

# Compressively sampling the plenacoustic function

Rémi Mignot<sup>a,b</sup>, Gilles Chardon<sup>a</sup> and Laurent Daudet<sup>a</sup>

<sup>a</sup>Institut Langevin, ESPCI ParisTech - Univ Paris Diderot - UPMC Univ Paris 06 - CNRS UMR 7587, 10, rue Vauquelin, 75005 Paris, France.

<sup>b</sup>Institut Jean Le Rond d'Alembert, UPMC Univ Paris 06 - CNRS UMR 7190, 4, place Jussieu, 75005 Paris, France.

## ABSTRACT

Directly measuring the full set of acoustic impulse responses within a room would require an unreasonably large number of measurements. Considering that the acoustic wavefield is sparse in some dictionaries, Compressed Sensing allows the recovery of the full wavefield with a reduced set of measurements, but raises challenging computational and memory issues. Two practical algorithms are presented and compared: one that exploits the structured sparsity of the soundfield, with projections of the modes onto plane waves sharing the same wavenumber, and one that computes a sparse decomposition on a dictionary of independent plane waves with time / space variable separation.

**Keywords:** Compressed Sensing, Room Impulse Responses, Plane waves, Interpolation, Sparsity.

## 1. INTRODUCTION

The *Room Impulse Response* (RIR) characterizes the sound transmission in a room between a source and a receiver; it is associated to the reverberation of a room. Assembling all RIRs makes a function which depends on time, on the source position, on the receiver position and on the room characteristics (geometry and wall properties). In Ref. 1, this function is named the *Plenacoustic Function* (PAF). On one hand, in some applications the effect of room reverberation is undesirable and acoustic echo cancelers are used to estimate the anechoic sound. On the other hand, reverberation plays an important role in auditory scene synthesis, in virtual reality for example. In both cases, knowing the whole set of RIRs in a given room could potentially be used to improve their performance. Measuring the PAF is fundamentally a sampling problem: from a limited number of point measurements, the goal is to reconstruct (i.e. interpolate) the RIR at any position in space.

Standard acquisition of signals relies on a regular sampling of space and time with respect to Shannon-Nyquist theory. At a given temporal frequency, the space sampling has to be dense enough to avoid aliasing in reconstruction and interpolation<sup>1</sup>. However, the measurement of a time varying 3-D image requires a too high number of microphones to be realized as such in practice. Nevertheless, informed by the physical nature of the measured signal, we can reduce the number of sampling locations. This number is directly linked to the number of microphones if one wants to acquire the signals simultaneously, in a microphone array setting. For example, in Ref. 2 a method based on Dynamic Time Warping is used for the interpolation of the early part of the RIRs. Another example is given in Ref. 3 that uses an acoustic model of rooms. This model is based on the modal theory and assumes that all RIRs share the same damped complex sinusoids (associated to common poles) with different amplitudes (residues). After the estimation of poles, these residues are estimated for each source position on a line considering a space dependency as a cosine function. Whereas the first method can interpolate the early part of the RIRs, the second one is adapted to the interpolation of the whole RIRs only in low frequencies.

In this paper we study the sampling and the interpolation of RIRs in low frequencies within a 3-D domain  $\Omega$  of the space, using the *Compressed Sensing* paradigm (CS): this principle allows to reduce the number of measurements if the signal is sparse (even approximately) in some domain. Here, this sparsity property is based on the modal theory. Although based on very different principles, the proposed method can be seen as an extension of Ref. 3, adapted for 3-D domains.

The outline of this paper is as follows. Section 2 recalls the basics of uniform sampling. In section 3, we exhibit a sparsity property and we propose two approaches based on CS. In section 4, we give details on algorithm implementations. Results are presented in section 5, before concluding remarks in section 6.

---

Further author information:

R.M.: E-mail: remi.mignot@epsci.fr, G.C.: E-mail: gilles.chardon@epsci.fr,

L.D.: E-mail: laurent.daudet@epsci.fr, Telephone: +33 1 40 79 52 16,

L.D. is on a joint affiliation with Université Paris Diderot and the Institut Universitaire de France.

This work is supported by the Agence Nationale de la Recherche (ANR), project ECHANGE (ANR-08-EMER-006).

## 2. UNIFORM SAMPLING

The Plenacoustic Function (PAF) gives the acoustic information that is being transferred between any source and receiver. It can be described as the set of all Room Impulse Responses (RIRs), for all source / receiver positions. Note that, due to reciprocity properties, sources and receivers play symmetric roles. Then, considering a fixed source (as this corresponds to our experimental setup), in this section we recall how standard sampling of the PAF can be done, using a uniform 3-D microphone array, as a function of the position  $\vec{X} = [x, y, z]^T$  of the receiver within a volume  $\Omega$  of the space.

The primary design parameter is the temporal bandwidth that is required for the applications at hand. If the maximum frequency is fixed at  $f_c$  [Hz], higher frequencies are removed with an analog low-pass filter and, assuming ideal filters, the sampling in time is done at a rate  $F_s > 2f_c$ . Depending on this temporal frequency bandwidth, the distance between microphones (sampling in space) has to be small enough to avoid spatial aliasing. In this section we present the spectrum of the PAF to define a criterion for the sampling. The PAF will be denoted by the space/time dependent function:  $p(t, \vec{X})$ .

### 2.1 Spectrum and sampling

In Ref. 1, Ajdler studied the spectrum of the PAF on a line parallel to the  $(Ox)$  axis. With  $\omega$  [rad.s<sup>-1</sup>] the time frequency and  $\varphi_x$  [rad.m<sup>-1</sup>] the spatial frequency, he observed that the energy of the 2D-FT  $\hat{p}(\omega, \varphi_x) = \text{TF}\{p(t, x)\}$  is mainly concentrated within the triangle bounded by  $|\varphi_x| \leq |\omega|/c_0$ , which corresponds to the dispersion relation of propagative waves. Then, for growing  $\varphi_x$ , he demonstrated that  $\hat{p}$  decreases faster than an exponential for  $|\varphi_x| > |\omega|/c_0$ , which corresponds to evanescent waves. From this, he determines the *sampling theorem* which describes how to sample in space for a target Signal-to-Noise Ratio SNR<sub>0</sub>:

$$\phi_x = \frac{2\pi}{\delta_x} > \frac{2\omega_c}{c_0} + \varepsilon(\text{SNR}_0, \omega_c), \quad (1)$$

where  $\delta_x$  is the spatial sampling step on the line,  $\omega_c = 2\pi f_c$  is the maximal frequency,  $c_0$  is the sound velocity, and  $\varepsilon$  is given if Ref. 1. Under the far field assumption, evanescent waves are negligible, which leads to  $\varepsilon = 0$ ; the sampling theorem then becomes:  $\delta_x < \pi c_0 / \omega_c$ .

In the case of 2-D sampling (in a plane parallel to  $(Oxy)$ ), under the far field assumption, the 3D-FT of the PAF,  $\hat{p}(\omega, \varphi_x, \varphi_y)$ , has its support in the cone of equation  $\varphi_x^2 + \varphi_y^2 \leq \omega^2/c_0^2$ , where  $\varphi_x$  and  $\varphi_y$  are the spatial frequencies along axes  $(Ox)$  and  $(Oy)$ . We have a similar result in the case of a 3-D sampling: the support of the spectrum of the PAF  $\hat{p}(\omega, \varphi_x, \varphi_y, \varphi_z)$  is such that  $\varphi_x^2 + \varphi_y^2 + \varphi_z^2 \leq \omega^2/c_0^2$ .

Finally, in any case, to avoid spatial aliasing we have to choose sampling steps that satisfy the sampling theorem:

$$\delta_v < \frac{\pi c_0}{\omega_c}, \quad \forall v \in \{x, y, z\}. \quad (2)$$

### 2.2 Reconstruction

The sampling of the PAF gives  $p(t_n, \vec{X}_m)$  for  $t_n = n/F_s$  and  $\vec{X}_m$  on a spatial grid. The reconstruction of the PAF for any time and position is done using a 4-D interpolation filter, which may be separable in time and space.

In theory, the ideal reconstruction should be performed using convolution with a sinc function which has infinite support, therefore requiring an infinite number of sampling points, in time and space. Because of the exponential time decay of the RIRs, the responses can be truncated in time, and using finite length filters provides good approximations.

However, in space this problem persists because a precise interpolation requires an overly large number of microphones. Actually, in order to reconstruct the PAF within a sub-domain  $\Omega$  of the room, in practice there are 2 possible strategies:

- by fixing the spatial sampling step  $\delta$  according to Shannon-Nyquist requirements, one has to increase the order of the 3-D interpolation filter (in space) in order to improve the reconstruction. Consequently, the microphone array must be larger than  $\Omega$ , and according to the desired quality, the number of microphones may be unrealistic in practice.
- by fixing the size of the array, one can improve the quality by taking a finer grid. Even if the array does not become bigger, the number of microphones increases, and boundary effects may still be present. As in the previous case, the number of microphones may be too high in practice.

In section 5, we compute the reconstruction of the PAF in a cube with sides of 1.7m, the uniform sampling is then compared with two new algorithms. The next section presents these new algorithms.

### 3. PLENACOUSTIC AND SPARSITY

In this section, we study the acoustic propagation within an enclosed space, and we show that it exhibits some sparsity properties. This validate the use of Compressed Sensing (CS) techniques in the following.

#### 3.1 Structured sparsity

Considering linear acoustic propagation away from the sources, the acoustic pressure  $p(t, \vec{X})$  is governed by the Wave Equation  $c_0^2 \Delta p(t, \vec{X}) - \partial_t^2 p(t, \vec{X}) = 0$ , where  $\Delta = \nabla^2$  is the laplacian operator and  $\partial_t$  is the time derivative. Assuming a modal behavior (at low frequencies) for closed rooms with ideally rigid walls, the solution can be decomposed as a discrete sum of complex harmonic signals with the angular frequencies  $\omega_q$ :

$$p(t, \vec{X}) = \sum_{q \in \mathbb{Z}^*} A_q \phi_q(\vec{X}) g_q(t), \quad (3)$$

where  $g_q(t) = e^{j\omega_q t}$ ,  $\phi_q$  is the modal shape of the mode  $q$  and  $A_q$  is a related complex amplitude. Note that the  $\omega_q$ 's and the  $\phi_q$ 's depend on the boundary conditions (room geometry and wall properties), while the  $A_q$ 's depend on the initial conditions. With the wavenumber  $k_q = \omega_q/c_0$ , we get the Helmholtz equation for every mode:  $\Delta \phi_q + k_q^2 \phi_q = 0$ .

In the Helmholtz equation,  $\phi_q$  is the eigenmode of the laplacian operator with eigenvalue  $-k_q^2$ . If the room is star-shaped, previous studies<sup>4,5</sup> have shown that an eigenmode of the laplacian with a negative eigenvalue can be approximated by a finite sum of plane waves incoming from various directions, and sharing the same wavenumber  $k$ . Then

$$\phi_q(\vec{X}) \approx \sum_{r=1}^R a_{q,r} e^{j\vec{k}_{q,r} \vec{X}} \quad (4)$$

is the  $R$ -order approximation of  $\phi_q$ , with  $\vec{k}_{q,r}$  the 3-D wavevector  $r$  of the mode  $q$ , such that  $\|\vec{k}_{q,r}\|_2 = |k_q|$ .

In the case of non rigid walls, the modes are damped in time,  $k_q$  now has an imaginary part:  $k_q = (\omega_q - j\xi_q)/c_0$ , where  $\xi_q < 0$  is the damping coefficient. Therefore,  $g_q(t)$  of eq. (3) becomes:  $g_q(t) = e^{j k_q c_0 t} = e^{\xi_q t} e^{j\omega_q t}$ . In theory, these losses modify  $\phi_q$ , nevertheless we assume that the approximation (4) remains valid, at least for  $\vec{X}$  far from the walls.

Consequently, considering a finite frequency range  $[0, \omega_c]$  containing  $Q$  real modes, or equivalently  $2Q$  complex modes, and considering  $R$ -order approximations of the  $\phi_q$ 's, the PAF  $p(t, \vec{X})$  can be approximated by a sum of  $2QR$  damped harmonic plane waves,  $\exp(j(k_q c_0 t + \vec{k}_{q,r} \vec{X}))$ , with coefficients linked by the relation  $\alpha_{q,r} = A_q a_{q,r}$ .

Now, taking advantage of this *Structured Sparsity*, we present an algorithm previously proposed for the near-field acoustic holography of plates<sup>6</sup>. First, using an array of  $M$  microphones placed at the  $\vec{X}_m$  sampling points within  $\Omega$  (with uniform or random sampling, cf. sec. 3.2.2), we acquire the digital signals  $p(t_n, \vec{X}_m)$ , of length  $N$  samples each. Second, we can reconstruct the PAF using the following algorithm:

- (a) The shared wavenumbers  $k_q$  are estimated using a joint estimation of damped sinusoidal components (using for example the algorithms MUSIC<sup>7</sup>, ESPRIT<sup>8</sup>, or SOMP<sup>9</sup>). Note that this stage corresponds to a sparse decomposition of the  $M$  signals using a joint sparsity model with damped sinusoids.
- (b) The matrix  $P$  of signals  $p(t_n, \vec{X}_m)$  can be written as  $P = \Phi G$ , where  $\Phi$  is the matrix of modes,  $\Phi_{[m,q]} = A_q \phi_q(\vec{X}_m)$ , and  $G$  is the dictionary of damped exponentials,  $G_{[q,n]} = e^{j k_q c_0 t_n}$ ; see equation (3). Then  $\Phi$  is estimated using the  $\ell_2$  optimization:  $\tilde{\Phi} = P G^H (G G^H)^{-1}$ .
- (c) From (4),  $\phi_q \approx \psi_q \alpha_q$  where  $\psi_q$  is the matrix of the plane waves,  $\psi_{q[m,r]} = e^{j \vec{k}_{q,r} \vec{X}_m}$ , sharing the same wavenumber. The  $\vec{k}_{q,r}$ 's are chosen using a uniform sampling of the sphere<sup>10,11</sup> of radius  $|\omega_q|/c_0$ . Then, the coefficients  $\alpha_{q,r}$  are estimated using the projection of every  $\tilde{\phi}_q$  into the corresponding basis of  $\psi_q$  as follows:  $\tilde{\alpha}_q = (\psi_q^H \psi_q)^{-1} \psi_q^H \tilde{\phi}_q$ .
- (d) Finally, the PAF can be interpolated for any  $t \in [0, N/F_s]$  and at any position  $\vec{X} \in \Omega$  using the approximation:  $\tilde{p}(t, \vec{X}) = \sum_q \tilde{\alpha}_{q,r} e^{j(k_q c_0 t + \vec{k}_{q,r} \vec{X})}$ .

Note that  $P$  is a real matrix, hence the coefficients of the complex modes have to obey the hermitian symmetry. This implies:  $\alpha_{q,r} = \alpha_{-q,r}^*$ ,  $k_q = -k_{-q}^*$  and  $\vec{k}_{q,r} = -\vec{k}_{-q,r}$ , where the symbol  $*$  denotes the conjugate. Actually, this hermitian symmetry is used in stages (b) and (c) in order to reduce the size of matrices.

The number  $Q$  of modes is chosen according to a modal analysis of the room<sup>12</sup>. It increases with the bandwidth, and if  $Q > N$  stages (a) and (b) cannot be led. The reason is that only the time information is exploited in these stages. The next section presents a stronger sparsity property, which takes into account simultaneously the information of time and space.

## 3.2 Plane wave sparsity

In this section, we study the solutions of the wave equation in the simple case of a rectangular room. From this study, we exhibit a stronger property of sparsity which justifies the use of the Compressed Sensing framework (CS).

### 3.2.1 Modal analysis in a rectangular room

In the case of a rectangular room with rigid walls, we can make the variable separation<sup>12</sup> in cartesian coordinates  $(x, y, z)$ . Then, each modal shape is written as the product of 3 functions of one variable. With  $\vec{X} = [x, y, z]^T$ , the PAF becomes:

$$p(t, \vec{X}) = \sum_{q \in \mathbb{Z}^*} A_q F_{xq}(x) F_{yq}(y) F_{zq}(z) e^{jk_q c_0 t}. \quad (5)$$

For each mode  $q$ , these functions verify the 1-D Helmholtz equation  $\partial_v^2 F_v + k_v^2 F_v = 0$  for  $v \in \{x, y, z\}$ . With rigid walls, the  $k_v$ 's are real constants such that  $k_x^2 + k_y^2 + k_z^2 = k^2$  (cf. Ref. 12). According to the Helmholtz equation, for each cartesian coordinate  $v$  the  $F_v$ 's are the sum of 2 solutions:  $F_v(v) = A_v^+ e^{jk_v v} + A_v^- e^{-jk_v v}$ . Then, expanding  $F_x F_y F_z$ , the modal shape  $\phi_q(\vec{X})$  is written as the sum of 8 plane waves  $e^{\pm jk_x x \pm jk_y y \pm jk_z z} = e^{j\vec{k}\vec{X}}$ , with  $\vec{k} = [\pm k_x, \pm k_y, \pm k_z]^T$ .

In the case of non rigid walls, as the wavenumber  $k$  is complex:  $k = (\omega - j\xi)/c_0$ , the  $k_v$ 's are complex too. This implies a slight decrease of the  $F_v$ 's near the walls. Nevertheless, for  $\vec{X}$  far from the walls, we assume that the imaginary part of  $k_v$  is negligible, and that  $k_x^2 + k_y^2 + k_z^2 = \mathcal{R}_e(k)^2 = \omega^2/c_0^2$ .

Note that in the case of a rectangular room, the wavevectors  $\vec{k} = [\pm k_x, \pm k_y, \pm k_z]^T$  are at the vertices of an inscribed parallelepiped of the sphere with radius  $|\omega|/c_0$ . Moreover, whereas the modal density\* strongly increases with the frequency, all the wavevectors are uniformly spaced<sup>12</sup> in the  $\vec{k}$ -space (with coordinates  $k_x, k_y, k_z$ ).

Consequently, in a bandwidth containing  $2Q$  complex modes, the PAF can be written as the sum of  $16Q$  harmonic plane waves in the case of rectangular rooms. Note that in the previous section, each modal shape was approximated by  $R$  fixed plane waves sampling uniformly the sphere of radius  $|\omega|/c_0$ , whereas here, with the assumption of rectangular room, only 8 plane waves are required by mode. Then, this stronger sparsity property justifies the use of CS techniques.

Although this model doesn't hold for arbitrary geometries (cylindrical rooms for example), it can nevertheless be extended to non rectangular rooms. Indeed, if all walls are plane, we can assume that the modal shapes are still sparse on a dictionary of plane waves. The corresponding wavevectors are not necessarily at the vertices of a parallelepiped, but they are always on the sphere.

### 3.2.2 Compressed Sensing framework

The general problem consists in the reconstruction of a signal  $y \in \mathbb{R}^{\mathcal{N}}$  from  $\mathcal{M}$  observations  $x_m$ , linked by the linear system  $x = \Phi y$ . *Compressed Sensing* (CS) deals with the underdetermined case, for which there are more unknowns than equations ( $\mathcal{N} > \mathcal{M}$ ), cf. e.g. Refs. 13, 14. As such a problem cannot be solved without additional hypothesis, the underlying idea is that if  $y$  lives in a subspace of dimension  $\mathcal{K}$  and with basis  $\psi$ , for  $\mathcal{K} < \mathcal{M}$ , we can solve  $y = \psi a$  writing  $x = \Phi y = \Phi \psi a = \theta a$ . However, in general we do not know  $\psi$ .

Then, we define  $\mathcal{L}$  vectors  $\psi_l$ , forming the matrix  $\Psi$  with  $\mathcal{L} \gg \mathcal{K}$ , and we look for a basis which explains  $y$ . In other words, we look for a vector  $\alpha \in \mathbb{R}^{\mathcal{L}}$   $K$ -sparse (where no more than  $K$  coefficients are non-zero), such that  $y = \Psi \alpha$ . Unfortunately this problem is not convex and difficult to solve. However, we can change it into a convex problem by considering the following *Basis Pursuit Denoising* approach:

$$\min_{\alpha \in \mathbb{R}^{\mathcal{L}}} \|\alpha\|_{\ell_1} \quad \text{subject to} \quad \|x - \Phi \Psi \alpha\|_{\ell_2} \leq \varepsilon, \quad (6)$$

where the norm  $\ell_n$  is given by  $\|y\|_{\ell_n} = (\sum_i |y_i|^n)^{1/n}$ , and  $\varepsilon$  is a fidelity parameter. A high  $\varepsilon$  allows a stronger sparsity of  $\alpha$ , and a small  $\varepsilon$  improves the reconstruction of  $y$ .

\*The modal density is related to the number of modes per frequency range.

Some theoretical results (cf. e.g. Refs. 15–17) give a sufficient condition for reconstructing  $y$  in the case of sparse signals, by the so-called Restricted Isometry Property (RIP). It quantifies how  $\Phi$  and  $\Psi$  are mutually incoherent with respect to their use on sparse signals. In practice, the RIP is difficult to compute, but it is verified with high probability for some random sampling matrices. This encourages the use of randomly selected observation points in practice, which are here the microphone positions in the 3-D space. Note that, conversely, a regular sampling grid might lead to a strong correlation with plane waves, whenever the wavenumber gets close to be aligned to one of the  $x$ ,  $y$  or  $z$  axis : such standard sampling scheme is therefore likely to be suboptimal in the CS framework.

### 3.2.3 Reformulation of the problem in a Compressed Sensing framework

Now, we can reformulate our problem as follows: let us define  $\mathcal{S}_x$  the signal vector of the measurements  $p(t_n, \vec{X}_m)$ , and  $\mathcal{S}_y$  the signal vector that we wish to reconstruct (interpolate) on a uniform grid of the space:

$$\mathcal{S}_x_{[(n+1)+(m-1)N]} = p(t_n, \vec{X}_m), \quad \text{and} \quad \mathcal{S}_y_{[(n+1)+(s-1)N]} = p(t_n, \vec{Y}_s), \quad (7)$$

where the  $\vec{X}_m$ 's are the positions of the  $M$  microphones of the array, and the  $\vec{Y}_s$ 's are the positions of the 3-D grid. Considering the ideal reconstruction using a uniform array (cf. sec. 2),  $\mathcal{S}_x$  and  $\mathcal{S}_y$  are linked by  $\mathcal{S}_x = \Phi_{xy}\mathcal{S}_y$ , where  $\Phi_{xy}$  is an interpolation matrix representing the spatial convolution for interpolating the PAF at  $\vec{X}_m$  starting from the signals on the grid of the  $\vec{Y}_s$ 's.

Since the number of microphones is limited in practice, we cannot directly reconstruct  $\mathcal{S}_y$  from  $\mathcal{S}_x$ . However, thanks to the sparsity property of the PAF as described in sec. 3.2.1, it is possible to solve this problem using CS. The rough idea is to define an oversized dictionary  $\Psi_y$  with harmonic plane waves which are ‘‘virtually’’ sampled on the grid. Then, writing  $\mathcal{S}_y = \Psi_y\alpha$ , in principle the problem might be solved with  $\mathcal{S}_x = \Phi_{xy}\Psi_y\alpha$ . Unfortunately because of the space dimensionality (4-D), standard  $\ell_1$  optimization algorithms of (6) would require too much memory and cannot be run on standard computers. Hence in the next section we propose a greedy algorithm for the interpolation of the PAF.

## 4. ALGORITHMIC DETAILS

When  $\ell_1$  optimization procedures cannot be processed because of computational issues, greedy algorithms such as Matching Pursuit are commonly used. However, with the size of data in this work, even this algorithm is too cumbersome to be computed in practice. In this section, first standard Matching Pursuit is presented, then we propose a derived version which can be applied for the sampling of the PAF in 3-D.

### 4.1 Matching Pursuit

*Matching Pursuit*<sup>18</sup> consists in iteratively subtracting from the signal the atom that best approximates it. This atom  $g$  is chosen among the columns of a dictionary matrix  $\Psi$ , of size  $(\mathcal{M}\times\mathcal{L})$ . Then the process is iterated on the residual which is, at the iteration  $i+1$ :

$$r_{i+1} = r_i - \alpha_i g_i, \quad (8)$$

with  $r_1$  the signal to approximate, and where the vector  $g_i$  and the coefficient  $\alpha_i$  are chosen to minimize  $\|r_{i+1}\|_{\ell_2}$ . If the column vectors of  $\Psi$  are normalized, the optimal atom is  $g_i = \arg \max_{g \in \Psi} |\langle g, r_i \rangle|$  and the optimal coefficient is given by the correlation  $\alpha_i = \langle g_i, r_i \rangle := g_i^H r_i$ . The symbol  $\cdot^H$  denotes the conjugate transpose of a complex matrix or a vector.

A similar method consists in searching at each iteration a group of  $P$  atoms simultaneously minimizing the norm of the residual  $r_{i+1} = r_i - G\alpha$ , where  $G$  is a  $(\mathcal{M}\times P)$  matrix of  $P$  atoms, and  $\alpha$  is a  $(P\times 1)$  vector. If the atoms are normalized, and if  $\text{rank}(G) = P$  with  $P < \mathcal{M}$ , the optimal matrix  $G_i$  minimizes

$$\|r_{i+1}\|_{\ell_2}^2 = \|r_i\|_{\ell_2}^2 - r_i^H G (G^H G)^{-1} G^H r_i, \quad (9)$$

and the weight vector is then  $\alpha = (G^H G)^{-1} G^H r_i = G^\dagger r_i$ , where the symbol  $\cdot^\dagger$  denotes the pseudo-inverse of a matrix.

In the present work, we first considered the application of Matching Pursuit considering groups of  $P$  harmonic plane waves which share the same wavenumber. For example, the rectangular room considered in section 3.2.1 led to  $P = 8$ . Unfortunately, because of the dimensionality of the problem, it is not possible to use this algorithm as such. Indeed, among a high number of possible wavenumbers  $k = (\omega - j\xi)/c_0$  (that belong to a subspace of dimension 2), we would have to test a wider number of possible combinations of  $P$  plane waves on the sphere of radius  $\omega/c_0$  (in a subspace of dimension  $2P$ ). Consequently, the matrices  $G$  live in a subspace of dimension  $2 + 2P$ , and exhaustive search for the most correlated vector is in practice impossible. In the next section, we propose a modified algorithm which alleviates this problem.

## 4.2 Modified algorithm

Let us define  $S$  the  $(N \times M)$  signal matrix such that  $S_{[n,m]} = p(t_n, \vec{X}_m)$ , and  $\mathcal{S}$  its vectorized version as in equation (7),  $\mathcal{S} = \mathcal{S}_x$ . The residual vectors will be noted  $\mathcal{R}_i$ , and their  $(N \times M)$  matrix versions  $R_i$ .

### 4.2.1 Analysis

The principle of the proposed algorithm is as follows: at every iteration  $i$ , first we choose the damped complex exponential which best approximates the  $M$  columns of  $R_i$  (time signal vectors), and so a wavenumber  $k_i = (\omega_i - j\xi_i)/c_0$  is estimated. Then we choose a group of  $P$  harmonic plane waves (on the sphere of radius  $\omega_i/c_0$ ) which efficiently explains the residual  $\mathcal{R}_i$ . For more details, the 4 stages of the iteration  $i$  are detailed here:

- (A) This stage is similar to the search of poles of SOMP<sup>9</sup>. We define the  $(N \times L_t)$  time dictionary matrix  $\Theta$  with  $L_t$  columns  $\theta_\ell$  which are damped complex exponentials:  $\Theta_{[n,\ell]} = \theta_{\ell[n]} = e^{\xi_\ell t_n} e^{j\omega_\ell t_n}$ , with  $\omega_\ell \in [0, \omega_c]$  and  $\xi_\ell < 0$ . Then defining the  $(L_t \times M)$  correlation matrix  $\eta_i := |\bar{\Theta}^H R_i|$ , we choose the index  $\ell_i$  which maximizes the sum of energies:  $\sum_m (\eta_{i[\ell,m]})^2$ . Here,  $\bar{\Theta}$  corresponds to  $\Theta$  where the columns are individually normalized:  $\bar{\theta}_\ell = \theta_\ell / \|\theta_\ell\|_{\ell_2}$ .
- (B) With the estimated wavenumber  $k_{\ell_i}$ , we define an  $(MN \times L_s)$  dictionary matrix  $\Delta_i$  with  $L_s$  columns  $\delta_{i,\ell}$  which are harmonic plane waves:  $\delta_{i,\ell} = e^{\xi_{\ell_i} t_n} e^{j\omega_{\ell_i} t_n} e^{j\vec{k}_\ell \vec{X}_m}$ , with  $\|\vec{k}_\ell\|_2 = \omega_{\ell_i}/c_0, \forall \ell \in [1, L_s]$ . Then, with  $W > P$ , we isolate  $W$  atoms  $\delta_{i,\ell}$  which are the maxima of  $\rho_{i[\ell]} := |\langle \delta_{i,\ell}, \mathcal{R}_i \rangle|$ . Note that  $\rho_i$  can be written  $\rho_i = |\bar{\Delta}_i^H \mathcal{R}_i|$ . Actually, because of possible lobes,  $\ell$  must index a 2-D grid of the sphere of radius  $\omega_{\ell_i}/c_0$ , and the chosen atoms are the  $W$  higher local maxima.
- (C) Among these  $W$  atoms, we test all combinations of  $P$  atoms (there are  $\binom{W}{P}$  possible combinations). Then we choose the combination  $G_i$  which minimizes (9):  $\|\mathcal{R}_{i+1}\|_{\ell_2}^2 = \|\mathcal{R}_i\|_{\ell_2}^2 - \mathcal{R}_i^H G_i G_i^\dagger \mathcal{R}_i$ . with  $G$  an  $(MN \times P)$  matrix of one combination of  $P$  vectors.
- (D) Finally, the best combination  $G_i$  is subtracted. Actually, here we have to consider the hermitian symmetry for real signals, and so defining  $\mathcal{G}_i = [G_i, G_i^*]$ , the residual  $i+1$  is:  $\mathcal{R}_{i+1} = \mathcal{R}_i - \bar{\mathcal{G}}_i \alpha_i$ , with  $\alpha_i = \bar{\mathcal{G}}_i^\dagger \mathcal{R}_i$ .

Compared to the standard Matching Pursuit algorithm presented in section 4.1, this new algorithm requires less time: in the first stage, no space information is used, and  $\Theta$  is a relatively small matrix  $(N \times L_t)$ , independent from  $i$ . Moreover, if the frequency grid of the  $\omega_\ell$ 's is uniform between 0 and  $F_s/2$ , using the Fast Fourier Transform makes the computation of  $\eta_i$  faster. Then in the second stage, the atoms are individually tested on a sphere (subspace of dimension 2), which facilitates the process; and only the  $W$  best atoms are selected for the stage (C). If  $W$  is properly chosen, the number  $\binom{W}{P}$  of possible combinations remains reasonable.

Moreover, in stages (B), (C) and (D), taking advantage of the variable separation, we can significantly reduce the sizes of the matrices. For example in stage (B),  $\Delta_i$  is an  $(MN \times L_s)$  matrix, while  $\rho_i$  can be computed by  $\rho_i = |\bar{\theta}_{\ell_i}^H R_i \bar{\Phi}_i^*|$ , where  $\theta_{\ell_i}$  is the  $(N \times 1)$  vector chosen in stage (A),  $\theta_{\ell_i[n]} = e^{\xi_{\ell_i} t_n} e^{j\omega_{\ell_i} t_n}$ , and  $\bar{\Phi}_i$  is an  $(M \times L_s)$  space dictionary matrix such that:  $\bar{\Phi}_{i[m,\ell]} = e^{j\vec{k}_\ell \vec{X}_m}$ , with  $\|\vec{k}_\ell\|_2 = \omega_{\ell_i}/c_0$ .

Knowing the number of theoretical modes  $Q$ , which depends of the room geometry (cf. e.g. Ref. 12), as in principle every group of  $P$  plane waves corresponds to a mode, in this work we choose the number of iterations equal to  $Q$ .

### 4.2.2 Projection and interpolation

At the end of the  $Q$  iterations, we get  $V = QP$  estimated harmonic plane waves, with wavenumber  $k_v$  and wavevector  $\vec{k}_v$ . They define the atoms of the  $(MN \times V)$  basis matrix  $B_x$ :

$$B_x_{[(n+1)+(m-1)N, v]} = e^{jk_v t_n} e^{j\vec{k}_v \vec{X}_m}, \text{ with } \|\vec{k}_v\|_2 = \omega_v/c_0 > 0, \text{ and } k_v = (\omega_v - j\xi_v)/c_0. \quad (10)$$

Then with  $A_x := [B_x, B_x^*]$ , considering positive and negative frequencies, we could solve the optimal solution  $\tilde{\mathcal{S}}_x = A_x a$  in the least mean square sense with  $a = A_x^\dagger \mathcal{S}_x$ , which would require complex calculus. In order to reduce memory requirements and makes computation faster, it is preferable to manipulate only real coefficients. Then  $a$  is obtained as follows:

$$a_{[v]} = a_{[v+V]}^* = \mu_{[v]} + j\mu_{[v+V]}, \quad \text{with } v \in [1, V] \quad \text{and } \mu = \frac{1}{2} \left[ \mathcal{R}_e\{B_x\}, -\mathcal{I}_m\{B_x\} \right]^\dagger \mathcal{S}_x. \quad (11)$$

If the problem of (11) is ill-conditioned, in practice we remove some atoms of  $B_x$  which are linearly close to some others. Note that the use of an orthogonal projection in stage (D) (as with the *Orthogonal Matching Pursuit*<sup>18</sup> algorithm) would partly solve this issue, but the associated computational cost would be prohibitive for the problem at hand.

Finally, the interpolation at any position  $\vec{Y} \in \Omega$  and any time  $t \in [0, N/F_s]$ , is done by:

$$\tilde{p}(t, \vec{Y}) = \sum_{v=1}^V a_{[v]} e^{jk_v t} e^{j\vec{k}_v \vec{Y}}, \quad (12)$$

or  $\tilde{S}_y = A_y a$  where  $A_y$  corresponds to the matrix basis of harmonic plane waves at the position  $\vec{Y}$ . Note that while the matrices are normalized in section 4.2.1, in (11) and (12),  $B_x$  and  $A_y$  are not normalized.

## 5. EXPERIMENTS AND RESULTS

In the following, we present some results of the two algorithms of sections 3.1 and 4.2. They will be named method CS1 and method CS2 respectively. Note that here, the first stage of CS1 uses SOMP<sup>9</sup>. The interpolation is evaluated using the Signal-to-Noise Ratio (SNR) [dB] and the Pearson correlation coefficient  $c$  [%]. With  $s$  the  $(N \times 1)$  vector of the target RIR, such that  $s_{[n]} = p(t_n, \vec{X})$ , and  $\tilde{s}$  its interpolation:  $\text{SNR}_{dB} = 20 \log(\|s\|_{\ell_2} / \|s - \tilde{s}\|_{\ell_2})$  and  $c = 100 |\langle s, \tilde{s} \rangle| / (\|s\|_{\ell_2} \cdot \|\tilde{s}\|_{\ell_2})$ .

Table 1 compares the uniform sampling to the methods CS1 and CS2. Here, we aim at reconstructing the RIRs within a cube  $\Omega$  of side 1.7m, starting from  $M$  simulated RIRs<sup>19,20</sup> of a virtual array (regular for the uniform sampling, random for CS1 and CS2). The cutoff frequency is  $f_c = 250\text{Hz}$ , the sampling rate is  $F_s = 750\text{Hz}$ , and the SNRs are averaged over 2744 interpolation positions in  $\Omega$ . Note that the regular array respects the Shannon theorem in any case:  $\delta_x < 2c_0/f_c$ . We observe that, for a given number of microphones, method CS1 significantly outperforms uniform sampling (cf.  $M = 64$  or 125). Equivalently, method CS1 can obtain equivalent performance as regular sampling, but with a smaller number of microphones (see for instance  $M = 64$  for CS1 and  $M = 216$  for the uniform sampling). According to these preliminary results, method CS2 does not seem to be competitive; we will see its benefits at a later stage.

| Methods                     | uniform sampling |      |      | method CS1 |      |      | method CS2 |      |      |
|-----------------------------|------------------|------|------|------------|------|------|------------|------|------|
| $M$ (number of microphones) | 64               | 125  | 216  | 64         | 96   | 125  | 64         | 96   | 125  |
| SNR [dB]                    | 15.9             | 21.7 | 25.0 | 24.2       | 29.6 | 30.6 | 13.6       | 14.8 | 15.4 |

Table 1. Comparison between the uniform sampling and methods CS1 and CS2.

We have then designed a real 3-D array with 120 electret microphones, randomly positioned within a cube of size 2m. The room has dimensions (3.8, 8.15, 3.6)m, it was emptied but still had features that made it non-ideal: a doorway, two windows, a cornice, concrete walls, wood panels, etc. The source is a baffled loudspeaker placed far from the array, and the RIRs have been measured using sine sweeps<sup>21</sup> in the bandwidth [50, 1000]Hz. The sine sweeps were long enough in order to reduce the noise of measurements. In order to isolate the modes below a cutoff frequency  $f_c$ , we have used a low-pass filter, and a downsampling at  $F_s > 2f_c$ . The microphones are placed at random positions within  $\Omega$ , with a statistical distribution close to uniform - up to mechanical constraints. The set of microphone positions has been calibrated using a optimization procedure<sup>22</sup>, with the measured positions as initial estimates.

As mentioned earlier, when the cutoff frequency  $f_c$  increases, the modal density strongly increases, and the sparsity assumption becomes less and less valid. Indeed, the number of theoretical modes has been calculated<sup>12</sup>, and table 2 shows that it increases faster than  $f_c$ . However, the experimental reconstruction SNR remains surprisingly stable with respect to  $f_c$ . Note that the analysis for more than 363 modes requires too much memory for our computer, and therefore we could not test for  $f_c > 350\text{Hz}$ .

| $f_c$ [Hz]           | 250  | 275  | 300  | 325  | 350  |
|----------------------|------|------|------|------|------|
| Number of modes      | 120  | 160  | 216  | 280  | 363  |
| SNR [dB]: method CS1 | 17.8 | 19.0 | 20.1 | 20.6 | 20.3 |
| SNR [dB]: method CS2 | 15.9 | 16.0 | 16.3 | 16.1 | 15.8 |

Table 2. Results for different cutoff frequencies.

In figure 1, two interpolated RIRs are displayed (for the two methods CS1 and CS2). One microphone of the array has been isolated for the interpolation, and the analysis has been done using the 119 others. Here  $f_c = 300\text{Hz}$  and  $F_s = 750\text{Hz}$ . Results shown that the interpolations are similar to the measured RIRs, both from the SNR and correlation performance measures.

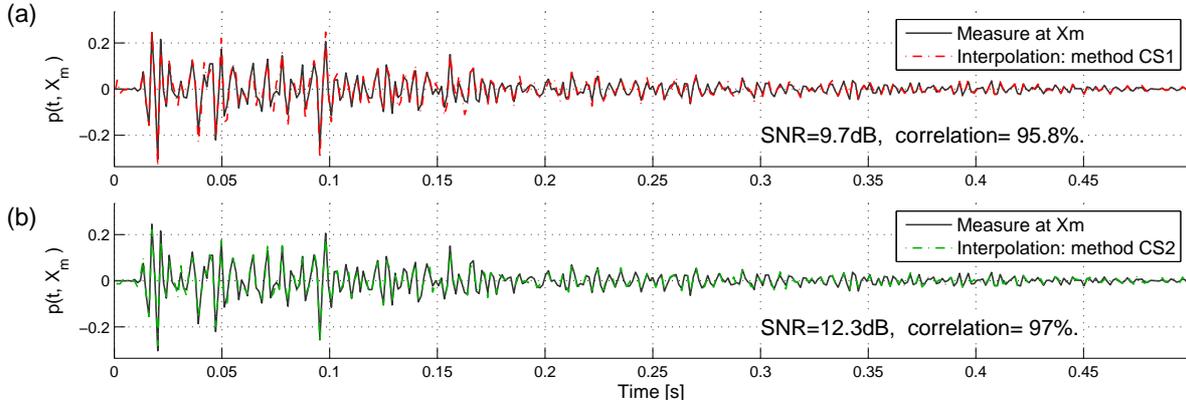


Figure 1. Measured and interpolated RIRs. (a): Method CS1, (b): Method CS2.

Figure 2a shows the performance of the interpolation according to the distance between the interpolation position and the center of the array. Here, each measured RIR has been interpolated using the model parameters from the analysis of the remaining 119 measured signals. We used  $f_c = 300\text{Hz}$ , and  $F_s = 750\text{Hz}$ . The microphones are grouped according to their distance from the center of the array. In each group corresponding to a distance range, we estimate the average reconstruction error and the corresponding standard deviation. As expected, performance decreases when the interpolation position moves away from the center, although it can be noticed that with method CS2 they decrease slower than with method CS1.

Figure 2b shows the performance of the interpolation when synthetic noise  $\epsilon_n$  is added to the measurement signals. It can be observed that, as expected, performances decrease when the noise level increases. At high noise levels, method CS2 appears more robust than method CS1: the least-square projection of method CS1 tries to fit the whole (noisy) signal with the model, while the “sparse” method CS2 is intrinsically a denoising framework.

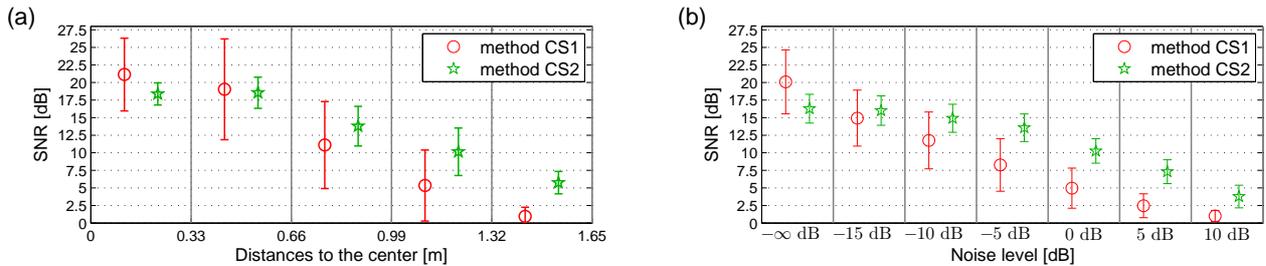


Figure 2. (a): Evaluation according to the distance from the center of the array. (b): Evaluation according to the level of the additional noise. The x-axis is, on a dB scale, the energy of the additional noise:  $\|\epsilon\|_{\ell_2}/\|s\|_{\ell_2}$  over the energy of the measured signals. These statistical representations show the average of the SNRs and the standard deviation.

We have checked the robustness of both methods to the geometry of the room, in particular when the measured room gets further away from the “ideal” rectangular room. This has been made by opening the windows and the door, and by placing a chair and a wood panel. Experimental results for RIR interpolation show that the performance of both methods was not significantly affected by this change of geometry.

In figure 3, the performances are evaluated according to the number of microphones for the analysis. Here, we have randomly selected 16 microphones close to the center of the array (distance smaller than 80cm). As a general trend, performance decreases with  $M$ . It is however interesting to notice that with less than 40 microphones, method CS2 outperforms method CS1, and with only 32 microphones the SNRs average is approximately 10dB.

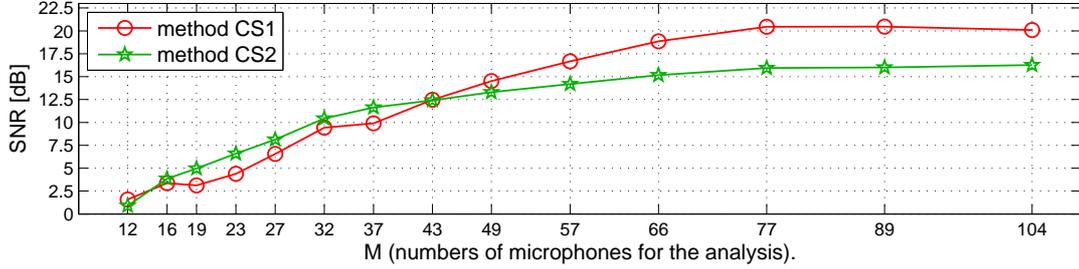


Figure 3. Evaluation according to the number of microphones of the array.

## 6. CONCLUSION

Using the modal theory, it is possible to use Compressed Sensing in order to sample and to reconstruct the Plenacoustic Function at low frequencies, with a number of sensors significantly lower than would be required by Shannon-Nyquist sampling theorem. More precisely, we have shown that our method results in a tradeoff between the number of microphones and the precision of the interpolation. Moreover, our method allows for a flexible array configuration : we experimented here with microphones randomly placed in the space volume of interest  $\Omega$ , with a spatial distribution close to uniform probability. Other antenna geometries could be tested, possibly resulting in better performance.

The reduction in the number of measurements / microphones allowed by Compressed Sensing can be important in practical applications. However, it comes with a computational cost that can rapidly become prohibitive. The two algorithms presented in this paper have been tuned so that they still can run in reasonable time on a PC: the MATLAB analyses of section 5 spent almost one hour with a 6 core CPU at 3GHz and a 24Gb of RAM.

As shown in section 5, the first algorithm (method CS1) gives excellent results in favorable cases, whereas the second one (method CS2) seems more robust to noise and the distance from the center of the array.

Here, the PAF reconstruction is limited to the lower frequencies of the spectrum, but on a wide time interval. A complementary approach can similarly interpolate the early part of the RIRs over a wide frequency range with a reduced set of measurements, using the sparsity of the early reflexions (cf. Ref. 23). By fusing both approaches, we hope to extrapolate the full RIRs in time and frequency.

## ACKNOWLEDGMENTS

The authors want to thank François Ollivier, Dominique Busquet and Christian Ollivon, of the Jean Le Rond d'Alembert Institute, for their precious help in the realisation of the microphone array and the acquisition of the RIRs.

## REFERENCES

- [1] Ajdler, T., Sbaiz, L., and Vetterli, M., "The plenacoustic function and its sampling," *IEEE Transaction on Signal Processing* **54**, p. 3790–3804 (October 2006).
- [2] Masterson, C., Kearney, G., and Boland, F., "Acoustic impulse response interpolation for multichannel systems using dynamic time warping," in *[AES 35th Conferences]*, (February 2009).
- [3] Haneda, Y., Kaneda, Y., and Kitawaki, N., "Common-acoustical-pole and residue model and its application to spatial interpolation and extrapolation of a room transfer function," *IEEE Transaction on Speech and Audio Processing* **7**, p. 709–717 (November 1999).
- [4] Perrey-Debain, E., "Plane wave decomposition in the unit disc: Convergence estimates and computational aspects," *J. Comp. App. Math.* **193**, p. 140–156 (August 2006).

- [5] Moiola, A., Hiptmair, R., and Perugia, I., “Approximation by plane waves,” Research Report 2009-27, Swiss Federal Institute of Technology Zurich (ETH Zürich) (2009).
- [6] Chardon, G., Leblanc, A., and Daudet, L., “Plate impulse response spatial interpolation with sub-nyquist sampling,” Submitted for publication.
- [7] Schmidt, R., “Multiple emitter location and signal parameter estimation,” *IEEE Transaction on Antennas and Propagation* **34**, p. 276–280 (March 1986).
- [8] Roy, R. and Kailath, T., “Estimation of signal parameters via rotational invariance techniques,” *IEEE Transaction on Acous., Speech, and Signal Processing* **37**, p. 984–995 (July 1989).
- [9] Chardon, G. and Daudet, L., “Optimal subsampling of multichannel damped sinusoids,” in [*Proceedings of the 6th IEEE Sensor Array and Multichannel Signal Processing workshop (SAM 2010)*], 4 pages (October 2010).
- [10] Zotter, F., “Sampling strategies for acoustic holography/holophony on the sphere,” in [*NAG-DAGA*], 4 pages (March 2009).
- [11] Leopardi, P., “A partition of the unit sphere into regions of equal area and small diameter,” *Electronic Transactions on Numerical Analysis* **25** (2006).
- [12] Kuttruff, H., [*Room Acoustics*], Spon press, 4 ed. (October 2000).
- [13] Candès, E. and Wakin, M., “An introduction to compressive sampling,” *IEEE Signal Processing Magazine* **25**(2), p. 21–30 (2008).
- [14] R.G.Baraniuk, “Compressive sensing,” *IEEE Signal Processing Magazine* **24**(4), p. 118–121 (2007).
- [15] Candès, E. and Tao, T., “Decoding by linear programming,” *IEEE Transaction Information Theory* **51**(12), p. 4206–4215 (2005).
- [16] Candès, E., “Compressive sampling,” in [*International Congress of Mathematicians*], (August 2006).
- [17] Candès, E., Romberg, J., and Tao, T., “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Transaction on Inform. Theory* **52** (February 2006).
- [18] Elad, M., [*Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*], Springer, 1 ed. (2010). 376 pages.
- [19] Allen, J. and Berkley, D., “Image method for efficiently simulating small-room acoustics,” *J. Acoust. Soc. Am.* **65**, p. 943–950 (April 1976).
- [20] Peterson, P., “Simulating the response of multiple microphones to a single acoustic source in a reverberant room,” *J. Acoust. Soc. Am.* **80**, p. 1527–1529 (November 1986).
- [21] Müller, S. and Massarani, P., “Transfer-function measurement with sweeps,” *J. Audio Eng. Soc* **49**(6), p. 443–471 (2001).
- [22] Ono, N., Kohno, H., Ito, N., and Sagayama, S., “Blind alignment of asynchronously recorded signals for distributed microphone array,” in [*IEEE WASPAA*], p. 161–164 (2009).
- [23] Mignot, R., Daudet, L., and Ollivier, F., “Compressed sensing for acoustic response reconstruction: interpolation of the early part,,” in [*IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA’11)*], (October 2011).