvist, and E. Shabalina, "Active room compensation for sound reinforcement using sound field separation techniques," *J. Acoust. Soc. Am.* **143**(3), 1346–1354 (2018).
- ⁶D. S. Talagala, W. Zhang, and T. D. Abhayapala, "Efficient multi-channel adaptive room compensation for spatial soundfield reproduction using a modal decomposition," *IEEE/ACM Trans. Audio, Speech, Lang. Process.* **22**(10), 1522–1532 (2014).
- ⁷T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *IEEE Trans. Signal Process.* **54**(10), 3790–3804 (2006).
- ⁸F. Jacobsen and P. M. Juhl, "Sound in enclosures," in *Fundamentals of General Linear Acoustics* (Wiley, London, 2013), Chap. 8.
- ⁹Y. Haneda, Y. Kaneda, and N. Kitawaki, "Common-acoustical-pole and residue model and its application to spatial interpolation and extrapolation of a room transfer function," *IEEE Trans. Audio, Speech, Lang. Process.* **7**(6), 709–717 (1999).
- ¹⁰R. Mignot, G. Chardon, and L. Daudet, "Low frequency interpolation of room impulse responses using compressed sensing," *IEEE/ACM Trans. Audio, Speech, Lang. Process.* **22**(1), 205–216 (2014).

- ¹¹R. Mignot, L. Daudet, and F. Ollivier, “Room reverberation reconstruction: Interpolation of the early part using compressed sensing,” *IEEE Trans. Audio, Speech, Lang. Process.* **21**(11), 2301–2312 (2013).
- ¹²N. Antonello, E. D. Sena, M. Moonen, P. A. Naylor, and T. van Waterschoot, “Room impulse response interpolation using a sparse spatio-temporal representation of the sound field,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.* **25**(10), 1929–1941 (2017).
- ¹³W. Jin and W. B. Kleijn, “Theory and design of multizone soundfield reproduction using sparse methods,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.* **23**(12), 2343–2355 (2015).
- ¹⁴P. Samarasinghe, T. Abhayapala, M. Poletti, and T. Betlehem, “An efficient parameterization of the room transfer function,” *IEEE/ACM Trans. Audio, Speech, Lang. Process.* **23**(12), 2217–2227 (2015).
- ¹⁵E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, “Compressed sensing with coherent and redundant dictionaries,” *Appl. Comput. Harmon. Anal.* **31**, 59–73 (2011).
- ¹⁶A. Moiola, R. Hiptmair, and I. Perugia, “Plane wave approximation of homogeneous Helmholtz solutions,” *Z. Angew. Math. Phys.* **62**, 809–837 (2011).
- ¹⁷E. G. Williams, *Fourier Acoustics—Sound Radiation and Nearfield Acoustical Holography* (Academic, New York, 1999), Chaps. 2 and 6.
- ¹⁸T. Nowakowski, J. de Rosny, and L. Daudet, “Robust source localization from wavefield separation including prior information,” *J. Acoust. Soc. Am.* **141**(4), 2375–2386 (2017).
- ¹⁹E. J. Candès and M. B. Wakin, “An introduction to compressive sampling,” *IEEE Signal Process. Mag.* **25**(2), 21–30 (2008).
- ²⁰J. B. Allen and D. A. Berkley, “Image method for efficiently simulating small-room acoustics,” *J. Acoust. Soc. Am.* **65**(4), 943–950 (1979).
- ²¹E. Fernandez-Grande, A. Xenaki, and P. Gerstoft, “A sparse equivalent source method for near-field acoustic holography,” *J. Acoust. Soc. Am.* **141**(1), 532–542 (2017).
- ²²A. Xenaki, P. Gerstoft, and K. Mosegaard, “Compressive beamforming,” *J. Acoust. Soc. Am.* **136**(1), 260–271 (2014).
- ²³C. F. Gaumont and G. F. Edelmann, “Sparse array design using statistical restricted isometry property,” *J. Acoust. Soc. Am.* **134**(2), 191–197 (2013).
- ²⁴G. Chardon, W. Kreuzer, and M. Noisternig, “Design of spatial microphone arrays for sound field interpolation,” *IEEE J. Select. Topics Signal Process.* **9**(5), 780–790 (2015).
- ²⁵A. Semechko, “Particle sample sphere code for uniform tessellation of a unit sphere,” <http://www.mathworks.com/matlabcentral/fileexchange/37004-uniform-sampling-of-a-sphere> (Last viewed April 1, 2017).
- ²⁶M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming, version 2.1,” <http://cvxr.com/cvx> (Last viewed April 1, 2017).
- ²⁷M. Pastor, M. Binda, and T. Harcarik, “Modal assurance criterion,” *Procedia Eng.* **48**, 543–548 (2012).