MICROPHONE ARRAY POSITION CALIBRATION IN THE FREQUENCY DOMAIN USING A SINGLE UNKNOWN SOURCE

Thibault Nowakowski, Laurent Daudet, Julien de Rosny

Institut Langevin, ESPCI CNRS UMR 7587, Université Paris Diderot,
1 rue Jussieu, 75005 Paris, FRANCE

ABSTRACT

We study the problem of microphone array localization in a strongly reverberant room, where time of arrivals (TOA) or time difference of arrivals (TDOA) cannot always be measured precisely. Instead, we use frequency-domain measurements to calibrate the array position, based on the modes of the room, excited by a wide-band single source, that can be unknown. By using the fact that each measured mode can be decomposed as a sum of model-based polynomials, we build a cost function whose minimum indicates the positions of the microphones. A simple Block Coordinate Descent algorithm can be used to minimize this cost function. Numerical results indicate that this algorithm converges to the right solution, and therefore that using frequency measurements for position calibration is a valid concept for dense arrays, as an alternative to time-domain methods in reverberant domains.

***Index Terms*—** Array position calibration, modal interpolation, reverberation

1. INTRODUCTION

A general trend in microphone arrays is to massively increase the number of microphones, especially with the increase of computational power and the availability of cheap digital MEMS microphones [1]. For most signal processing applications, such as acoustic imaging, it is essential to determine the geometry of the array, i.e. the position of each microphone. However, a direct measurement of the absolute position of each microphone is sometimes difficult to achieve, or extremely cumbersome - this is typically the case for very large arrays, ad-hoc arrays, or random arrays that have become more popular with the paradigm of Compressed sensing [2]. Therefore, in the past few years, a number of methods have been developed, whose goal is to determine the relative position of each microphone using purely acoustical measurements, with a number of point-like calibration sources emitting around the array. Most of these approaches are based on time-of-flight measurements between sources and microphones, so-called Time Of Arrival (TOA) or Time Difference Of Arrival (TDOA). The array geometry is then determined with multidimensional scaling (MDS) [3]. Often, the challenge of such methods is to find tractable closed-form solutions to the calibration problem [4, 5, 6, 7, 8, 9].

Although the formulation of these time-domain methods are more and more efficient in simple environments, they are difficult to exploit in complex domains, such as rooms, especially when reverberation, noise, limited bandwidth of the emitted signals, sampling precision, or heterogeneities in the environment make the measurement of TOA / TDOA difficult. Furthermore, TOA / TDOA models assume point-like sources and microphones. While this assumption is reasonable given the minimum wavelength $\lambda = .017$ m (i.e. $f = 20000$ Hz) in the audible bandwidth and the typical size of measurement microphones (1/2" or 1/4"), it is often not the case for most wide-band sources : for many loudspeakers, the position of the centre of phase (equivalent point source) is frequency-dependent. Finally, the computation of the positions from TOA / TDOA might involve strong nonlinearities, and therefore small errors on the TOA / TDOA estimations can translate into large errors on the estimated array geometry.

Alternatively, a few models based on measurements in the frequency domain have recently emerged. They are based on a model of diffuse acoustic fields, which can be seen as an infinite superposition of plane waves coming from all directions, or explained by a uniform density of energy inside the propagation domain [10]. The intercorrelations of these fields measured at pairs of microphones provide estimates of covariance matrices. Each element $(i\ ;\ j)$ of the covariance matrix can be analytically modeled [11, 12, 13] as a *sinc* function, whose argument is the product of the wavenumber k and the distance r_{ij} between microphones i and j. Fitting the analytic covariance matrix on the measurements, it is possible to retrieve all pairwise distances between microphones, and then again to apply MDS techniques in order to deduce the array geometry. Nevertheless, diffuse field calibration techniques have their own limitations. *Prima facie*, it is difficult to establish a real isotropic diffuse field into a given environment. Furthermore, heterogeneities and boundary conditions of the environment modify the covariance, that cannot always be modeled as a *sinc* function. Finally, this covariance model can only be used in practice for close microphones, as measurement noise renders all estimations unreliable for distant microphones.

In this article, we investigate a different paradigm for array calibration in reverberant rooms, with frequency-domain measurements. Unlike TOA-based methods which require the measurement of several sources, we only use a single source to emit a wide-band signal that excites the modes of the room. Based on harmonic polynomial expansions of solutions to the Helmholtz equation, it is possible to spatially interpolate the measured modes. As these harmonic polynomials depend on the microphone positions, a cost function can be built that compares the modeled values with their measured counterparts. Minimizing this cost function provides the positions of the microphones.

To summarize our work, we make the following contributions :

- a new method for array position calibration using only one source, based on the measurements of the eigenfrequencies (modes) of a room, the interpolation of the homogeneous Helmholtz equation and the minimization of a cost function ;

This work is supported by DGA (Direction Générale de l'Armement) and by LABEX WIFI (Laboratory of Excellence within the French Program "Investments for the Future") under references ANR-10-LABX-24 and ANR-10-IDEX-0001-02 PSL*.

- an efficient iterative algorithm, in order to reach the global minimum of this non-convex function ;
- a discussion on the limitations of this calibration technique ;
- some numerical simulations in the case of two-dimensional domains, that show the relevance of the method.

2. CALIBRATION MODEL

Let $\mathcal{D}$ be a closed domain with unknown boundary conditions (B.C.) (Neumann, Dirichlet, etc.), modeling a reverberant room. An acoustic field is created by a wide-band source S (impulse, white noise, sweep sine), and is sampled by a set of microphones at positions $\mathbf{x_n}$, distributed within a sub-domain $\Omega \in \mathcal{D}$. The measurement vector is noted $\mathbf{p}$, expressed in the Fourier domain. Depending on the dimensions of the room, the emitted signal excites some specific frequencies, that are called the eigenfrequencies of the room, and that correspond to the modes of the domain. The modes of $\mathcal{D}$ are natural solutions to the homogeneous Helmholtz equation measured in the set Ω:

$$\begin{cases} \Delta p_0 + k^2 p_0 = 0 & \text{in} \quad \Omega \\ \text{+ unknown B.C.} & \text{on} \quad \partial\Omega \end{cases} \tag{1}$$

From the set of point measurements, a given mode can be interpolated within the convex hull of the set of measurement points thanks to the Vekua theory [14, 15], that demonstrates that harmonic polynomials can be used to approximate solutions of the homogeneous Helmholtz equation. In two-dimensional cases, these polynomials are Fourier-Bessel (F-B) functions of order l whose expression depends on the polar coordinates $(r_n\ ;\ \theta_n)$ of a microphone n, and on the wavenumber k of the mode (they have to be replaced by spherical harmonics in three-dimensional cases [16]). Special care has to be taken in the choice of the origin for the coordinate system, as F-B functions must be centered in the measurement set. Taking the center of mass of the measurement domain Ω as origin minimizes the number of F-B functions needed to describe the field.

In [17], F-B functions were used to interpolate the modes of a plate inside the convex hull of the measurement points, delineating the domain Ω (this also applies in the three-dimensional case). The authors used the fact that the pressure field measured at frequency f and wavenumber k can be decomposed as :

$$\mathbf{p} = \mathbf{W}\boldsymbol{\alpha} \tag{2}$$

where $\mathbf{W}$ is a dictionary of F-B functions of order $-L \leq l \leq L$, L being the maximal order of the decomposition, and $\boldsymbol{\alpha}$ is the vector of coefficients of the projection.

Generally, given a number N of microphones, a mode of wavenumber k and a measurement set Ω whose convex hull has a radius R, the minimal order L of F-B functions needed to interpolate the mode, in a 2D case, is approximately equal to :

$$L \simeq \lceil kR \rceil \tag{3}$$

where $\lceil \cdot \rceil$ denotes the ceiling integer rounding function.

For two-dimensional domains (plates, membranes), an atom of the $\mathbf{W}$ basis, for microphone n and at order l then writes :

$$W_{nl} = J_l(kr_n)e^{j \cdot l\theta_n} \tag{4}$$

where $J_l(z)$ is the Bessel function of first kind.

Here, it is important to stress that the basis $\mathbf{W}$ is parametrized by the positions of the microphones. When these positions are known, the representation of the field as a linear combination of the atoms of $\mathbf{W}$ is consistent with the measurements, up to the modeling error (i.e., the finite order of F-B functions). On the contrary, if the sensing positions are unknown or inaccurate, there is no guarantee that the decomposition will be possible, as the F-B functions for the basis might be built upon inaccurate positions. In that case, the model will not be able to explain the measured field at the sensing positions. We use this fact to build our calibration model.

Let $\mathbf{W}_\varepsilon$ be the Fourier-Bessel dictionary built with incorrect positions $\mathbf{x}_\varepsilon$. It is still possible to decompose the measured field $\mathbf{p}$ on $\mathbf{W}_\varepsilon$, by inverting equation (2) to calculate the $\boldsymbol{\alpha_\varepsilon}$ coefficients :

$$\boldsymbol{\alpha_\varepsilon} = \mathbf{W}_\varepsilon{}^\dagger \mathbf{p} \tag{5}$$

with $.^\dagger$ denoting the Moore-Penrose pseudo-inverse. Now, by observing that the dictionary is not built upon the actual sampling positions, the reconstructed field $\mathbf{p}_\varepsilon$

$$\mathbf{p}_\varepsilon = \mathbf{W}_\varepsilon \boldsymbol{\alpha_\varepsilon} \tag{6}$$

is such that $\mathbf{p}_\varepsilon \neq \mathbf{p}$.

Then we address the following optimization problem "*find the positions* $\mathbf{x}$ *of microphones such that* $\mathbf{p}_\varepsilon = \mathbf{p}$", which can be rewritten, using joint penalization over eigenfrequencies, as :

$$\mathbf{x} = \min_{\mathbf{x_\varepsilon}} \sum_k \|\mathbf{p}_\varepsilon - \mathbf{p}\|^2 = \min_{\mathbf{x_\varepsilon}} \sum_k \|(\mathbf{W}_\varepsilon \mathbf{W}_\varepsilon{}^\dagger - \mathbb{1})\mathbf{p}\|^2 \tag{7}$$

As visible on equation (7), this least-squares optimization is equivalent to minimizing the projection of the measurements $\mathbf{p}$ onto the kernel of the orthogonal space defined by span$(\mathbf{W}_\varepsilon)$.

3. RESOLUTION METHOD

This cost function is highly non-convex, with a lot of local minima. Furthermore, the number of variables, i.e. the number of positions of all microphones, can be large. In order to reach the global minimum, we have used a simple algorithm based on block-coordinate descent, optimizing the position of one microphone, in all directions, at a time. A discrete grid of candidate positions is built, at the smallest spatial resolution of the problem, with one microphone set as the origin. The microphones positions are first initialized at random onto the grid. Then, all positions are blocked except one that is replaced successively by all the test positions. The position that corresponds to the minimum of the cost function in this actual array configuration replaces the initial position. The center of mass of the array is updated and the F-B basis is recentered accordingly. Then the next microphone position is optimized, etc.

To illustrate the non-convexity of the cost function, we give here a simple example with a two-dimensional rectangular array (square Shannon sampling), with only one misplaced microphone. As the modes of a room are of infinite expansion and draw a regular pattern, several positions exist where the microphone will measure the same field for the mode of wavenumber k. However, only one position is a joint minimizer for all the modes. This demonstrates the necessity of using multi-modal measurements (at different frequencies), as seen on figure 1, where the use of several modes helps the convergence to the true global minimum. In practical cases, when more than one position is incorrect, this issue gets all the more important.

A further refinement is used to ease the optimization, based on the fact that two local minima are roughly distant from a wavelength $\lambda = \frac{2\pi}{k}$. At low frequencies, λ is large (for example in air, for $f = 340$ Hz the wavelength is $\lambda = 1$ m). Then, the minimum of the cost function should be more spatially extended. In return, it will be

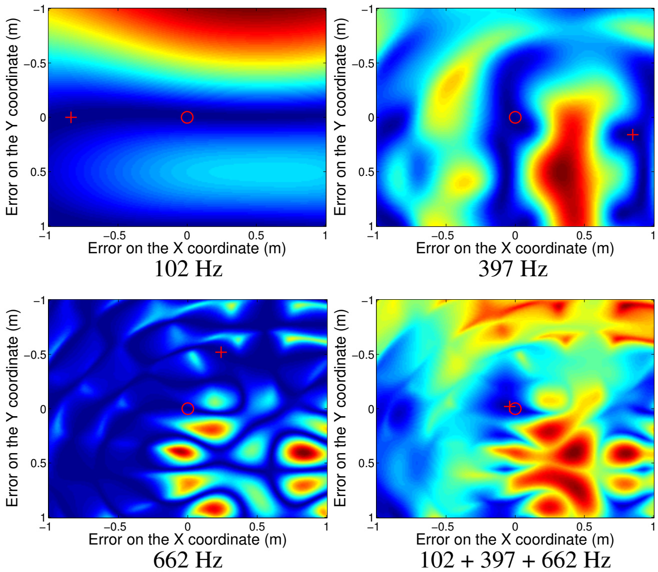

Fig. 1. Cost functions for 3 frequencies for a rectangular array (square sampling), when moving only one microphone around its true position, and sum (bottom right plot) of the 3 cost functions. Circle : true global minimum. Cross : position of the minimum at convergence.

more difficult to have a good accuracy on the estimated positions. On the contrary, high frequencies give a better precision on the estimated positions but have cost functions with a lot of local minima, as the wavelength gets smaller. By starting the optimization process using only modes at low frequencies, it is possible to have a first coarse estimation of the microphones positions. Then higher modes are added progressively (once the array configuration is trapped into a minimum for a given number of modes), for better accuracy. Thus, several local minima are avoided and the iterative method is more likely to converge to the global minimum.

4. LIMITATIONS OF THE MODEL

The fact that this method is based on the spatial interpolation of the modes implies some constraints on the array. For example, the spatial distribution of microphones impinges on the interpolation error. It was shown in [18, 19] that the optimal way to sample solutions of the Helmholtz equation in a domain Ω of radius R is to place most of the measurements on the border $\partial\Omega$ of the domain of interest, so that the distribution satisfies on average the spatial Shannon-Nyquist criterion for the mode of maximal frequency that will be interpolated. Some extra measurements points have to be added in the interior of Ω, so that the interpolation in the least-squares sense is regularized on the eigenfrequencies of the array. This implies that the method will better work with arrays that are "well-sampled" on the domain border (but not necessarily uniformly distributed), or with dense random arrays with enough microphones to compensate the fact that $\partial\Omega$ is not optimally sampled. As with all the other methods for position self-calibration, the domain Ω also needs to be homogeneous, i.e. with no obstacle nor velocity change inside the array.

It follows from these constraints and from (3) that the total number of functions $\mathcal{L}$ used to build $\mathbf{W}$ is :

$$\mathcal{L} = 2L + 1 \quad (8)$$

As the maximal value of $\mathcal{L}$ to guarantee the stability of this least-squares approximation is $\mathcal{L} = N$, and it follows from (3) and (8) that the highest frequency for interpolation is approximately :

$$f_{max} \simeq \frac{(N-1) \cdot c_0}{4\pi R} \quad (9)$$

with c_0 the sound velocity in the medium.

5. NUMERICAL EXPERIMENTS

To confirm the validity of the method, different numerical scenarios of antenna are investigated in two dimensions, simulating the calibration experiment on plates or membranes. Using the FDTD method [20], we simulate the propagation of a wide-band source (20 Hz - 3000 Hz), inside a rectangular domain of dimensions 6 m $\times$ 5 m. The sampling frequency is set to $f_s = 28590$ Hz so that the spatial grid is sufficiently fine. The emitted signal is recorded on 65 points of the grid representing the locations of the 65 microphones, distributed in a small sub-domain of the room.

After taking the Fourier transform of the recorded signals, we select manually the increasing frequencies of the modes visible in the power spectrum (this task could easily be automated), starting from the lowest frequency, here approximately 30 Hz, to 1000 Hz.

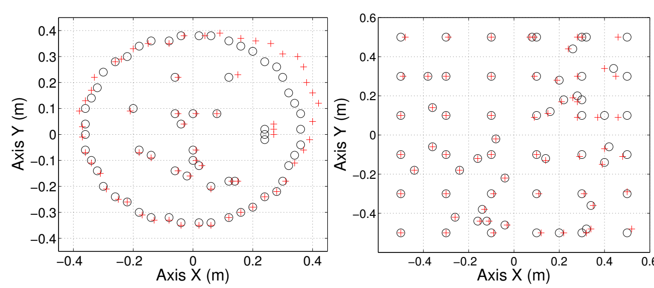

Fig. 2. (Left) Array 1 : calibration state at iteration 60. (Right) Array 2 : calibration state at iteration 120. Black circles : real positions. Red crosses : current estimated positions. At convergence, markers are superimposed.

We compare two array geometries, both with 65 microphones. The first array is quasi-circular (circle quantized on grid positions), of radius $r = .35$ m. It approximates the "ideal" sampling scheme described above. In the second array, 29 microphones are randomly placed with uniform distribution, within a square regular array of $6 \times 6 = 36$ microphones and size 1 m. The grid of test positions is of dimensions 1.2 m $\times$ 1.2 m with a precision of .01 m. The radius parameter used to set the order L of the F-B expansion is chosen as $R = .8$ m, slightly larger than the spatial grid.

For the sake of simplicity, we fix the array coordinate system by assuming that the positions of 3 microphones are known. It should be acknowledged that this gives more constraints than needed as, minimally, one only has to fix the position of one microphone (avoiding translation indeterminacies), the direction of a second microphone (avoiding rotations) and the sign of the scalar product of a third microphone onto the line defined by the first 2 microphones (avoiding a mirror symmetry).

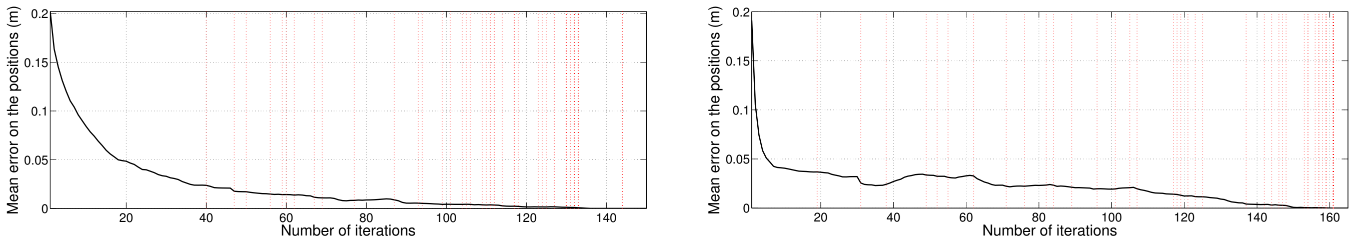

Fig. 3. Mean position error as a function of the number of iterations, (left) array 1, (right) array 2. Each vertical red dashed line represents the addition of a new modal frequency into the cost function.

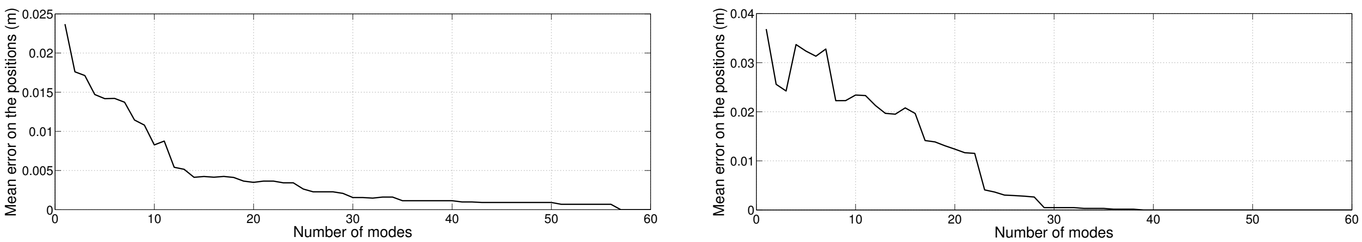

Fig. 4. Mean position error as a function of the number of modes, (left) array 1, (right) array 2.

Then, the positions of all microphones are estimated with the optimization algorithm. In figure 2, the real and estimated positions of the two arrays are represented during the optimization phase, before convergence (at convergence, exact positions are retrieved). Figure 3 represents the mean error of microphones positions as a function of the number of iterations. Each red dashed line represents the addition of a new frequency into the cost function. As expected, the error decreases quickly for the first few low-frequency modes but needs more and more frequencies to reach the global minimum.

The mean error of calibration is also represented as a function of the number of modes, on figure 4. For the quasi-circular array, only 57 modes were needed to minimize the cost function, the highest mode having a frequency $f = 328$ Hz. This value is not very high, compared to the highest frequency that can be interpolated with this array. Indeed, considering equation (9), $f_{max} \simeq 4950$ Hz. For the second array, 39 modes were needed, the highest frequency being $f = 262$ Hz, whereas $f_{max} \simeq 3460$ Hz.

Here the reference known microphone was chosen at random but these simulations confirm that, in a general case, it is better to choose this reference microphone near the center of gravity of the array. This helps the regularization of the modes interpolation, especially with non-dense arrays. Figure 5 displays the error of calibration for each position of array 2, averaged over all the iterations of the optimization (before convergence), as a function of the distance to the reference microphone. For this application, the reference microphone was chosen at one corner of the array. It can be seen that the mean error increases with the distance to the origin : the furthest microphones are the last to be correctly located. Choosing the reference microphone around the centre of the array would ease the optimization, especially for non-dense arrays.

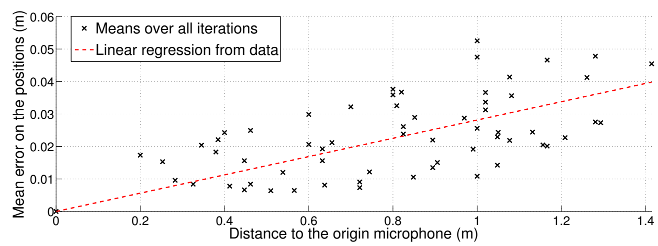

Fig. 5. Array 2 : mean error of calibration over all iterations of the optimization step (before convergence) in function of the distance between each microphone and the reference microphone (N.B. : at the end of the optimization step, estimated positions converged to the true positions).

6. CONCLUSION

In this work, we have used modal interpolation techniques to propose a new method for the calibration of the geometry of a microphone array placed in a reverberant room. It is based on frequency measurements, instead of the TOA / TDOA used in most standard methods - this can be useful when TOA / TDOA cannot be precisely measured. Using only one wide-band source, that does not need to be known nor point-like, and the decomposition of the measured modes of a room onto a basis of Fourier-Bessel functions, we have built a cost function that is minimized at the position of the microphones.

However, the optimization of this cost function is non-trivial, as it is highly non-convex. Here, a simple Block Coordinate Descent iterative method was used, and was made more robust by progressively increasing the bandwidth. Numerical tests in 2D confirm the relevance of this approach, which must now be generalized to 3D, where the same theoretical arguments hold, by replacing the F-B functions by spherical harmonics. Further effort should also be spent on more efficient optimization methods, both in terms of computation time and probability of success. Finally, the robustness to noise must be thoroughly investigated. However, as our algorithm relies on least-squares optimization, it should handle well gaussian noise.

Another extension of this work will be the study of cases where the room geometry is known. In that case, the relative position of the microphones could be identified with much less measurements, alleviating the requirement for dense arrays. It is important to stress that this would also give the absolute position of the array in the room, which can be very useful in practical scenario (e.g. for source localization) and is not addressed by most TOA / TDOA - based methods, with the notable exception of [8].

7. REFERENCES

[1] I. Hafizovic, C.I.C Nilsen, M. Kjølerbakken, V. Jahr, "Design and implementation of a MEMS microphone array system for real-time speech acquisition", *Applied Acoustics*, Vol. 73, No 2, pp. 132–143, 2012.

[2] E.J. Candès, M.B. Wakin, "An introduction To Compressive Sampling", *IEEE Signal Processing Magazine*, Vol. 25, No 2, pp. 21–30, 2008.

[3] W.S. Torgerson, "Multidimensional Scaling : I. Theory and Method", *Psychometrika*, Vol. 17, No 4, pp. 401–419, 1952.

[4] R.C. Raykar, R. Duraiswami, "Automatic Position Calibration of Multiple Microphones", *Proc. ICASSP*, 2004.

[5] M. Crocco, A. Del Blue, M. Bustreo and V. Murino, "A Closed Form Solution to the Microphone Position Self-Calibration Problem", *Proc. ICASSP*, 2012.

[6] S.T. Birchfield, A. Subramanya, "Microphone Array Position Calibration by Basis-Point Classical Multidimensional Scaling", *IEEE Transactions on Speech and Audio Processing*, Vol. 13, No. 5, pp. 1025–1034, 2005.

[7] T-K. Le, N. Ono, "Numerical Formulae TOA-Based Microphone and Source Localization", *Proc. IWAENC*, 2014.

[8] I. Dokmanic, L. Daudet, M. Vetterli, "How to Localize Ten Microphones in One Fingersnap", *Proc. EUSPICO*, 2014.

[9] P. Pertilä, M. Mieskolainen, M.S. Hämäläinen, "Passive Self-Localization of Microphones Using Ambient Sounds", *Proc. EUSIPCO*, 2012.

[10] T. J. Schultz, "Diffusion in reverberation rooms", *J. Sound Vib.*, Vol. 16, No 1, pp. 17–28, 1971.

[11] I. McCowan, M. Lincoln, I. Himawan, "Microphone Array Shape Calibration in Diffuse Noise Fields", *IEEE Transactions on Audio Speech and Signal Processing*, Vol. 16, No 3, pp. 666–670, 2008.

[12] M.J. Taghizadeh, R. Parhizkar, P.N. Garner, H. Bourlard, "Euclidean matrix completion for ad-hoc microphone array calibration", *Proc. DSP*, 2013.

[13] M.J. Taghizadeh, P.N. Garner, H. Bourlard, "Enhanced diffuse field model for ad-hoc microphone array calibration", *Signal Processing*, Vol. 101, pp. 242–255, 2014.

[14] I.N. Vekua "New methods for solving elliptic equations", *North-Holland*, 1967.

[15] A. Moiola, R. Hiptmair, I. Perugia. "Plane Wave Approximation of Homogeneous Helmholtz Solutions", *Z. Angew. Math. Phys.*, No 62, Issue 5, pp. 809-837, 2011.

[16] E.G. Williams, "Fourier Acoustics, Sound Radiation and Nearfield Acoustical Holography", *Academic Press*, 1999.

[17] G. Chardon, A. Leblanc, L. Daudet, "Plate impulse response spatial interpolation with sub-Nyquist sampling", *Journal of Sound and Vibration*, Vol. 330, No 23, pp. 5678–5689, 2011.

[18] G. Chardon, A. Cohen, L. Daudet, "Sampling and reconstruction of solutions to the Helmholtz equation", *accepted in Sampling Theory in Signal and Image Processing*, 2014.

[19] G. Chardon, W. Kreuzer, M. Noisternig, "Design of a robust open spherical microphone array", *Proc. ICASSP*, 2014.

[20] A. Taflove, "Computational Electrodynamics: The Finite-Difference Time-Domain Method, 3^{rd} ed.", *Artech House Publishers*, 2005.