ICA 2019 AACHEN

PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

An application of multi-scale directional dictionaries to RIR interpolation

Elias ZEA[1]

[1] The Marcus Wallenberg Laboratory for Sound and Vibration Research, KTH Royal Institute of Technology, Sweden

ABSTRACT

The spatio-temporal sparsity of room impulse responses (RIRs) degrades in the late part due to the stronger wave interference compared with the early part. Meanwhile, such an interference decays in amplitude as time progresses due to absorption in the room, resulting in a decreasing dynamic range of the measurement. Together these aspects pose challenging conditions for compressive sensing applications, such as the interpolation of RIRs measured at sparse microphone positions. In the search of sparse transformation spaces, this paper examines the application of a multi-scale directional dictionary (known as shearlets) to interpolate RIR measurements. These redundant dictionaries consist of multiple curved elementary functions, which offer a decomposition of the acoustic wavefronts into various wavelengths, propagation directions, and times of arrival. Results reported in this paper demonstrate the potential these dictionaries have to interpolate RIRs in both convex and nonconvex rooms, motivating further examination under experimental conditions and in broader frequency ranges.

Keywords: Room impulse responses, Interpolation, Shearlet dictionaries

1. INTRODUCTION

Room impulse responses (RIRs) can be regarded as spatio-temporal imprints carrying all the relevant information about the acoustic response a sound source has in a given environment, e.g. in a factory hall or a vehicle cabin. Measuring RIRs is therefore a task of considerable importance these noisy days, as it provides a means to analyze, compensate, and control sound fields in acoustical spaces – let alone the many applications it has in teleconferencing, virtual/augmented reality, and auralization [1]. In essence, RIRs are measured by recording the pressure response of the acoustical space to an impulsive source, at a number of positions with microphones.

The execution of such measurements is classically limited by the number of positions (by Nyquist-Shannon sampling): beyond 1.6 million microphone positions per cubic meter for frequencies within the audible human range. For this reason, one can today find novel sparse sampling approaches, based on compressive sensing (CS) [2,3], to reconstruct RIRs that have not been measured using a reduced set of microphone positions. The success of CS relies on the sparsity –a measure of the no. significant coefficients– of the RIRs in some appropriate vector space, known as dictionary; and on an incoherent sensing operation, generally linked with the randomness of the microphone positions. In this way, one can think of CS as the action of sampling and compressing the RIRs at once.

Among the pioneering papers in this topic has been written in 2013 by Mignot *et al.* [4], in which the authors exploit the temporal sparsity of the early part of the RIRs by using a dictionary of monopole sources. The room in which the measurements were carried out is empty, and it is assumed that no diffraction phenomena occurs. Few years later, Antonello *et al.* published a paper in which they use time-varying equivalent source dictionaries and various kinds of sparsity domains: spatial, spatio-spectral, and spatio-temporal; and presented an experimental validation with RIRs measured in an empty, rectangular room [5]. Plane-wave dictionaries have also been subject of study by Mignot *et al.*, as well as, more recently, by Verburg and Fernandez-Grande, who have investigated the

[1] zea@kth.se

interpolation [6,7] and extrapolation [7] of the responses in empty, rectangular rooms. A common result from using plane-wave dictionaries is that the associated sparsity goes hand-in-hand with the modal density of the room [7]: sparse plane-wave representations are most accurate in convex rooms and at frequencies below Schroeder's limit [8]. Overall, the sparsity of the representation systems studied so far seems to be strongly linked with the geometry of the room, its damping, and its contents.

In this way, the motivation driving this paper is the search for sparse representation systems that can account for a broader variety of room properties; conditions closer to the acoustic environments one can encounter. This paper examines the application of a multi-scale directional representation system, known in the image processing community as *shearlets* [9], thereby exploiting the *sparsity of the classes of functions* in the RIRs: curved singularities (wavefronts) of various shapes, propagation directions, and times of arrival. By approximating these curved functions directly in space-time, no prior knowledge of the properties of the various objects and boundaries in the room is needed. In particular, this paper studies the interpolation of synthetic RIRs in two rooms, using 2D planar array data obtained with finite elements and a 3D shearlet dictionary [10].

2. BACKGROUND

2.1 Shearlets and RIR data

Shearlets can be understood as elongated wavelets, with anisotropic directional properties. Harmonic analysts have designed them to optimally represent edges (e.g. discontinuities of bounded curvature) in multidimensional data [11]. Here optimality means the fewest coefficients to represent the edges, which has a beneficial impact in the tasks of compression, transmission, denoising, as well as restauration of image/video signals. Shearlets are in essence curved elementary functions of various scales, orientations, and translations, and together constitute a redundant representation system; with a redundancy factor in the order of tens to hundreds [11]. See examples of shearlets in Figure 1.

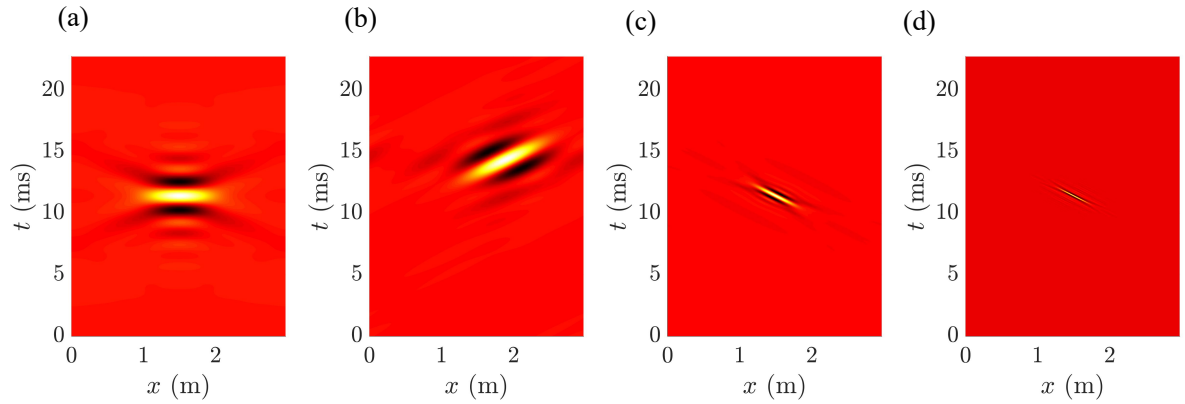

Figure 1 – Examples of 2D shearlet elements in space-time, drawn from a 3-scale dictionary [12]. *Shearlet scales*: elements (a) and (b) belong to scale 1 (broader scale), whereas (c) and (d) belong to scales 2 and 3 (finer scales). *Shearlet orientations*: there are various orientations per scale, obtained from shears of the same element. *Shearlet translations*: elements (a), (c) and (d) have no translation (centered in the image), whereas (b) has some offset in time and in space.

As Figure 2 illustrates, the principal idea in this paper is to apply shearlet dictionaries as sparse representation systems for spatio-temporal RIR measurements, data which consists of curved wavefronts of various shapes, propagation directions, and times of arrival. For instance, consider the spatio-temporal RIRs are arranged into a column vector $\mathbf{p} \in \mathbb{R}^T$, with T as the total number of time samples and microphone positions, and define the shearlet dictionary matrix $\mathbf{\Theta} \in \mathbb{C}^{T \times F}$, with $F > T$ as the total number of shearlet coefficients. Then, the RIRs can be synthesized from its corresponding shearlet expansion coefficients, $\mathbf{s} \in \mathbb{C}^F$, via the expression

$$\mathbf{p} = \mathbf{\Theta s}. \tag{1}$$

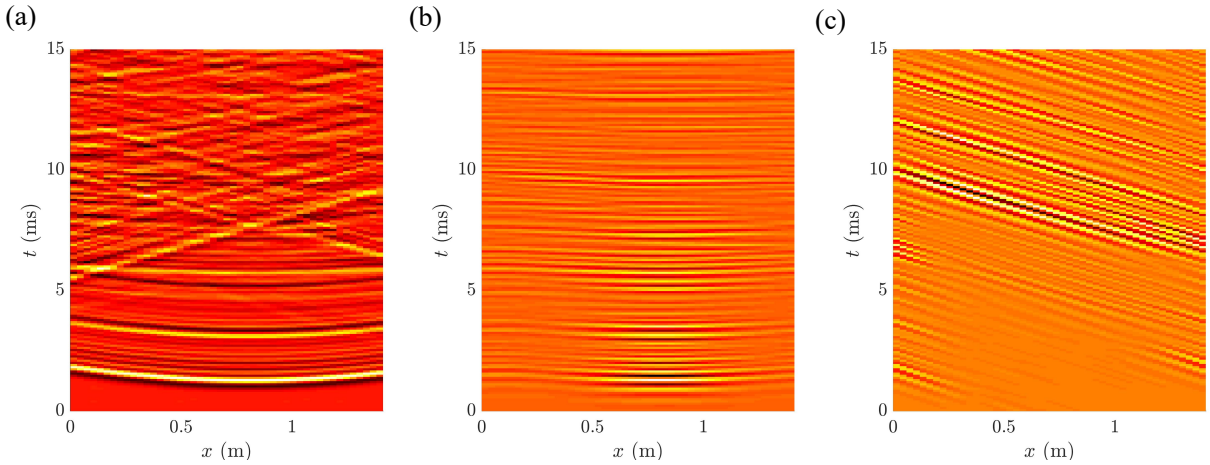

Figure 2 – Shearlet decompositions of RIR data. (a) RIRs measured in a small office room with a 1D array, with microphone spacing 3 cm and sampling frequency 11250 Hz. (b) Decomposition of the RIRs into a "horizontal" shearlet [like the one in Fig. 1(a)]. (c) Decomposition of the RIRs into a "diagonal" shearlet [like the one in Fig. 1(c)]. Images are normalized for illustration purposes.

2.2 RIR interpolation

Let us begin by defining the RIRs measured at sparse positions, $\tilde{\mathbf{p}} \in \mathbb{R}^U$, as the action of under-sampling the target RIRs, $\mathbf{p}$, with an appropriate masking operator $\mathbf{\Pi} \in \mathbb{R}^{U \times T}$, that is $\tilde{\mathbf{p}} = \mathbf{\Pi p}$ (see Figure 3). Using a shearlet dictionary as representation system via Eq. (1) leads us to the under-determined system of equations

$$\tilde{\mathbf{p}} = \mathbf{\Phi s} + \mathbf{n}, \tag{2}$$

where the compressive sensing operator $\mathbf{\Phi} = \mathbf{\Pi\Theta} \in \mathbb{C}^{U \times F}$ models the joint action of shearlet synthesis and spatial under-sampling, and the vector $\mathbf{n} \in \mathbb{R}^U$ accounts for measurement noise and approximation errors. The system under-determination is caused by the less observations than shearlet coefficients: $U \ll F$.

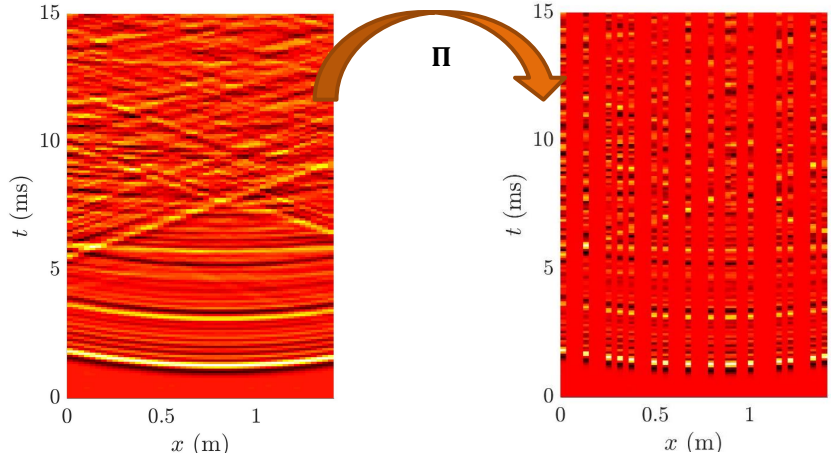

Figure 3 – Applying the mask $\mathbf{\Pi}$ to an RIR dataset (left), to obtain the under-sampled RIR dataset (right). The vertical stripes in the right correspond to the missing microphone positions. Images are normalized for illustration purposes.

Rooted on the assumption that the shearlet dictionary serves a reasonably sparse representation system of the RIRs, the search for the original (target) RIRs $\mathbf{p}$ can be mathematically formulated as a convex optimization problem [13]

$$\min_{\mathbf{s}} \|\tilde{\mathbf{p}} - \mathbf{\Phi s}\|_2^2 + \mu \|\mathbf{s}\|_1, \tag{3}$$

where the ℓ_q-norm is defined as $\|\cdot\|_q = [\sum|\cdot|^q]^{1/q}$, and $\mu > 0$ is a regularization parameter that balances the sparsity of the solution and the model misfit. In order to solve (3) efficiently with the use of fast Fourier transforms, one can run an iterative soft-thresholding algorithm [14]

$$\mathbf{s}^{(\gamma)} = \Im_\mu \left\{\mathbf{s}^{(\gamma-1)} + \mathbf{\Phi}^H\left(\tilde{\mathbf{p}} - \mathbf{\Phi}\mathbf{s}^{(\gamma-1)}\right)\right\}, \tag{4}$$

with some initial shearlet coefficient vector $\mathbf{s}^{(0)}$, e.g. a zero vector, H denotes Hermitian transpose, and the action of a soft-thresholding operator over the i-th entry of the vector $\mathbf{s}$ is defined as

$$\Im_\mu \{s_i\} = \operatorname{sgn}(s_i) \cdot \max(0, |s_i| - \mu/2). \tag{5}$$

At the γ-th iteration step, the RIRs are interpolated by computing the synthesis equation (1) with $\mathbf{s} = \mathbf{s}^{(\gamma)}$. The value of μ can be chosen, for instance, from the corner of the Pareto frontier curve [15]. In this work we shall consider the computer implementation of the 3D discrete shearlet transform *ShearLab* [10], which allows us to process time-domain responses measured with 2D planar arrays. 3D shearlets can be thought of as 2D shearlets with additional orientations and translations due to the additional spatial dimension.

2.3 A note on sensing coherence

Shearlets that are aligned with the time axis constitute artificial information, as their coefficient amplitudes are only caused by (coherent with) the missing microphone responses. On the contrary, shearlets that are aligned towards the spatial axes constitute natural information, as they are aligned with the acoustic wavefronts. In this way, it is crucial to remove (from the dictionary representation) the shearlets that are aligned with the time axis, such that the sensing coherence of $\mathbf{\Phi}$ is decreased and, as a consequence, the interpolation accuracy is increased. This issue, illustrated in Figure 4 below, was first described by Herrmann and Hennenfent, in their work on interpolating seismic traces with curvelets [16].

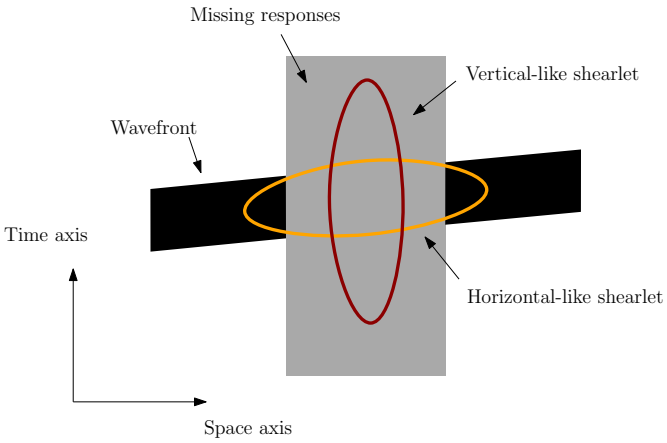

Figure 4 – Artificial (and natural) information carried by “vertical” (and “horizontal”) shearlets

3. SIMULATION SETUP

The sound field in two rooms is predicted with finite elements in COMSOL. The first room is a 2D shoebox [see Fig. 5(a)], of dimensions 8.5m x 10m, with acoustically hard walls. The second room has nonconvex geometry [see Fig. 5(b)], also with hard walls. The reason hard walls are considered is that they are associated with a less sparse (more challenging) sound field to reconstruct [7]. A point source is located at $(x, y) = (0,0)$, which emits a Gaussian pulse at $t = 100$ μs, with 500 Hz bandwidth, and 10^{-2} m^2/s amplitude.

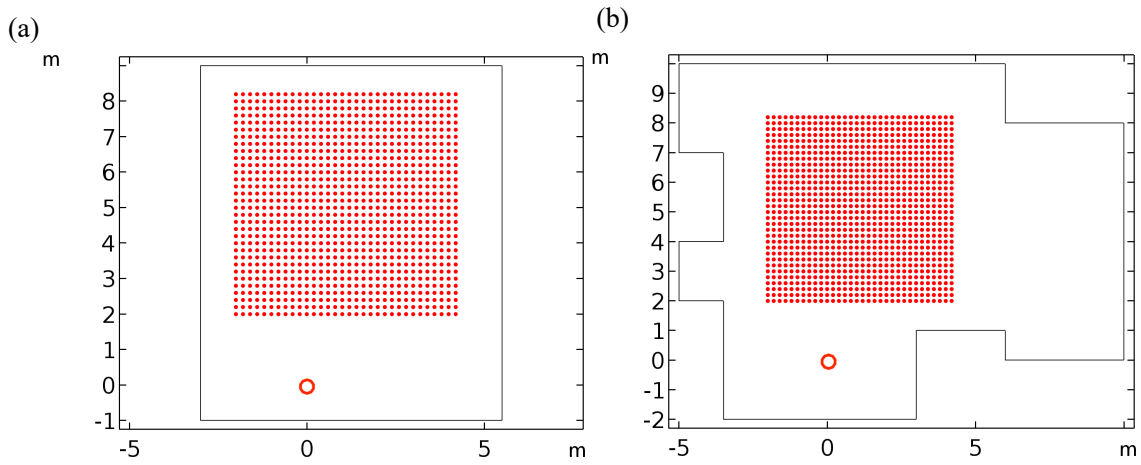

Figure 5 – Illustration of the simulation setup in (a) room no. 1, shoebox, and (b) room no. 2, nonconvex. The source position is shown with the bigger circle, whereas the 32 x 32 array positions are shown with the smaller circles.

The reference sound field is measured at 32 x 32 positions (see Fig. 5), separated by 20 cm. The responses are 274 ms, with sampling frequency of 1000 Hz. The under-sampled sound field is obtained from application of the masks –shown in Figure 6 below– at all time instants, resulting in spatial information sampled at 1/3 and 1/5 of the Nyquist-Shannon rate. Gaussian noise is also added to the sound field, so as to have a signal-to-noise ratio of 30 dB.

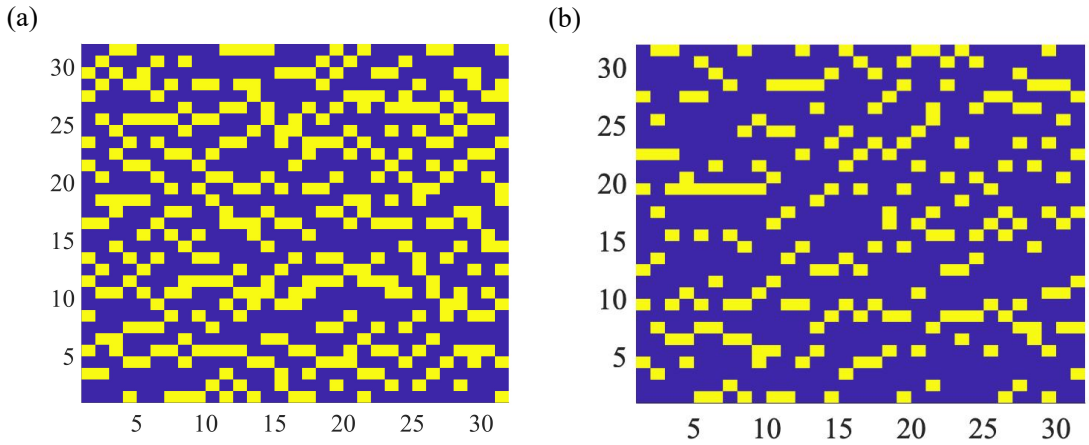

Figure 6 – 2D spatial masks [measured (yellow) and masked (blue)] which are applied to the measured sound field at all time instants, in order to get the under-sampled RIRs. The masks are designed via jittered under-sampling [17], corresponding to (a) 67% and (b) 80% missing responses.

The interpolation accuracy is quantified with the relative reconstruction error in dB

$$\varepsilon = 10\log_{10}\frac{\left\|\mathbf{p}_{rec}-\mathbf{p}_{ref}\right\|}{\left\|\mathbf{p}_{ref}\right\|}, \tag{6}$$

where the subscripts “rec” and “ref” denote recovered and reference RIRs. In addition, the modal assurance criterion (MAC) [18]

$$\mathrm{MAC}(f_j) = \frac{\left|\langle \widetilde{\mathbf{p}_{rec}}(:,f_j), \widetilde{\mathbf{p}_{ref}}(:,f_j)\rangle\right|}{\left|\widetilde{\mathbf{p}_{rec}}(:,f_j)\right|\left|\widetilde{\mathbf{p}_{ref}}(:,f_j)\right|} \tag{7}$$

is also computed, which indicates the spatial similarity between the Fourier spectra (denoted with $\widetilde{\ }$) of the recovered and reference RIRs at the j-th frequency.

4. INTERPOLATION RESULTS

Figure 7 shows the interpolation results in the two rooms (arranged in rows), running a total of 200 thresholding iterations, and using a 3D shearlet dictionary of 3 scales [10]. The relative error is in this case -13.2 dB in the shoebox room, and -12.7 dB in the nonconvex room. Overall there is good agreement between the interpolated and reference RIRs; however, some artifacts due to the mask can be observed in the interpolated responses (middle column of Fig. 7). These artifacts can, on the one hand, be attributed to too large a gap between missing positions compared with the size of the shearlets [19]; but, on the other hand, to the lack of spatial periodicity in the RIRs, which can cause shearlets to wraparound at the edges of the array – similarly to the leakage problem in Fourier acoustics [20].

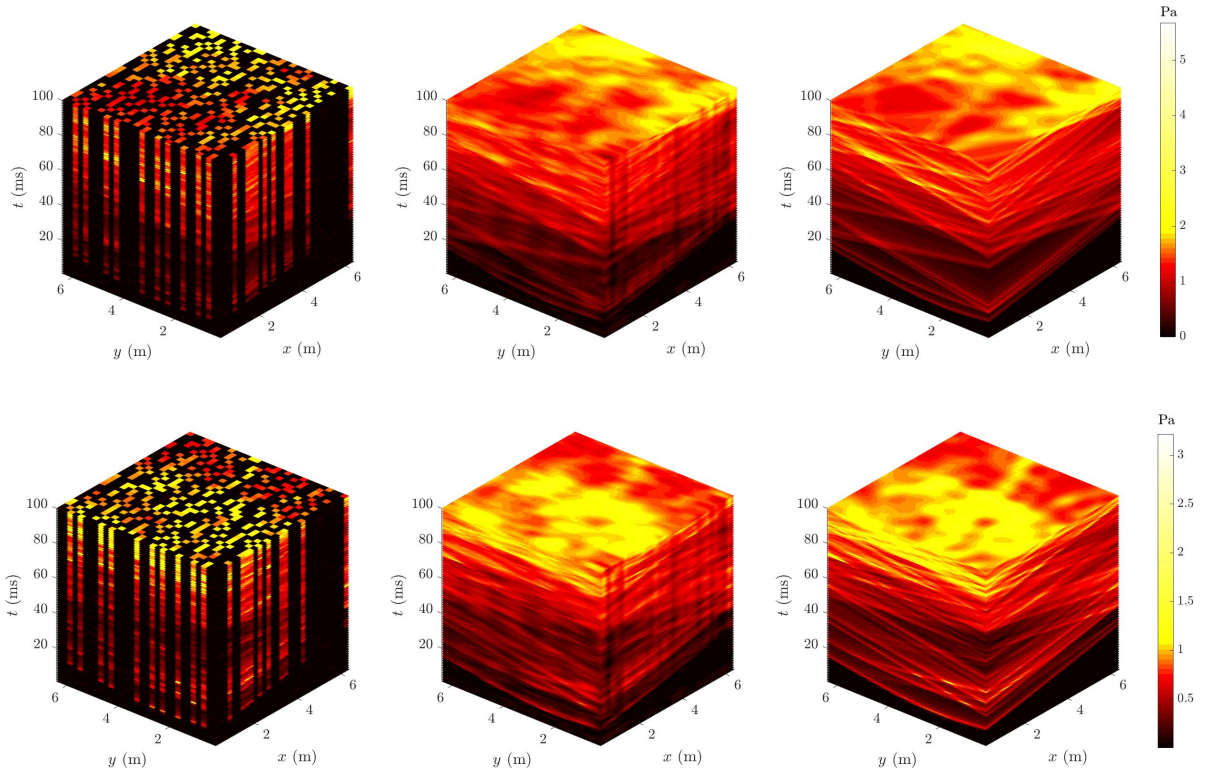

Figure 7 – RIR interpolation results, with spatial under-sampling factor of 3, and during the first 100 ms of responses. Top row: results in room no. 1, shoebox [see Fig. 5(a)]. Bottom row: results in room no. 2, nonconvex [see Fig. 5(b)]. Left column: under-sampled RIRs. Middle column: interpolated RIRs. Right column: reference RIRs.

Figure 8 shows the same kind of results as Figure 7, but with the spatial under-sampling factor of 5. As more information is missing, conditions become more challenging to recover the original responses: lower probability of information recovery [2]. Nevertheless, some of the wavefronts are reasonably interpolated in the narrower gaps between missing positions. In this case, the relative errors are -5.1 dB and -5.2 dB in room no. 1 and 2, respectively.

To complement these results, the MAC is plotted in Figure 9, for RIRs interpolated in the two rooms, and with the two spatial under-sampling factors (3 and 5). It is clear to see that the MAC values are closer to 1 in a broader bandwidth as less information is missing. For instance, the MAC values obtained from 67% under-sampling seem to remain above 0.6 up to 400 Hz, which means the interpolated RIRs are not far from the reference. Also, the results do not seem to be largely influenced by the convexity of the acoustical space, which can be considered promising towards the search of sparse representation systems for RIRs measured in complex acoustical spaces.

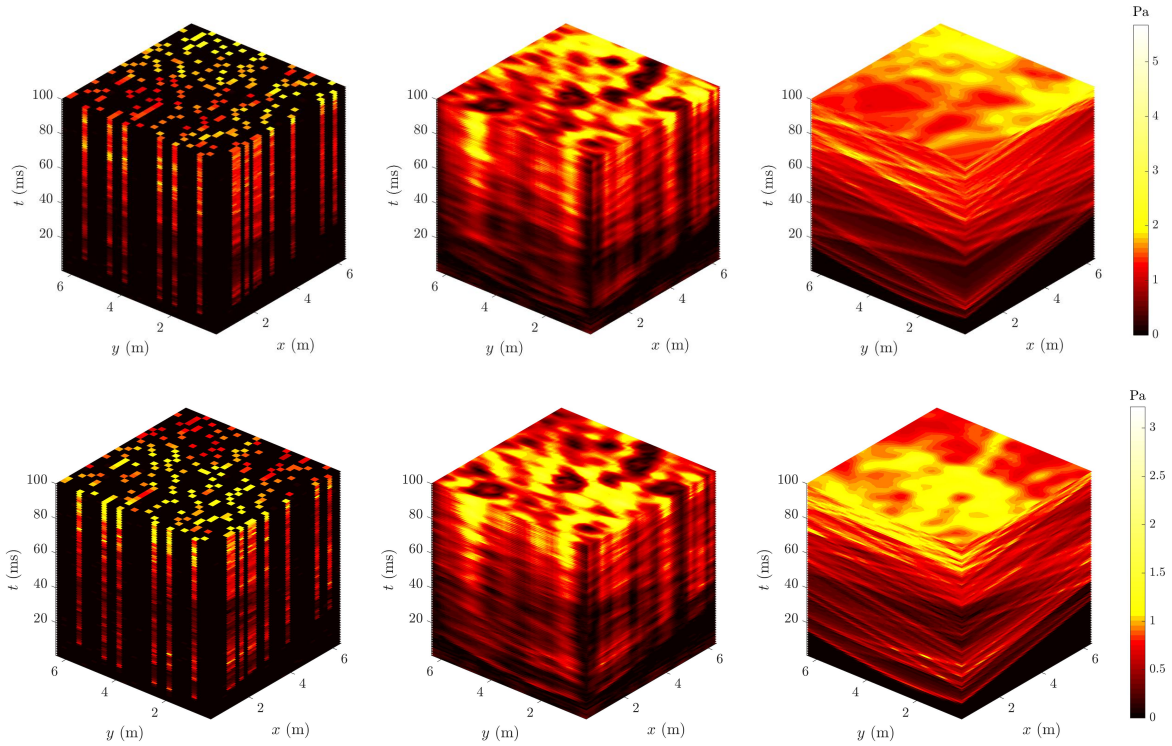

Figure 8 – RIR interpolation results, with spatial under-sampling factor of 5, and during the first 100 ms of responses. Top row: results in room no. 1, shoebox [see Fig. 5(a)]. Bottom row: results in room no. 2, nonconvex [see Fig. 5(b)]. Left column: under-sampled RIRs. Middle column: interpolated RIRs. Right column: reference RIRs.

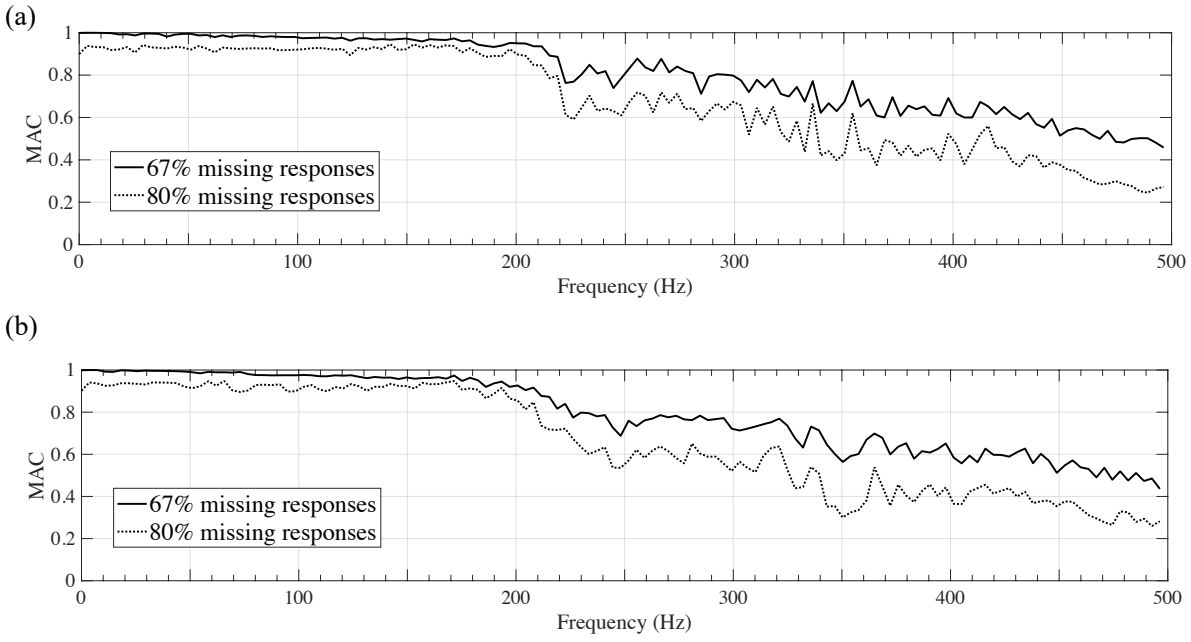

Figure 9 – Modal assurance criterion (MAC) values against frequency, for 67% (solid) and 80% (dotted) missing responses. (a) Room no. 1 (b) Room no. 2.

5. CONCLUSIONS

This paper presents the application of shearlet dictionaries to the problem of reconstructing room impulse response (RIR) measurements with sparse microphone arrays. The idea can be understood as approximating the spatiotemporal wavefronts into a sparse set of curves of various sizes, orientations,

and translations. Then, based on the assumption of RIR sparsity in the shearlet domain, the interpolated solution can be approximated by running an iterative thresholding algorithm. For validation purposes, the sound field is numerically computed in two rooms using finite elements, and thereafter under-sampled in space by a factor of 3 and 5; resulting in 67% and 80% missing responses, respectively. The results demonstrate a reasonably accurate interpolation of the 67% missing responses in a frequency range up to 400 Hz, and in the presence of background noise 30 dB below the signal level. In the case of 80% missing responses, too much information is lost and the interpolation RIRs disagree with the reference RIRs. Overall, the promising results obtained here motivate further investigation under experimental conditions and in a broader frequency range.

ACKNOWLEDGEMENTS

The present work is supported financially by the Swedish Research Council, under grant agreement No. 2016-04366. Thanks to Prof. I. Lopez Arteaga for the useful comments and suggestions.

REFERENCES

1. Vorländer. M. Auralization: Fundamentals of Acoustics, Modelling Simulation, Algorithms, and Acoustic Virtual Reality. Springer-Verlag, Berlin (Germany), 2008.
2. Candès, E.J.; Wakin, M.B. An introduction to compressive sampling, IEEE Signal Process Mag 25(2), 2008, 21–30.
3. Gerstoft, P.; Mecklenbräuker, C.F.; Seong, W.; Bianco, M. Introduction to compressive sensing in acoustics. J Acoust Soc Am 143(6), 2018, 31-48.
4. Mignot, R.; Daudet, L.; Ollivier, F. Room reverberation reconstruction: Interpolation of the early part using compressed sensing. IEEE Trans Audio Speech Lang Process 21(11), 2013, 2301–2312.
5. Antonello, N.; De Sena, E.; Moonen, M.; Naylor, P.A.; van Waterschoot, T. Room impulse response interpolation using a sparse spatio-temporal representation of the sound field. IEEE/ACM Trans Audio Speech Lang Process 25(10), 2017, 1929–1941.
6. Mignot, R.; Chardon, G.; Daudet, L. Low frequency interpolation of room impulse responses using compressed sensing. IEEE/ACM Trans Audio Speech Lang Process 22(1), 2014, 205–216.
7. Verburg, S.; Fernandez-Grande, E. Reconstruction of the sound field in a room using compressive sensing. J Acoust Soc Am 143, 2018, 3770–3779.
8. Schroeder, M.R.; Kuttruff, K.H. On frequency response curves in rooms. Comparison of experimental, theoretical, and Monte Carlo results for the average frequency spacing between maxima. J Acoust Soc Am 34(1), 1962, 76–80.
9. Labate, D.; Lim, W.-Q.; Kutyniok, G.; Weiss, G. Sparse multidimensional representation using shearlets SPIE 5914, 2005, 254–262.
10. Kutyniok, G.; Lim, W.-Q.; Reisenhofer, R. ShearLab 3D: Faithful digital shearlet transforms based on compactly supported shearlets. ACM Trans Math Software 42, 2014, 1–39.
11. Kutyniok, G.; Labate, D. Shearlets: Multiscale Analysis for Multivariate Data. Springer, London (UK), 2012.
12. Häuser, S.; Steidl, G. Fast finite shearlet transform: a tutorial. ArXiv 1202.1773, 2012, 1–41.
13. Elad, M. Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing. Springer, London (UK), 2010.
14. Daubechies, I.; Defrise, M.; De Mol, C. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Comm Pure Appl Math 57, 2004, 1413–1457.
15. Van den Berg, E.; Friedlander, M.P. Probing the Pareto frontier curve for basis pursuit solutions. SIAM J Sci Comput 31(2), 2008, 890–912.
16. Herrmann, F.J.; Hennenfent, G. Non-parametric seismic data recovery with curvelet frames. Geophys J Int 173, 2008, 233–248.
17. Hennenfent, G.; Herrmann, F.J. Simply denoise: Wavefield reconstructon via jittered undersampling. Geophys 73(3), 2008, V19–V28.
18. Pastor, M.; Binda, M.; Harcarik, T. Modal assurance criterion. Proc Eng 48, 2012, 543–548.
19. Genzel, M.; Kutyniok, G. Asymptotic analysis of inpainting via universal shearlet systems. SIAM J Imag Sci 7(4), 2014, 2301–2339.
20. Williams, E.G. Fourier Acoustics: Sound Radiation and Nearfield Acoustic Holography. Academic Press, San Diego (USA), 1999.