JOINT ESTIMATION OF THE ROOM GEOMETRY AND MODES WITH COMPRESSED SENSING

Helena Peić Tukuljac, Thach Pham Vu, Hervé Lissek, Pierre Vandergheynst

School of Computer and Communication Sciences
Ecole Polytechnique Fedérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland
{helena.peictukuljac, thach.phamvu, herve.lissek, pierre.vandergheynst}@epfl.ch

ABSTRACT

Acoustical behavior of a room for a given position of microphone and sound source is usually described using the room impulse response. If we rely on the standard uniform sampling, the estimation of room impulse response for arbitrary positions in the room requires a large number of measurements. In order to lower the required sampling rate, some solutions have emerged that exploit the sparse representation of the room wavefield in the terms of plane waves in the low-frequency domain. The plane wave representation has a simple form in rectangular rooms. In our solution, we observe the basic axial modes of the wave vector grid for extraction of the room geometry and then we propagate the knowledge to higher order modes out of the low-pass version of the measurements. Estimation of the approximate structure of the k-space should lead to the reduction in the terms of number of required measurements and in the increase of the speed of the reconstruction without great losses of quality.

Index Terms— compressed sensing, k-space, plane waves, room modes, room transfer function

1. INTRODUCTION

In 2006 Ajdler et al. [1] have defined the Plenacoustic function (PAF) as the function that contains the room impulse responses (RIRs) for all the possible pairs of microphone and source positions in a room with the given acoustical properties. Without having any prior knowledge involved, it is extremely hard to estimate the PAF. As shown by Moiola et al. [2] the acoustical behavior of the room can be described by a discrete sum of plane waves that can exist inside a given room which are tightly related to the resonant frequencies. This plane wave approximation holds for any star-convex room and is independent of boundary conditions, domain of propagation, type of the source or proximity to the source or the walls [3].

Sparse plane wave approximation in the low frequency domain introduces an assumption required for sparse analysis of room's complex wavefield which further opens the door to compressed sensing [4, 5]. Mignot et al. [3] have started the trend of the sparse modal analysis. They have designed a greedy approach which uses space decomposition based on iterative alternating projections for the estimation of the wave number and wave vectors that fully determine the acoustical behaviour of the given room. Due to the high dimensionality of data acquired by microphones, greedy methods such as Simultaneous Orthogonal Matching Pursuit (SOMP) [6] (simultaneous, since we are fitting measurements from multiple microphones at once) have shown better performance than the relaxation of the minimization of ℓ_0 norm [7].

Our solution focuses on the structured sparsity of the plane wave representation for the reconstruction of parameters of the Room Transfer Function (RTF). In literature, sparse plane wave representation has been used not only for the representation of the wavefield in a room in low frequency domain, but also for efficient storage of highly correlated recordings of dense microphone arrays [8]. Besides sparse plane wave representation an interesting sparse approach to the estimation of RTF is a recent approach with orthonormal basis functions based on infinite impulse response filters (IIR) [9]. Though not exploring plane wave sparsity, the solution relying on the weighted spatio-temporal representation [10] also gives promising room impulse response interpolations.

On the other hand, the solutions for estimating the shape of the room usually rely on knowing the location of early reflections [11], [12], but finding the true reflections within an echogram is not a trivial problem and is still an open research question.
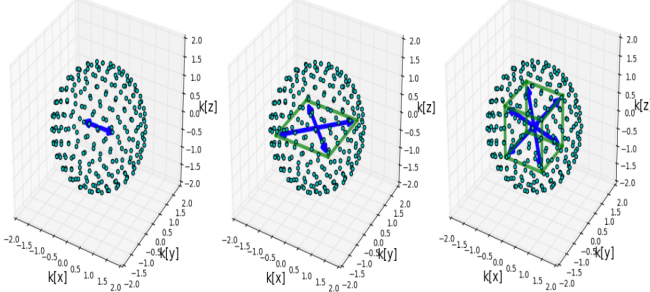
2. PROBLEM SETUP

When sampling sound we need to take into account two types of possible aliasings: temporal and spatial. Depending on the highest frequency that we want to capture f_c, we define our temporal sampling step Δt in such a way that the sampling frequency satisfies $f_s = \frac{1}{\Delta t} > 2f_c$ [13]. Once the
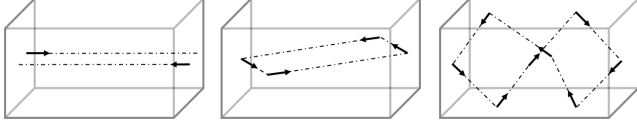
(a) Plane waves inside a rectangular room.

(b) Wave vectors in the search space.

Fig. 1. Plane wave types and corresponding structured sparsity of wave vectors. In theory these vectors form a parallelepiped inscribed into a sphere with radius $\frac{\omega_n}{c}$, resulting in structured sparsity. From left to right: x-axial mode, xy-tangential mode and oblique mode.
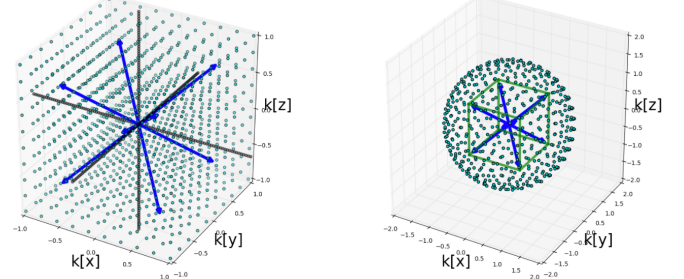

Fig. 2. The left hand side shows the periodicity of the wave vector grid with respect to $\mathbf{k} = [\pm k_x, \pm k_y, \pm k_z]$ with period over the axes equal to: $\frac{\pi}{L_x}$, $\frac{\pi}{L_y}$ and $\frac{\pi}{L_z}$. Here we see an example of an oblique wave vector. The right hand side shows the search space on our uniformly sampled sphere.

temporal sampling step is fixed, we determine the appropriate sampling step in space either by the limits imposed by the Courant–Friedrichs–Lewy condition [14] for finite difference time domain (FDTD) schemes, or by a contemporary view of the problem observed through the sampling of the PAF [1].

The support of the spectrum of the PAF $\hat{p}(\omega, \varphi_x, \varphi_y, \varphi_z)$, where $\omega = \frac{2\pi}{\Delta t}$ is the temporal angular sampling frequency $[rad.s^{-1}]$ and $\varphi_i = \frac{2\pi}{\Delta i}$ is the spatial angular frequency $[rad.m^{-1}]$ over each of the ith observed axis, lays inside a hypercone: $\varphi_x^2 + \varphi_y^2 + \varphi_z^2 \leq \frac{\omega^2}{c^2}$, where c is the celerity of sound. This gives the following condition for the sampling step over each of the axes: $\Delta i < \frac{\pi c}{\omega_c}$, $\forall i \in \{x, y, z\}$. The following question emerges: can we acquire the targeted information at lower sampling rates, both in time and in space, by exploiting the underlying structure of the data without introducing significant losses?

2.1. Plane wave representation of wavefield

Acoustic propagation is governed by the wave equation:

$$\Delta p(t, \mathbf{X}) - \frac{1}{c^2}\frac{\partial^2}{\partial^2 t}p(t, \mathbf{X}) = 0. \tag{1}$$

Solution of the wave equation can be approximated in the low frequency domain as a discrete sum of damped complex harmonics [2]:

$$p(t, \mathbf{X}) = \sum_{q \in I} A_q \phi_q(\mathbf{X}) g_q(t), \tag{2}$$

where $I \subset \mathbb{Z}$, ϕ_q represents the spatial dependency of mode shape whose shape is illustrated in [15] and g_q is corresponding time evolution of the mode. Temporal functions are orthogonal. This expression emphasizes the separability of the analysis and the estimation of the temporal and spatial parameters, which can greatly reduce the computational complexity of the parameter analysis [3].

Temporal functions take the form of $g_q(t) = e^{jk_q ct}$, where $k_q = \frac{\omega_q - j\xi_q}{c}$ is the *wave number* of the qth room mode. ω_q is the resonant frequency and $\xi_q < 0$ is the corresponding damping factor. On the other hand, in the spatial functions, the room modes can be decomposed as a sum of plane waves: $\phi_q(\mathbf{X}) \approx \sum_{r=1}^{R} a_{q,r} e^{j\mathbf{k}_{q,r} \cdot \mathbf{X}}$, where $\mathbf{k}_{q,r}$ is the rth *wave vector* of the qth mode and $R = 8$ for a rectangular room case. In Figure 1. we see an example of all types of plane waves in a rectangular room: axial, tangential and oblique determined by the wave vectors. In the case of a room with *low damping*, the length of the wavevectors can be approximated by the real part of the corresponding wavenumber: $\|\mathbf{k}_{q,r}\| = |k_q|$, since in that case $k_q \approx \frac{\omega_q}{c}$. This gives us an intuition for the spherical vector search which will be explained more in detail later. For a rectangular room the wave vectors are on the vertices of a parallelepiped inscribed into the sphere. Through modal decomposition (3) and plane waves approximation, the final form of the RIR is composed of the modal wave numbers k_q's, the corresponding wave vectors $\mathbf{k}_{q,r}$'s and their expansion coefficients α_q's:

$$p(t, \mathbf{X}) = \sum_{q,r} \alpha_q e^{j(k_q ct + \mathbf{k}_{q,r} \cdot \mathbf{X})}. \tag{3}$$

As in the theory of modal decomposition [16, 2], we will focus on the coupling of the pressure field with the standing waves at room's resonant frequencies.

2.2. Periodicity of the wave vector grid

In our solution we will be focusing only on the rectangular rooms with the regular wave vector grid (regular eigenvalue lattices in the wave vector space) [16] as shown in Figure 2. The *k-space* is an array of numbers representing spatial frequencies. According to theory, as long as we know the periodicity of the grid over each of the axes, it will provide us the knowledge on the room geometry as well as the values of the wave vectors of higher order. So the goal of our approach is the estimation of these three periods along each of the axes. Under the assumptions that the room is lightly damped, the 3 fundamental axial modes can be used as a basis to find all higher order modes. This will reduce the cutoff frequency of the analyzed data, which further reduces the density of the required grid of microphones, due to the dependencies between the temporal and spatial sampling as shown earlier.

3. PARAMETER ESTIMATION WITH PARTIAL COMPRESSED SENSING FOR STRUCTURED DATA

In 1985, Richardson et al. [17] have proposed a curve fitting algorithm allowing the reconstruction of the RTF curve from discrete measurements using room mode shaped functions as basic fitting elements. For different positions of the microphones/sound sources across the room, some parameters stay the same - *common parameters*: eigenfrequencies which depend on the room geometry, and the room mode damping which depends on the damping of the walls. The attenuation and the phase of the room modes are position dependent parameters - *specific parameters* which are expressed by different expansion coefficients and different spatial coordinates.

3.1. Acoustical properties of rectangular rooms

There are two key points for our parameter estimation procedure: how many room modes N do we expect up to a given cutoff frequency f_c and what are their approximate resonant frequencies ω_n? These are dependent on the room shape and size [16]. For a rectangular room of size $L_x \times L_y \times L_z$ angular eigenfrequencies are given by the expression: $\omega_n = \pi c \sqrt{\left(\frac{n_x}{L_x}\right)^2 + \left(\frac{n_y}{L_y}\right)^2 + \left(\frac{n_z}{L_z}\right)^2}$ where $(n_x, n_y, n_z) \in \mathbb{N}_0^3 \setminus (0,0,0)$ and approximate number of modes up to the cutoff frequency f_c is given by: $\tilde{N}_{f_c} \approx \frac{4\pi}{3} V \left(\frac{f_c}{c}\right)^3$ where $V = L_x L_y L_z$.

Of special interest will be the basic axial resonant frequencies: $\omega_{[1,0,0]} = \frac{\pi c}{L_x}$, $\omega_{[0,1,0]} = \frac{\pi c}{L_y}$ and $\omega_{[0,0,1]} = \frac{\pi c}{L_z}$, because they will provide the data about the shape of the room.

3.2. Reconstruction procedure

Our goal is to reconstruct spatial periods of the wave vector grid from low-pass room impulse responses over each of the axes. The size of the room is assumed to be unknown and is

---

Algorithm 1 ReSEMblE algorithm (Algorithm for the joint estimation of Room SizEs and ModEs)

Input: A set of measurements $\{m(x,y,z,t)\}_{i=1}^M$ at M known locations $\mathbf{X} = [x,y,z]^T$ in space and T points in time. $\mathbf{R} \in \mathbb{C}^{T \times M}$ are measurements in matrix form and $\mathbf{r} \in \mathbb{C}^{TM}$ are measurements in a vectorized form. f_p is frequency that separates data into 2 analysis procedures.

Output: Estimated room size $\tilde{L}_x, \tilde{L}_y, \tilde{L}_z$ and estimated room transfer function parameters:
- expansion coefficients $\{\alpha\}_{n=1,v=1}^{N,V}$,
- resonant frequencies $\{\omega\}_{n=1}^N$ and damping $\{\xi\}_{n=1}^N$
- wave vectors $\{\mathbf{k}\}_{n=1,v=1}^{N,V}$

N: number of modes,
V: number of wave vectors per wave number.

procedure RESEMBLE($\mathbf{R}$, $\mathbf{X}$)
  Separate the measurements with f_p: $\mathbf{R} = \mathbf{R}^l + \mathbf{R}^h$.
  for $i_l \in \{1, ..., N_l\}$ **do**
    step 1: estimate $(\omega_{i_l}, \xi_{i_l})$ from $\mathbf{R}_{i_l}^l$
    step 2: estimate $\mathbf{k}_{i_l}$ from $\mathbf{r}_{i_l}^l$
    step 3: compute new residual $\mathbf{R}_{i_l+1}^l$
  end for

  Recover the room size $\tilde{L}_x, \tilde{L}_y, \tilde{L}_z$ from basic axial room modes and form the regular wave vector grid.
  for $i_h \in \{N_l + 1, ..., N\}$ **do**
    step 1: get ω_{i_h} and $\mathbf{k}_{i_h}$ from the wave vector grid
    step 2: estimate ξ_{i_h} from $\mathbf{R}_{i_h}^h$
    step 3: compute new residual $\mathbf{R}_{i_h+1}^h$
  end for
  Estimate the expansion coefficients $\{\alpha\}_{n=1,v=1}^{N,V}$ using least square approach.
end procedure

---

jointly estimated. All measured signals are separated into two components: low-pass $\mathbf{R}^l$ and high-pass $\mathbf{R}^h$. Analysis procedure is first applied to the low-pass component, which includes the estimation of the wave numbers and corresponding wave vectors. The bandwidth of this low-pass analysis is chosen in such a way that it covers reasonable sizes of rooms and removes the false modes that can appear below the first mode in RTF. With $f \in [20, 70]$Hz we cover room dimensions $L_x, L_y, L_z \in [2.45, 8.575]$m for $c = 343\frac{\text{m}}{\text{s}}$. This can easily be adjusted for rooms of unusual sizes.

3.2.1. Estimation of ω_{i_l}, ξ_{i_l} and ξ_{i_h}

In the *low* part we define a unit-norm temporal dictionary with atoms of form: $\Theta[:, i] = \frac{\theta[i]}{\|\theta[i]\|}$, where $\theta[i] = e^{\xi_n[i]t} e^{j\omega_n[i]t}$ and i is an index on a 2D grid of possible (ω_n, ξ_n), $\omega_n \in [0, \pi f_s]$ and $\xi_n \in [10\xi_0, 0.1\xi_0]$, $\xi_0 = -3\frac{\ln 10}{RT_{60}}$. The atoms

with the highest correlation contains the solution pair. In the *high* part the frequency is known, so we have only a 1D grid of possible values for the damping, which leads to a much simplified search.

3.2.2. *Estimation of* $\mathbf{k}_{i_l}$

The estimation of wave vectors is done with a structured group sparsity assumption - after estimating the wave number, we construct a sphere with a radius $\frac{\omega_n}{c}$ which follows from the assumption of lightly damped modes. We define a non-unit-norm spatio-temporal dictionary with atoms of form: $\Sigma[:,i] = e^{\xi_{i_l}t}e^{j\omega_{i_l}t}e^{\mathbf{X}\cdot\mathbf{k}[i]}$, where $\mathbf{k}$ are samples on this uniformly sampled sphere [18].

On the surface of this sphere we search for a group of 8 wave vectors $[\pm k_x, \pm k_y, \pm k_z]^T$ which form a parallelepiped and which are aligned with the residual the most. In a case of tangential modes, the parallelepiped collapses over 1 dimension and shrinks to 4 wave vectors (e.g. $[\pm k_x, \pm k_y, 0]^T$), and axial modes are defined by 2 wave vectors (e.g. $[\pm k_x, 0, 0]^T$).

In each iteration the best subgroup of 8 atoms has been estimated by applying a simultaneous version of matching pursuit (MP) [19] and the new residual is estimated by an orthogonal projection onto the space spanned by the union of all of the subgroups that were previously selected.

4. RESULTS FOR RECONSTRUCTING THE K-SPACE OF A RECTANGULAR ROOM

In our solution we have relied on two types of structured sparsity expected in theory [16]: wave vector sparsity as nodes of parallelepiped and wave vector periodicity in the wave vector grid. How does this structured approach affect the data retrieval? As shown in [3, 10] efficient interpolation of the sound field is expected only within the part of the room surrounded by microphones used for training of the parameters.

We will present the performance of our approach on measurements made in a rectangular room with an approximate size 3m × 5.6m × 3.53m. Properties of the chosen room are observed in [20]. Microphones are distributed randomly inside a 1m side cube in one half of the room and the sound source is in the other half of the room. Since we were processing real measurements, in order to have an idea about the approximate value of some of the parameters we want to estimate, we have applied the rational fraction polynomial curve fitting [17] based on the room mode shaped polynomials as basic fitting elements. This way we have retrieved approximate resonant frequencies and mode damping factors. During the curve fitting process, our wave numbers $k_n = \frac{\omega_n + j\xi_n}{c}$ appear in the poles of the fitted function [21]: $p_\omega(\mathbf{X}) = \rho c^2 \omega q \sum_n \frac{\phi_n(\mathbf{X})\phi_n(\mathbf{X}_0)}{K_n[2\xi_n\omega_n + j(\omega^2 - \omega_n^2)]}$.

Figure 3. shows the results for the estimation of the room mode resonant frequencies and their position in the k -space in the *low* part of the algorithm with 20 microphones. Here
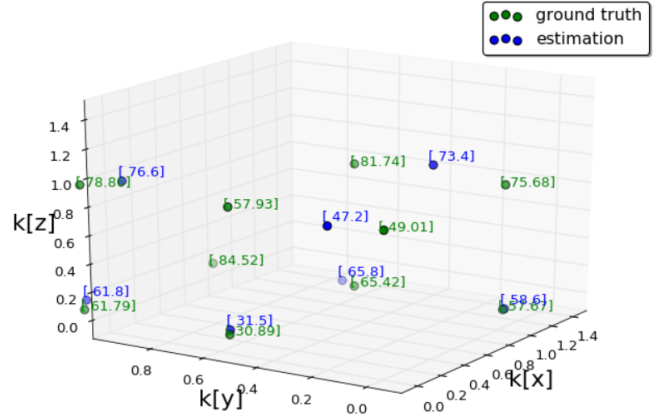

Fig. 3.   The estimation of wave vectors in k-space. The numbers next to the points indicate the corresponding eigenfrequencies (Hz). What we expect from theory in a case with perfectly rigid walls is plotted against the values we get from the measurements.

the f_p frequency was set to be 70Hz. The basic axial modes are easily recognized and they give a fine approximation of the room size up to a few cm away from ground truth. We can notice that the k_x and k_y component of the estimated wave vectors give a good approximation, but there is a slight deviation in the k_z direction. This is attributed to the fact that in the room where the measurements were performed the floor is made from wood and ceiling is made of concrete. Also the slight deviation of the eigenfrequencies can be attributed to the fact that the search of the wave vectors was performed with a rigid wall model $\|\mathbf{k}_{q,r}\| = |k_q|$.

After applying the *high* part of the algorithm, the Pearson correlation coefficient showed that the approximation is good (e.g. 82% for only 19-microphone setting and $f_c = 200$Hz), but it should be further improved once the nature of the deviation of the wave vectors is efficiently characterized.

5. CONCLUSION

The proposed solution is suitable only for rectangular shaped rooms that are lightly damped, which was confirmed by the experiments. Also, the sound source has to be put in a position such that it excites all the axial modes. Although that the solution requires the N, RT_{60} and c parameters to be know, solution is not sensitive to their slight perturbation. The estimation of approximate structure of the k-space has lead to the reduction in the terms of number of required measurements and in the increase of the speed of the reconstruction without great losses of quality, but not for a broad range of frequencies. The higher we take the frequencies, the greater become the deviations. In the spirit of reproducible research, we have decided to open our data and code.

6. REFERENCES

[1] T. Ajdler, L. Sbaiz, and M. Vetterli, "The plenacoustic function and its sampling," *IEEE Transactions on Signal Processing*, vol. 54, no. 10, pp. 3790–3804, Oct 2006.

[2] A. Moiola, R. Hiptmair, and I. Perugia, "Plane wave approximation of homogeneous helmholtz solutions," *Zeitschrift für angewandte Mathematik und Physik*, vol. 62, no. 5, pp. 809, Jul 2011.

[3] R. Mignot, G. Chardon, and L. Daudet, "Low frequency interpolation of room impulse responses using compressed sensing," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 22, no. 1, pp. 205–216, Jan 2014.

[4] D. L. Donoho, "Compressed sensing," *IEEE Trans. Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[5] E. J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theor.*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[6] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation: Part i: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572–588, Mar. 2006.

[7] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 948–958, June 2010.

[8] Y. Koyano, K. Yatabe, Y. Ikeda, and Y. Oikawa, "Physical-model based efficient data representation for many-channel microphone array," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2016, pp. 370–374.

[9] G. Vairetti, E. De Sena, M. Catrysse, S. H. Jensen, M. Moonen, and T. van Waterschoot, "A scalable algorithm for physically motivated and sparse approximation of room impulse responses with orthonormal basis functions," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 7, pp. 1547–1561, July 2017.

[10] N. Antonello, E. De 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 Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 10, pp. 1929–1941, Oct 2017.

[11] M. Kreković, I. Dokmanić, and M. Vetterli, "Omnidirectional bats, point-to-plane distances, and the price of uniqueness," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 3261–3265.

[12] Ivan Dokmanić, Reza Parhizkar, Andreas Walther, Yue M. Lu, and Martin Vetterli, "Acoustic echoes reveal room shape," *Proceedings of the National Academy of Sciences*, vol. 110, no. 30, pp. 12186–12191, 2013.

[13] H. Nyquist, "Certain topics in telegraph transmission theory," *Transactions of the American Institute of Electrical Engineers*, vol. 47, no. 2, pp. 617–644, April 1928.

[14] H. Lewy, K. Friedrichs, and R. Courant, "Über die partiellen differenzengleichungen der mathematischen physik," *Mathematische Annalen*, vol. 100, pp. 32–74, 1928.

[15] H. Peić Tukuljac, H. Lissek, and P. Vandergheynst, "Localization of sound sources in a room with one microphone," *SPIE, Wavelets and Sparsity XVII*, vol. 10394, 2017.

[16] H. Kuttruff and E. Mommertz, *Room Acoustics*, pp. 239–267, Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[17] Mark H. Richardson and David L. Formenti, "Global curve fitting of frequency response measurements using the rational fraction polynomial method," 1985.

[18] A. Semechko, "Suite of functions to perform uniform sampling of a sphere v 1.3, online," https://ch.mathworks.com/matlabcentral/fileexchange/37004-suite-of-functions-to-perform-uniform-sampling-of-a-sphere, 2015, [Online; accessed 05-October-2017].

[19] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *Trans. Sig. Proc.*, vol. 41, no. 12, pp. 3397–3415, Dec. 1993.

[20] R. Boulandet, J. Mosig, and H. Lissek, *Tunable Electroacoustic Resonators Through Active Impedance Control of Loudspeakers*, EPFL, PhD Thesis, 2012.

[21] E. T. J. L. Rivet, *Room Modal Equalisation with Electroacoustic Absorbers*, EPFL, PhD Thesis, 2016.