Sound Field Reconstruction In Low-Frequency Room Acoustics: A Benchmark Study With Simulation

T. Pham Vu, E. Rivet, H. Lissek
Acoustic Group, Signal Processing Laboratory LTS2, EPFL, Lausanne, Switzerland

1. **Introduction**

Sound field reconstruction is the experimental process of reconstructing the sound field distribution in an open or confined space. For room acoustics, sound field reconstruction consists of recovering the entire enclosed sound field in a room using various measurements. This characterization process requires a thorough knowledge of either the Room Frequency Response (RFR) or its time domain equivalent, Room Impulse Response (RIR). These quantities express the acoustic transmission characteristics between a source and a receiver in a room and hence are dependent on the positions of the sound source and receivers in the room. For a fixed source position, the reconstruction of sound field in a room requires measurements at multiple locations in the room. The challenge to this process is how to faithfully reconstruct the sound field using the least possible measurements. To this end, the problem of sound field reconstruction can be studied under the scope of a sampling problem.

An accurate reconstruction of sound fields in rooms using a regular space and time sampling would generally result in a dense microphones placement. In [1], it has been shown that through analyzing the sparse 4D spectrum of the Plenacoustic Function (PAF), a general sampling regime can be deduced to acquire the recovery of the PAF with a limited number of measurements. Recently, it has been shown in [2] that by observing certain sparsity in low frequency room acoustics, ones can reduce the number of RIR measurements even more under the concept of Compressive Sensing (CS) which is a series of techniques that exploit sparsity to recover signals using the least measurements possible. Multiple experiments and simulations have been performed to analyze the accuracy and efficiency of the compressive sensing framework in interpolating the RIRs[2]. Many of these tests are towards the interpolation of RIRs neighboring the center of the arrays.

In this paper, we focus on using this aforementioned Compressive Sensing framework to obtain the low frequencies acoustic information of a room under a more expansive point of view, where the spatialization of sound pressure in a general non-rectangular room can be recovered and visually analyzed using numerical simulations. The framework can be tested and analyzed through different room settings by varying the parameters in COMSOL Multiphysics® software, especially in terms of absorbing characteristics. Through these analysis, the performance of the CS framework can be assessed to verify its validity and robustness under different circumstances.

The outline of the paper is as follows. First, we tackle the governing equations regarding low-frequency room acoustics in Section 2. From there, different sparse properties that are present can be extracted and used as the foundation for the Compressive Sensing framework. The reconstruction framework composed of the Simultaneous Orthogonal Matching Pursuit (SOMP) and Least-square estimation method can then be explained in Section 3. Section 4 is dedicated towards the validation of the framework in reconstructing the sound field in the room using simulation. First, the configuration in COMSOL is explained, focusing on the input and output as well as the models and studies used. Then, the CS framework can be analyzed by comparing different aspects of the reconstruction result for a lightly damped room to the ones obtained from the simulation. It is worth noting that under our framework, it is possible to reconstruct the sound pressure distribution in the room at any particular frequency of interest. The robustness of the framework can further be studied by extending the comparison to other room settings by varying the absorbing characteristics of the walls. Finally, some concluding remarks are presented in Section 5.

2. Sparsities in room acoustics

A. Modal decomposition

The acoustic wave equation in a room at low frequencies can be decomposed as a discrete sum of damped harmonic eigenmodes:

$$p(t,\vec{X}) = \sum_{n\epsilon\mathbb{Z}^*} A_n \Phi_n(\vec{X}) g_n(t) \qquad (1)$$

where Φ_n is the mode shape (eigenfunction), $g_n(t)$ is the decaying function and A_n is the corresponding complex expansion coefficient of mode n. Each eigenmode n is uniquely represented by a complex wavenumber $k_n = (\omega_n + j\delta_n)/c_0$ with ω_n as the angular frequency and $\delta_n > 0$ as the modal damping. The harmonic decaying function $g_n(t)$ can then be expressed as:

$$g_n(t) = e^{jk_n c_0 t} = e^{j(\omega_n + j\delta_n)t} = e^{j\omega_n t} e^{-\delta_n t} \quad (2)$$

Note that $\vec{X}$ is the position of the receiver which is present in (1) while the source position is hidden but included in the complex coefficients A_n's and will remain so for the rest of our derivation. This is due to the fact that in here, we only consider the set-up with one fixed source in the room and the objective is to reconstruct the whole sound pressure field within this limiting case.

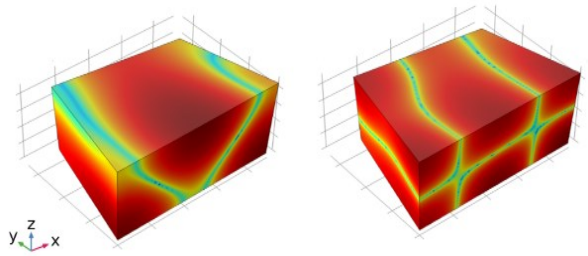

Figure 1. Typical room modes for a non-rectangular room

The mode shape function, on the other hand, is a space dependent function. Assuming the walls are ideally rigid, each mode shape function Φ_n is the exact solution of the Helmholtz equation [3]:

$$\Delta\Phi_n + k_n^2 \Phi_n = 0 \quad (3)$$

In the case of a room with non-rigid walls, the above equation is assumed to hold for receiver's positions not too close to the wall or in the case of rooms with low wall damping.

B. Mode shape approximation

From (1), it can be seen that the mode shape functions Φ_n's need to be linked directly to the position of the receiver $(\vec{X})$ in order to produce a closed-form expansion of the wave equation. It has been shown in [4] that using a combination of spherical harmonics and spherical Bessel functions, the mode shape functions can be approximated as a finite sum of plane waves with the same wavenumber pointing in various directions. Each individual mode shape, can then be formulated using the R-th order approximation:

$$\Phi_n(\vec{X}) \approx \sum_{r=1}^{R} B_{n,r} e^{j\vec{k}_{n,r}\cdot\vec{X}} \quad (4)$$

where $\vec{k}_{n,r}$'s are the wavevectors sharing the same wavenumber $\|\vec{k}_{n,r}\|_2 = |k_n|$. It can be seen later that for rooms with random shape, it is best to provide $\vec{k}_{n,r}$'s as the results of spherically sampling a sphere which radius equals to $|\omega_n/c_0|$. Equation (4) not only gives a good approximation for Φ_n, but more importantly, makes it possible to have a closed-form interpretation of the mode shape function independent on the nature of the modes. Assuming now that for a chosen frequency limit, an R-th order approximation would be enough to closely depict every mode shape function within this limit, using (2) and (4), the equation in (1) could be expanded as:

$$p(t, \vec{X}) = \sum_{n,r} C_{n,r} e^{j\omega_n t} e^{-\delta_n t} e^{j\vec{k}_{n,r}\cdot\vec{X}} \quad (5)$$

Where $C_{n,r} = A_n B_{n,r}$. Through a series of derivations, we now can interpret the equation in (1) as a closed-form finite sum of space-time damped harmonics with the expansion coefficients $C_{n,r}$'s. This expansion form of the equation could be in fact interpreted in multiple ways under the matrix form, which is ideal for the CS framework in the next section.

3. Compressive sensing framework

3.1 The inputs

The algorithm in this research tackles a general case with a non-rectangular room and hence the modal behavior of the room would be more difficult to predict (Figure 1). Inside this room, M microphones are randomly spread in order to acquire the RIR measurements. These unprocessed RIRs can then go through a low pass filter and be down-sampled based on the frequency range of interest in order to reduce the computational cost later on in the algorithm. Calling N_t the length of the time vector for these processed RIRs, we will end up with an $(N_t \times M)$ matrix $\boldsymbol{S}$ of signals which will be taken as the input of our algorithm.

3.2 The outputs

The output of the algorithm should be all the unknowns in (5) except some pre-chosen parameters such as the number of modes N and the wavevectors $\vec{k}_{n,r}$. Therefore, the outputs include the angular eigenfrequency ω_n and damping δ_n of each mode as well as the $N \times R$ expansion coefficients $C_{n,r}$. With all these quantities found, we can interpolate the RIRs at

any position $\vec{X}_{int}$ by simply plugging it into equation (5).

3.3 Algorithm

The algorithm consists of two parts. The first part is making use of the Matching Pursuit method to find the modal wavenumbers k_n's for the N room modes. With these results, the second part focuses on using the least-squares method to approximate the expansion coefficients $C_{n,r}$ for a set of predefined wavevectors $\vec{k}_{n,r}$.

A. Wavenumber identification

This step is based on the Simultaneous Orthogonal Matching Pursuit method [5] (SOMP) for damped sinusoids [6]. From a redundant set of damped sinusoids, it estimates the ones that are highly matched with the matrix of input signals. At the beginning, two sets of $\omega_{min} < \omega < \omega_{max}$ and $\delta_{min} < \delta < \delta_{max}$ can be formed. The range of these sets are roughly defined based on accessible knowledge on the room such as its reverberation time and the geometrical properties. Forming all possible combination between the entries from these two sets will produce an overly redundant set of complex components $(j\omega_q - \delta_q)$ in which $q \in [1, Q]$ with Q as the total number of possible combinations. Note that only a small number of entries from this set are the correct eigenfrequency and damping of the modes in the room, hence the term 'overly redundant'. Each entry of this set is then used to form a time vector of length N_t of damped sinusoid $\theta_q = e^{j\omega_q t}e^{-\delta_q t}$. Using the normalized vectors $\bar{\theta}_q = \theta_q / \|\theta_q\|_2$ as column vectors, we have an $(N_t \times Q)$ matrix $\bar{\Theta}$.

The algorithm, in short, repetitively performs a pole searching procedure in loops. Each loop begins with an $(N_t \times M)$ residue matrix $\boldsymbol{R}_i$. At the first loop, $\boldsymbol{R}_1$ is set to be equal to the predefined signal matrix $\boldsymbol{S}$. Through the pole searching procedure, a damped sinusoid with the highest correlation to the residue matrix (represents a pair of ω_n and δ_n) is chosen. The residue matrix $\boldsymbol{R}_{i+1}$ of the next loop can then be formed by extracting the contribution of this chosen sinusoid from $\boldsymbol{R}_i$. The stages of a loop are listed below:

- Define a $(Q \times M)$ correlation matrix $\Xi_i = |\bar{\Theta}^H \boldsymbol{R}_i|$. Each row indexed q of Ξ_i represents a set of M correlation values between the q^{th} normalized damped sinusoid with each of the M measurements. Summing up the energy of this set of values gives an evaluation correlation value σ_q between the q^{th} damped sinusoid and the entire set of measurements: $\sigma_q = \sum_{m=1}^{M} \left(\Xi_{i[q,m]}\right)^2$. Out of the Q values of σ_q, we choose the maximum one, which points us to the pole with the highest correlation to the measurements. Consequently, its index (namely, q_i) provides the resulting chosen modal wavenumber of this loop which is $k_i = (\omega_{q_i} + j\delta_{q_i})/c_0$.
- After finding a pole, following the orthogonalization and projection of SOMP in [5], the residue matrix for the next loop can be interpreted as $\boldsymbol{R}_{i+1} = \boldsymbol{R}_i - \mathcal{P}_i \boldsymbol{R}_i$ in which $\mathcal{P}_i$ is the projection onto the chosen damped sinusoidal.
- Repeat with $i = i + 1$ until $i = N$

Consequently, we have a group of complex wavenumbers that corresponds to the eigenmodes of the room. It is worth noting that this is not the only way to identify the complex eigenfrequencies from the measurements and it is possible to replace this steps with any other pole searching [7].

B. Projection onto wavevectors

As we have found the ω_n's and δ_n's given in equation (5). The remaining parameters are the expansion coefficients $C_{n,r}$'s.

- The first stage consists of separating the current known and unknown parameters in terms of matrices. The temporal terms in (1) are now known and can be separated. Writing in matrix form with regards to the measurement matrix $\boldsymbol{S}$, we have: $\boldsymbol{S}^{\mathrm{T}} = \Psi G$ with G is the $(N \times N_t)$ matrix with each of its row a damped sinusoidal $g_n(t) = e^{j\omega_n t}e^{-\delta_n t} = e^{jk_n t}$ and Ψ is the $(M \times N)$ space-dependent matrix of modes which includes the expansion coefficients A_n's that appear in (1):

$$\Psi_{[m,n]} = A_n \Phi_n(\vec{X}_m) \tag{6}$$

  with $\vec{X}_m$'s the M position vectors for the input measurements of $\boldsymbol{S}$. One way of looking at (6) is that if $N_t > N$ (which usually is the case), it creates an over-determined matrix problem with $(M \times N)$ unknowns and $(M \times N_t)$ equations. Hence, we can estimate Ψ by computing the least-squares estimation:

$$\Psi \approx \boldsymbol{S}^{\mathrm{T}} G^H (G G^H)^{-1} \tag{7}$$

- Based on the derivation in (5) we can further expand Ψ using plane waves expansion:
  - First, using spherical sampling, for each wavenumber, we create a set of R wavevectors

$\vec{k}_{n,r}$ whose lengths and directions correspond to a uniform sampling of a sphere with radius $|\omega_n/c_0|$ (according to [2],[8], a value of $R \approx 3M/4$ is sufficient to avoid both over and under-fitting).

- Each column ψ_n of the matrix Ψ can now be tackled individually as they belong to different modes. If ρ_n is specified as the $(M \times R)$ matrix of the plane wave harmonics for mode n in which $\rho_{n[m,r]} = e^{j\vec{k}_{n,r}\cdot\vec{x}_m}$, then we can express each column vector ψ_n as:

$$\psi_n = \rho_n \boldsymbol{C}_n \tag{8}$$

where $\boldsymbol{C}_n$ is the $(R \times 1)$ vector containing the R expansion coefficients $C_{n,r}$ of mode n. With $M > R$ as previously chosen, taking ρ_n as the basis, we can perform a least-squares projection of ψ_n onto this basis to construct the coefficient vector $\boldsymbol{C}_n$:

$$\boldsymbol{C}_n \approx (\rho_n^H \rho_n)^{-1} \rho_n^H \psi_n \tag{9}$$

Repeating the procedure across every existing mode $n \leq N$ will return all the expansion coefficients required for reconstruction.

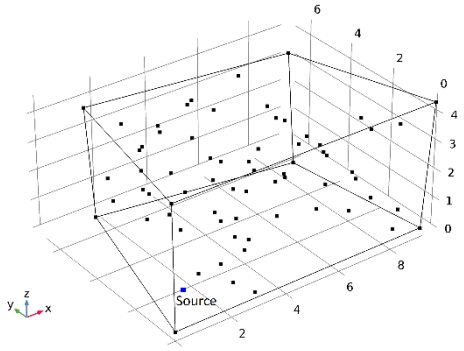

Figure 2. Geometry of the FEM model. The black dots represent the measurement points.

4. Numerical Validation

In this section, an FEM model of a non-rectangular room is first introduced for numerical analysis. The simulation results are then used as the inputs for the reconstruction framework suggested in the previous section. At the same time, they can also be used as a reference to validate the reconstruction result. The validation of the framework starts with the case of a lightly damped room by comparing the simulated sound field with the reconstructed ones by the CS framework. Then, the damping properties of the room is increased to make the problem more challenging for the algorithm. Further comparisons can then be conducted to understand the robustness of the framework.

4.1 Numerical simulation

COMSOL Multiphysics® simulation software is used for both creating the input measurements for the CS framework as well as for the evaluation of sound field that follows. In the ideal case, the input should be the RIRs measured at different positions in the room. However, as the algorithm tackles modal properties in the room, the input could as well be chosen to be the response of a broad-band excitation that can be defined in terms of the frequency domain. Our study is done using the Pressure Acoustics, Frequency Domain module from COMSOL® software. The model is a non-rectangular room with maximum height of $4.5m$, maximum width of $9.6m$ and maximum length of $6.8m$ which replicates the actual reverberation chamber of our laboratory. The walls are set to be non-resonating (that is to have no imaginary part in its impedance) and have the same absorption coefficient of $\alpha = 0.01$. The source is fixed near one corner of the room and set as a mono-pole point source which emits a broad-band excitation. The last step is to spread out the receivers randomly in the room. Knowing the room geometries, we use Matlab to generate M measurement points inside the room with the walls coordinates used as the limiting borders. These points data are then saved into a .dxf file that can be imported into the room model using the CAD Import Module. Figure 2 shows the geometries of the room as well as an example of spreading 50 measurement points inside the room. From the study, we obtain the band-limited frequency responses at each of those measurement points. These responses can then be transformed into the time domain using the Fourier transform and may also go through a low-pass filter to help the algorithm focus more on the low frequency regions. These time domain signals will serve as the input of our algorithm.

In our CS framework, we use a total of 50 measurement points. The frequency range is chosen to be below 80 Hz where around 20 eigenmodes exist within this range. The processed time signals will go through the series of algorithm detailed in the compressive sensing framework and will return 25 complex wavenumbers along with the gains for the 37 $(\approx 3M/4)$ wavevectors for each of these wavenumbers. The reason we use 25 as the number of modes is to be consistent with the practical case where an exact number of modes under a chosen frequency limit could not usually be precisely predicted beforehand, especially in case of a non-rectangular room.

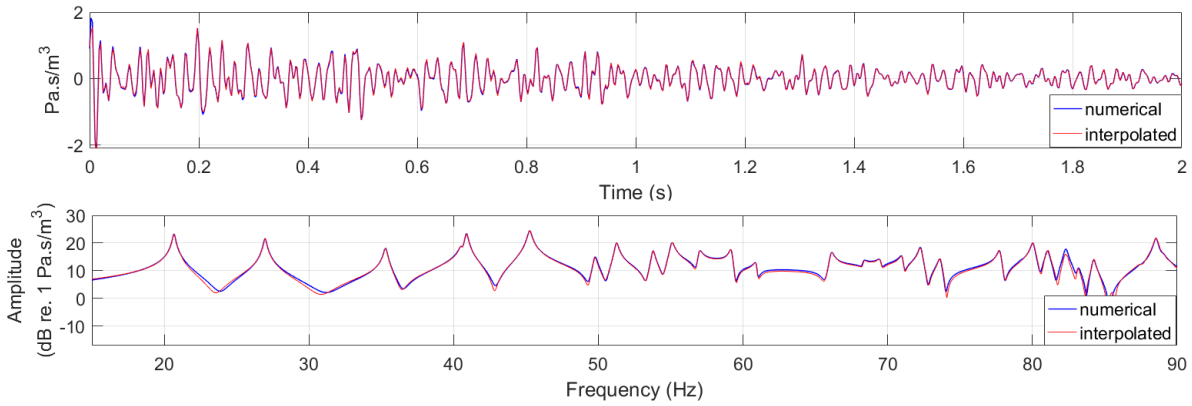

Figure 3. Interpolation of the RIR and RFR for a point inside the room.

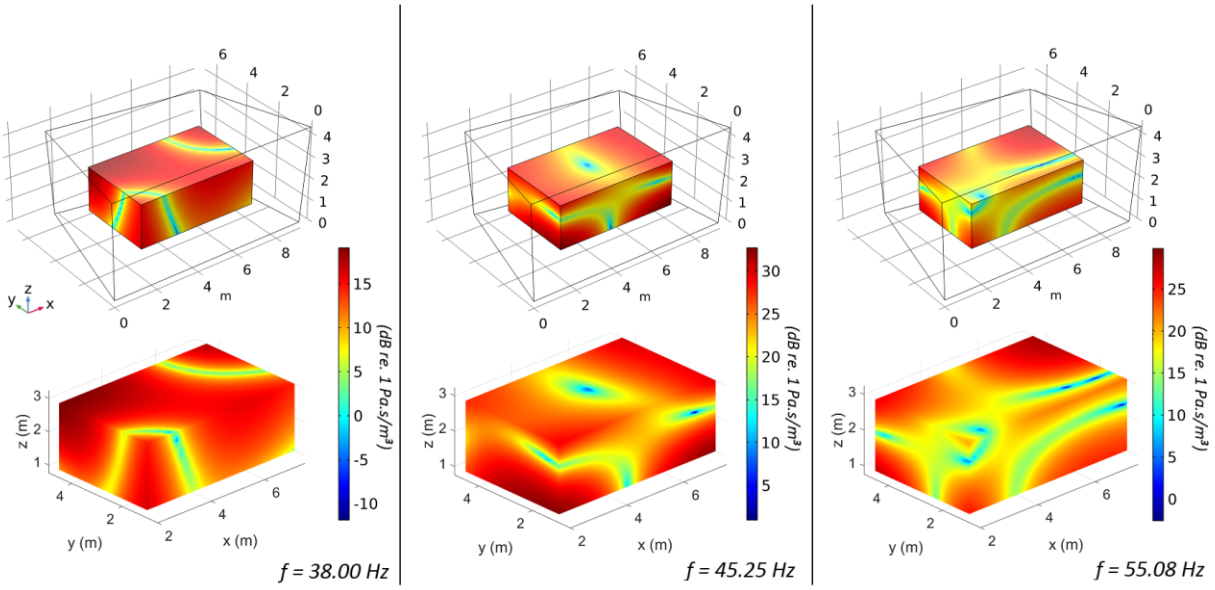

Figure 4. Sound field reconstruction (bottom) at different frequencies for a rectangular area inside the room in comparison with the referencing sound fields from numerical simulation (top).

4.2 Sound field reconstruction

A. Local interpolation

Using the outcomes of the algorithm, we can now interpolate the time response at any point inside the room. The processed RIRs in our case indicate the transmission between the source volume flow rate (in m^3/s) and the sound pressure (in Pa) acquired at the measurement points. An example can be seen in Figure 3 for a point far from the walls but also not too close to the center of the room. It can be observed that for 50 measurement points, the reconstruction is highly accurate for both the RIR and RFR in this case. It should be noted, however, that this level of accuracy is not guaranteed for every interpolated point in the room and the error could be higher depending on the position of the point and in relation with the precision of the eigenmodes searching results. This, once again, highlights the need for a spatial representation of the sound field to guarantee the validity of the framework in general.

B. Spatial sound pressure distribution

The interpolation process can now be done for multiple points to acquire a series of processed RIRs for the room. The RFRs of the room can then be produced through the Fourier transform of these time responses. These resulting RFRs will allow us to observe the spatial response of the room at any single frequency of interest.

It has been studied in [3] that for a room with non-rigid walls, the Helmholtz equation for the modeshape function is less valid for positions close to the walls. Hence, our visualization is done for a rectangular

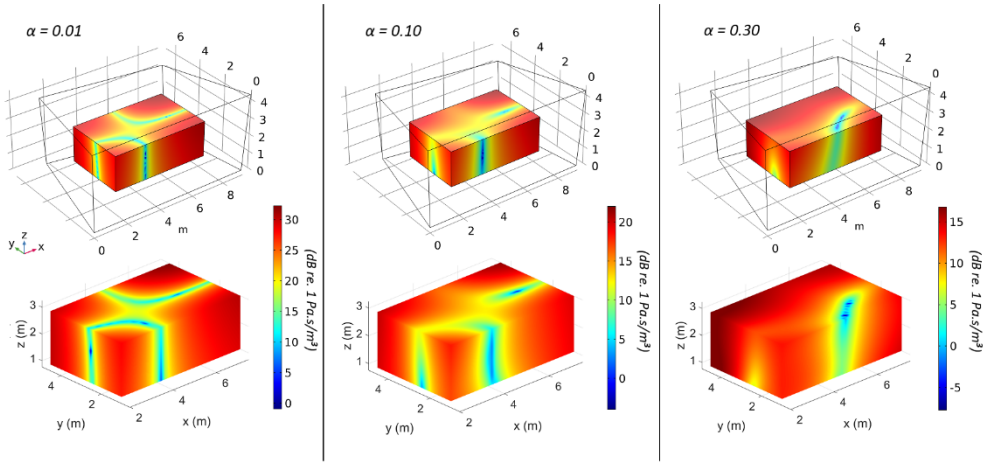

Figure 5. Sound field reconstruction (bottom) as compared to the reference (top) for different cases of wall damping at the eigenfrequency around 35.3 Hz

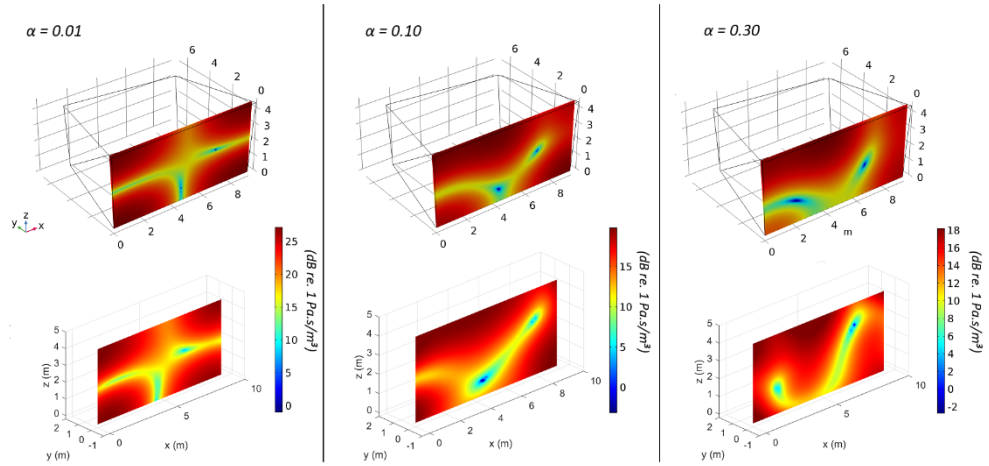

Figure 6. Same comparison as Figure 5 but for locations close to one wall of the room for the eigenfrequency around 45.2 Hz

volume inside the room but at least $1m$ away from its walls. This reconstruction can be compared with a frequency domain study using the same settings in COMSOL Frequency Domain study (refer to Figure 4). It can be observed that the reconstruction of the sound field using CS yields highly accurate results. Through this visualization of sound pressure distribution, we can verify that the algorithm not only performs well on a few individual interpolated positions but it is also capable of reconstructing the sound field accurately over wide areas. Moreover, the presence of the mode shapes are clearly depicted in all three examples. Furthermore, the high level of accuracy is maintained in every direction of the 3D depiction since the measurement points are spread randomly within the room. Some initial tryouts with a regular grid of microphones have not achieved such global precision in the reconstruction. This, once again highlights the importance of the much-recommended randomness in regular CS frameworks. From Figure 4, it should be also noticed that despite small differences occurring when comparing point by point in terms of sound pressure, the general shapes as well as the separation between high and low sound pressure areas are nevertheless precisely depicted. It could as well be seen in the comparison that the depiction of sound pressure field is accurate not just for the eigenfrequencies (at 45.25 Hz and 55.08 Hz) but also for frequencies in between two consecutive modes (38 Hz is an example).

C. Rooms with higher damping

It is worth noting that the previous results are for a lightly damped room where the walls have very low absorption coefficient ($\alpha = 0.01$). In this section, we study two more cases with $\alpha = 0.1$ and $\alpha = 0.3$ to challenge the robustness of the algorithm. Figure 5 shows the comparison of the reconstruction at the same eigenfrequency around 35.3 Hz between

different values of wall absorption. Surprisingly, the reconstructed sound pressure distribution for these cases can still maintain a fairly precise depiction of the simulated sound field. It can clearly be seen that the CS reconstruction also correctly depicts the smoothening effect on the spatial distribution as the room becomes more damped. This validation result is especially meaningful in terms of studies regarding room modes equalization. It shows that this particular CS framework can be used to assess the room sound field before and after a certain damping or equalization method has been used, especially the ones using active control.

The sound field reconstruction quality near the walls, however, degrades considerably, especially in the case of $\alpha = 0.3$ as can be seen in Figure 6. This is partially due to the fact that in our algorithm, we assume that the orthogonality of mode shape functions is valid regardless of the general damping situation. This result is hence understandable as the assumption is more valid when the analyzed sound field is far from the walls rather than when it is close to the boundaries. Future research can focus on how to improve the model taking into account the non-orthogonality of mode shape functions in highly damped case as the near-boundary sound field also has certain importance to room acoustics practitioners.

5. Conclusions

In this communication, we have been able to verify the validity of the compressive sensing approach in reconstructing sound fields of a general non-rectangular room. Under this framework, we are able to reconstruct the spatial sound pressure field inside the room at certain frequencies and analyze its precision in comparison to that from simulations. The comparisons show promising results not only for a lightly damped room but also for rooms with higher damping assuming the analyzed sound field is not too close to the walls. This could prove to be useful for further application of the algorithm by a wider audience of users. One of many potential applications is to use this method to evaluate the change in the sound fields due to different room mode treatment methods. Future studies will focus on this particular application as well as study further on other robust aspects of the algorithm such as how to improve its accuracy near the wall for highly damped rooms as well as how to reduce the number of measurement points even more.

References

[1] T. Ajdler, L. Sbaiz and M. Vetterli, "The Plenacoustic Function and Its Sampling," in IEEE Transactions on Signal Processing, vol. 54, no. 10, pp. 3790-3804, Oct. 2006.

[2] R. Mignot, G. Chardon and L. Daudet, "Low Frequency Interpolation of Room Impulse Responses Using Compressed Sensing", in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 22, no. 1, pp. 205-216, Jan. 2014.

[3] H. Kuttruff, 2000. Room acoustics Fourth. Spon Press.

[4] A. Moiola, R. Hiptmair, and I. Perugia, "Plane wave approximation of homogeneous Helmholtz solutions", Zeitschrift fur Angewandte Mathematik und Physik (ZAMP), vol. 62, pp. 809-837, 2011.

[5] Joel A. Tropp, Anna C. Gilbert, Martin J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit, Signal Processing", Volume 86, Issue 3, 2006, Pages 572-588, ISSN 0165-1684, https://doi.org/10.1016/j.sigpro.2005.05.030.

[6] G. Chardon and L. Daudet, Optimal subsampling of multichannel damped sinusoids, in Proc. 6th IEEE Sensor Array and Multichannel Signal Process. Workshop (SAM'10), pp. 25-28, 2010.

[7] T. Pham Vu, E. Rivet, H. Lissek, "Low Frequency Sound Field Reconstruction in Non-rectangular Room: A Numerical Validation", 11th European Congress and Exposition on Noise Control Engineering – Euronoise 2018

[8] R. Roy and T. Kailath, "Estimation of signal parameters via rotational invariance techniques," IEEE Trans. Acoust., Speech, Signal Process., vol. 37, no. 7, pp. 984-995, 1989.