Unsupervised geometry calibration of acoustic sensor networks using source correspondences

Joerg Schmalenstroeer[1], *Florian Jacob*[1], *Reinhold Haeb-Umbach*[1],
Marius H. Hennecke[2], *Gernot A. Fink*[2]

Department of Communications Engineering, University of Paderborn, Germany[1]
Department of Computer Science, TU Dortmund University, Germany[2]
{schmalen, jacob, haeb}@nt.uni-paderborn.de[1] {gernot.fink, marius.hennecke}@tu-dortmund.de[2]

Abstract

In this paper we propose a procedure for estimating the geometric configuration of an arbitrary acoustic sensor placement. It determines the position and the orientation of microphone arrays in 2D while locating a source by direction-of-arrival (DoA) estimation. Neither artificial calibration signals nor unnatural user activity are required. The problem of scale indeterminacy inherent to DoA-only observations is solved by adding time difference of arrival (TDoA) measurements. The geometry calibration method is numerically stable and delivers precise results in moderately reverberated rooms. Simulation results are confirmed by laboratory experiments.

Index Terms: unsupervised, sensor network, geometry calibration

1. Introduction

Spatially distributed microphone arrays are not only employed for acoustic beamforming but also for acoustic source localization. The latter can be used for advanced teleconferencing systems, ambient communication, or location based services [1], making microphone arrays an important component of smart environments. Acoustic source localization, however, requires that the position and orientation of the sensor arrays is known. Their determination, termed geometry calibration, is often a tedious manual task which is contradictory to the desire to have a quick and effortless system setup.

Some approaches have been proposed to conduct automatic geometry calibration for acoustic and/or visual sensors. They may be grouped according to the kind of observations available. The first category comprises methods that assume that all coordinates of the source signal's position are measurable. From these data, which are typically given in Cartesian coordinates, the network geometry can be inferred [2]. The calibration itself can, for example, be based on a singular value decomposition [3].

However many sensors cannot deliver the full coordinate information. For example, linear microphone arrays in the far field of the source can only determine the direction-of-arrival, but not the range. The second category therefore consists of methods that assume that only angle information is available. Here, a system of nonlinear equations is set up and solved iteratively for the relative source and sensor locations [4]. Only if at least one distance is known a priori the absolute positions can be determined. This, however, contradicts the goal of a fully unsupervised automatic calibration.

Our approach falls into this second category, but we will present a method to circumvent the need for an a priori known distance. In [4] it has been observed that the solution of the system of nonlinear equations is numerically very sensitive and that depending on the initial conditions and the geometric configuration the iterations do not converge. In this paper we present a reformulation which greatly improves numerical stability.

Some approaches use special calibration signals to identify a unique source for all sensors and to ease the source localization [5]. Our goal, however, was to employ the user's speech as calibration signal. The user is not required anything else than speaking and walking around.

The paper is organized as follows: In Section 2 the 2-dimensional calibration problem is introduced and the solution presented in [4] is described. Section 3 presents our reformulation and Section 4 illustrates how the missing distance information can be obtained by TDoA measurements. The calibration algorithm consisting of the iterative Newton algorithm embedded in a random sample consensus method for outlier rejection is shown in Section 5, which is followed by experimental results in Section 6 and conclusions drawn in Section 7.

2. Background

We are interested in geometry calibration in 2D where we assume that the sensors are placed near the walls such that the source is inside a polygon spanned by the sensor locations. Further we assume that the individual shapes of the microphone arrays are known (e.g. by the shape calibration method of [6]), such that each array can deliver an estimate of the direction-of-arrival of a desired source signal.

The i-th observation of a source at the unknown position $\boldsymbol{P}_i = [x_i^P, y_i^P]^T$ results in an observed angle ϕ_{ij} at the j-th sensor, which has the unknown position (x_j^S, y_j^S) and the unknown rotation θ_j. Neither are the positions and rotations of the sensors nor are the distances t_{jk} between the sensors known. Figure 1 shows a setup consisting of three sensors and a trajectory with six explicitly marked observations (red dots).

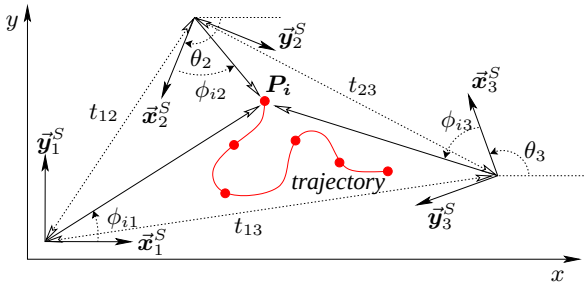

Figure 1: *Geometry calibration problem*

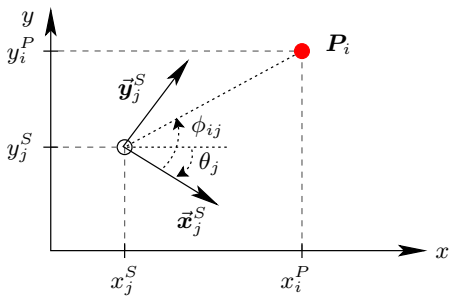

Figure 2: *Geometric relation between observation and sensor*

Figure 2 gives a detailed view of the geometric relations for the i-th observation at the j-th sensor. Following [4], the geometric relation can be formulated by

$$\tan\left(\theta_j + \phi_{ij}\right) = \frac{y_i^P - y_j^S}{x_i^P - x_j^S} \tag{1}$$

which can be reformulated as

$$\begin{aligned} &x_j^S \tan(\phi_{ij}) - y_j^S - (x_i^P + y_i^P \tan(\phi_{ij}))\tan(\theta_j) + y_i^P \\ &+ x_j^S \tan(\theta_j) + y_j^S \tan(\phi_{ij})\tan(\theta_j) - x_i^P \tan(\phi_{ij}) = 0. \end{aligned} \tag{2}$$

Let us assume that the sensor network consists of K sensors, which implies that we have $3(K-1)$ unkowns (coordinates (x_j^S, y_j^S) and rotation θ_j, $j \in [2, K]$). The position and the rotation of one sensor can be arbitrarily fixed, so the amount of unkowns is reduced by three. Each observation introduces two new unknowns (x_i^P, y_i^P) and results in K additional equations (see (2)). Thus at least

$$N \geq \frac{3(K-1)}{K-2} \tag{3}$$

independent observations are required to be able to find a solution.

Let $\mathbf{\Omega} = [x_2^S, y_2^S, \theta_2, \ldots, x_K^S, y_K^S, \theta_K, x_1^P, y_1^P, \ldots, x_N^P, y_N^P]$ be the vector of $3(K-1)+2N$ unknowns and $\boldsymbol{f}(\Omega)$ the system of equations formed by (2) if $j = 2, \ldots, K$ and $i = 1, \ldots, N$. Since a closed form solution cannot be derived, Newton's method is employed to solve it numerically:

$$\mathbf{\Omega}_{\kappa+1} = \mathbf{\Omega}_\kappa - \boldsymbol{J}(\mathbf{\Omega}_\kappa)^{-1} \cdot \boldsymbol{f}(\mathbf{\Omega}_\kappa), \tag{4}$$

with $\mathbf{\Omega}_\kappa$ denoting the estimate of the unknowns at the κ-th iteration and $\boldsymbol{J}(\mathbf{\Omega}_\kappa)^{-1}$ the (pseudo-)inverse of the Jacobian matrix containing the partial derivatives of $\boldsymbol{f}$.

3. Improvements

The approach proposed in [4] (Eq. (1), (2) and (4)) has the following disadvantages. First, Newton methods are sensitive towards the initial values. If the initial values $\mathbf{\Omega}_0$ are far away from the optimal solution the method may not converge or may get stuck in a local minimum. For the problem discussed here we would have to guess reasonable initial values of $3(K-1)+2N$ unknowns, certainly not an easy task. Second, the system of equations may cause numerical problems, since the $\tan$-functions reaches infinity for odd multiples of $\pm\frac{\pi}{2}$. Additionally, the partial derivatives of the $\tan$-functions which are required in Eq. (4) will be extremely large for angles close to $\pm\frac{\pi}{2}$.

Experiments revealed that the aforementioned disadvantages of the Newton method, especially the numerical problems caused by the $\tan$-functions, resulted in severe convergence problems: With random initializations, the iterations did not converge in almost half of all experiments. If artificial noise was added to observations from simulated setups, the convergence problems were aggravated, even for small values of the noise. Another observation we made was that the convergence itself depended on the angles ϕ_{ij}: If the predominant part of the observations lies in the vicinity of $\pm\frac{\pi}{2}$ the method only rarely converged. Adding an arbitrary value before starting the iterations eased the problem.

Stability can be improved by reformulating the system of equations. Multiplying Eq. (2) with $\cos(\phi_{ij})$ and $\cos(\theta_j)$ the following equation is obtained:

$$\begin{aligned} &+x_j^S \sin(\phi_{ij})\cos(\theta_j) - y_j^S \cos(\phi_{ij})\cos(\theta_j) \\ &-x_i^P \cos(\phi_{ij})\sin(\theta_j) - y_i^P \sin(\phi_{ij})\sin(\theta_j) \\ &+x_j^S \cos(\phi_{ij})\sin(\theta_j) + y_j^S \sin(\phi_{ij})\sin(\theta_j) \\ &-x_i^P \sin(\phi_{ij})\cos(\theta_j) + y_i^P \cos(\phi_{ij})\cos(\theta_j) = 0 \end{aligned} \tag{5}$$

The partial derivatives with respect to the unknowns only contain $\sin$- and $\cos$-functions and thus will be numerically more stable as the partial derivatives are now bounded. Simulations confirmed that the Newton method was now much more stable (see Section 6).

Please note that the system of equations still suffers from scale invariance. As is obvious from Eq. (1) all lengths can be multiplied by an arbitrary scaling factor ν without influencing the solution. In order to avoid the trivial solution (all unknowns equal to zero) an equation has to be added which describes a distance relation between two unknowns. In [4] the issue was solved by assuming a priori knowledge of one distance, e.g. the distance between two sensors. However, the need for a known distance can be avoided in the case of acoustic signals and at least two approximately synchronized sensors. Then the scale indeterminacy can be solved by using time difference of arrival (TDoA) information, as will be explained in the next section.

4. DoA and TDoA estimation

The proposed calibration method requires the relative observation angles ϕ_{ij} of the sensors. In the case of acoustic sensors, e.g. linear microphone arrays, the angles may be estimated using an adaptive beamforming approach [7] or a generalized cross-correlation method (e.g. GCC-PHAT) [8]. We prefer beamforming to GCC-PHAT since the required window sizes can be chosen smaller and thus the averaging effect of large windows during fast movements of the speaker is reduced.

The error of the position estimation depends on the distance between the speaker and the closest wall if the microphone arrays are placed at the walls. This known issue is caused by an underestimation of the DoA values, if they approach $\pm\frac{\pi}{2}$. Figure 3 shows the trajectory of an estimated angle for a room of size 4×4 m using a linear microphone array. In order to reduce the influence of this underestimation on the calibration results angles are incorporated only if the absolute values of all angles are smaller than 0.35π.

The estimation of the TDoAs requires larger observation windows compared to the DoA estimation, since the distance between arrays is significantly higher than the distance between sensors of the same array. Here, TDoAs are estimated by GCC-PHAT in combination with a state-space filter, e.g. Kalman or

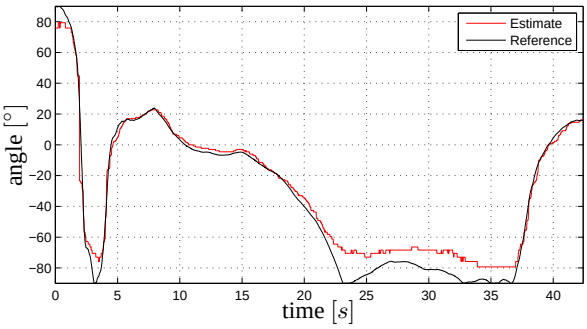

Figure 3: *Underestimation error of angles*

particle filter.

Each combination of sensors j and k delivers a TDoA value $\tau_{jk}(i)$ for the i-th observation which can be used in a distance equation

$$\sqrt{(x_j^S - x_i^P)^2 + (y_j^S - y_i^P)^2} - \sqrt{(x_k^S - x_i^P)^2 + (y_k^S - y_i^P)^2} = c \cdot \tau_{jk}(i) \quad (6)$$

with c being the velocity of sound. Replacing the positions $[x_i^P, y_i^P, x_j^S, y_j^S, x_k^S, y_k^S]$ by $[\nu x_i^P, \nu y_i^P, \nu x_j^S, \nu y_j^S, \nu x_k^S, \nu y_k^S]$ and using the results of the calibration procedure allows the estimation of the unknown scaling factor ν with Eq. (6). To improve the estimate, the scaling factor should be averaged across several spatially distributed observations.

5. Calibration procedure

The first step of the calibration is the DoA estimation using the adaptive beamformers, gathering the set of observations $O' = \{\phi_{ij}\}$ with $1 \leq i \leq N'$ and $1 \leq j \leq K$. From this set all observation pairs violating the condition $|\phi_{ij}| < 0.35\pi$ are removed, resulting in a new set O. At least N observations are required to solve the system of equations (5). If more than N observations are available, a least squares (LS) solution can be obtained. In the experiments we will compare the LS solution using all observations O against a random sample consensus (RANSAC, M rounds) method [9], which is known to be more robust against outliers. In the following we will briefly present the RANSAC method.

In each RANSAC round do:

1. Randomly select a set $\boldsymbol{C} \subset O$ of N observations
2. Generate R random initial values $\boldsymbol{\Omega}^{(r)}$ and determine

$$\boldsymbol{\Omega}_0 = \underset{\boldsymbol{\Omega}^{(r)}}{argmin}\{|\boldsymbol{f}(\boldsymbol{\Omega}^{(r)})|\} \ , r = 1 \ldots R \quad (7)$$

3. Iterate Newton $\boldsymbol{\Omega}_{\kappa+1} = \boldsymbol{\Omega}_\kappa - \boldsymbol{J}(\boldsymbol{\Omega}_\kappa)^{-1} \cdot \boldsymbol{f}(\boldsymbol{\Omega}_\kappa)$ until either
   (i) $|\boldsymbol{\Omega}_{\kappa+1} - \boldsymbol{\Omega}_\kappa| < \Delta$, or
   (ii) Maximum number of iterations reached, or
   (iii) $|\boldsymbol{f}(\boldsymbol{\Omega}_\kappa)|^2 < \epsilon$

   with appropriately chosen values for Δ and ϵ.
4. Extend the consensus set $\boldsymbol{C}$:
   a) Project observations O onto a common coordinate system using $\boldsymbol{\Omega}_\kappa$
   b) Calculate intersection points $\boldsymbol{W}_{jk}(i)$ of the DoA's of the j-th and k-th sensor regarding observation i (see Fig. 4)
   c) Compute average scatter $d(i) = \langle|\boldsymbol{W}_{jk}(i) - \widehat{\boldsymbol{P}}_i|^2\rangle$ of intersection points $\boldsymbol{W}_{jk}(i)$ from their mean $\widehat{\boldsymbol{P}}_i = \langle|\boldsymbol{W}_{jk}(i)|\rangle$, where $\langle\cdot\rangle$ denotes averaging operation
   d) Add all observations with $d(i) < \sigma$ to consensus set, with σ as an appropriately chosen threshold
5. Stop iteration if consensus set $\boldsymbol{C}$ contains more than 70% of the elements of O, else
   a) if size of $\boldsymbol{C}$ has changed $\Rightarrow$ Goto 3.
   b) if size of $\boldsymbol{C}$ has not changed $\Rightarrow$ Goto 1.

In each RANSAC round an estimate for $\boldsymbol{\Omega}$ is found. Subsequently, either an average value over all $\boldsymbol{\Omega}$ can be calculated or the set with the smallest error with respect to $\boldsymbol{f}(\boldsymbol{\Omega})$ can be selected.

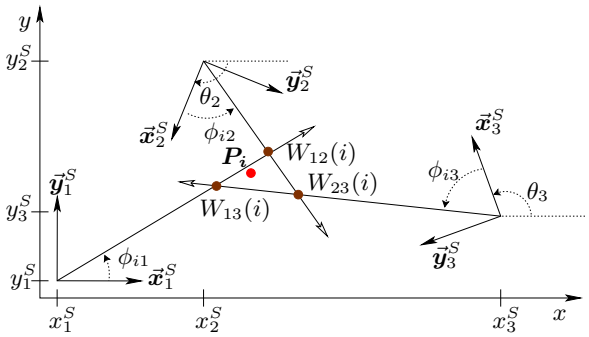

Figure 4: *Projection error used for consensus set extension*

Step two of the RANSAC (Eq. (7)) is a Monte-Carlo approach to solve the initialization problem discussed in Section 3. Although the random initialization and the reformulation of the system of equations considerably improved the convergence of the Newton method, in some runs the RANSAC method did not converge to a coherent solution and aborted after a maximum number of iterations, see the following section for quantitative results.

6. Experiments

We compiled a database consisting of artificially reverberated recordings with a minimum duration of 90 s per setup. Two rooms with reverberation times (T_{60}) between 50 ms and 450 ms were simulated using the image method from [10], including a speaker walking along a random path.

Experiments revealed that a symmetric placement of sensors (e.g. a sensor in each corner of a room [4]) considerably improves the calibration results. However, since we cannot assume that sensors are always arranged in a symmetric way, we will consider symmetric (room A: 4×4 m, sensors centered at each wall) and asymmetric arrangements (room B: 4×3.5 m) in our experiments. Fig. 5 illustrates the sensor placements.

The calibration error is measured in terms of the translation error T_{Err} which is the average difference between the estimated and the true sensor positions, and the rotation error R_{Err}, which is the average error of the estimated sensor orientations. We will compare the calibration results using Eq. (2) and ideal scaling factors (superscript *tan*) against the results obtained by using Eq. (5) and either ideal scaling factors (superscript *ideal*) or

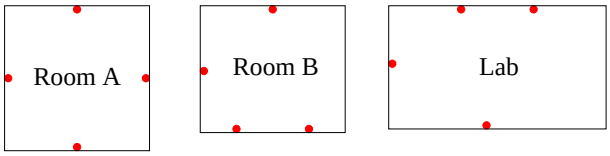

Figure 5: *Sensor positions (red dots) in rooms*

scaling factors estimated from TDoA values (superscript *real*). The calibration results are given for employing the RANSAC method (RS) (for *tan*, *ideal* and *real*) or for using all available observations within a least squares (LS) solution (*ideal* and *real* only) .

Since the speakers were moving during the recordings, the TDoA values had to be estimated by using a small window size of 4096 samples at 16 kHz sampling rate. Subsequently, the TDoA values were processed using a Kalman filter. Therefor the TDoA values were transformed to length differences $\chi_{jk}(i) = c \cdot \tau_{jk}(i)$ and a random walk process was assumed for the speaker movement. This approach achieved a TDoA error reduction of approximately 40%.

T_{60}	T_{Err}^{tan}	R_{Err}^{tan}	T_{Err}^{ideal} $[m]$		T_{Err}^{real} $[m]$		R_{Err} $[^\circ]$	
[ms]	[m]	[°]	LS	RS	LS	RS	LS	RS
50	0.12	3.70	0.04	0.05	0.28	0.27	1.03	1.47
100	0.14	4.28	0.09	0.09	0.10	0.12	1.19	0.97
150	0.26	8.18	0.12	0.13	0.33	0.14	3.05	3.57
200	0.20	6.13	0.06	0.07	0.20	0.37	1.34	1.35
250	0.38	11.00	0.13	0.14	0.36	0.45	2.90	3.32
300	0.17	3.00	0.15	0.11	0.17	0.26	3.08	3.07
350	0.30	9.09	0.11	0.09	0.23	0.36	2.67	1.71
400	0.20	3.61	0.20	0.23	0.32	0.44	3.07	4.19
450	1.30	89.42	0.08	0.07	0.20	0.47	1.26	0.32

Table 1: Room A: Symmetric sensor placement

Table 1 shows the experimental results for room A with the symmetric sensor placement, while Table 2 displays the results of the asymmetric sensor placement of room B. The calibration errors are higher than in the symmetric case, especially if the system of equations with the *tan*-functions is employed. It can be noted that the errors do not increase monotonically with the amount of reverberation. We attribute this to the random movement of the speaker which results in more or less disturbed observations.

T_{60}	T_{Err}^{tan}	R_{Err}^{tan}	T_{Err}^{ideal} $[m]$		T_{Err}^{real} $[m]$		R_{Err} $[^\circ]$	
[ms]	[m]	[°]	LS	RS	LS	RS	LS	RS
50	0.13	4.00	0.07	0.05	0.09	0.04	2.52	1.66
100	0.36	5.39	0.23	0.14	0.25	0.15	5.84	2.96
150	0.83	29.69	0.36	0.16	0.36	0.28	10.68	4.75
200	0.84	16.38	0.37	0.49	0.66	0.73	9.65	12.33
250	0.52	4.94	0.58	0.74	1.26	0.96	13.30	17.91
300	0.95	14.59	0.66	0.50	1.17	0.50	14.40	15.08
350	1.00	25.65	0.39	0.69	1.10	0.74	10.41	19.07
400	0.77	20.54	0.82	0.96	0.97	1.20	21.96	26.07
450	1.00	9.79	0.36	0.36	1.45	0.64	7.37	8.17

Table 2: Room B: Asymmetric sensor placement

For both setups, the new proposed system of equations (5) (T_{Err}^{ideal}, R_{Err}) significantly reduces the translation error and the rotation error compared to the previously proposed form (2) (T_{Err}^{tan}, R_{Err}^{tan}). The improved numerical stability can be seen by the fact that the *tan*-formulation (Eq. (2)) did not converge in 44.7% of the RANSAC rounds, while the new proposed formulation (Eq. (5)) failed only in 13.6% of all random initializations.

The experiments show that RANSAC has no notable advantage in the easy case of a symmetric sensor placement (room A). However, in the tricky case of an asymmetric sensor placement (room B), RANSAC mostly outperforms the least squares solution. Scaling the geometry with the estimated TDoA values increases the translation error (T_{Err}^{real} vs. T_{Err}^{ideal}). This can be attributed to an overestimation of the scaling factor caused by erroneous TDoA values.

We also conducted experiments in our laboratory (3.4×6 m, $T_{60} \approx 157$ ms), where we used an asymmetric sensor placement (see Fig. 5). The average translation error was $T_{Err}^{ideal} = 0.22$ m and the rotation error was $R_{Err} = 2.3^\circ$. When we used the TDoA estimates to scale the geometry the translation error was increased to $T_{Err}^{real} = 0.25$ m. These results are in line with the simulations.

7. Conclusions

We have presented a new approach for calibrating a sensor network consisting of spatially distributed microphone arrays. The approach utilizes observations from moving speakers and does not need any special calibration signal or artificial user behavior. It reduces the average translation and rotation error by roughly a factor of 2 compared to a recently published method. Additionally, we used TDoA values to solve the scale indeterminacy problem of calibration methods which are based on DoA-only observations. The findings from simulations were confirmed by an experimental setup in the laboratory.

8. Acknowledgements

This work has been supported by Deutsche Forschungsgemeinschaft (DFG) under contract no. Ha3455/7-1 and Fi799/5-1.

9. References

[1] J. Schmalenstroeer, R. Haeb-Umbach, “Online Diarization of Streaming Audio-Visual Data for Smart Environments”, IEEE Journal of Selected Topics in Signal Processing, vol. 4, no. 5, pp. 845-856, oct. 2010

[2] S. Valente, et. al., “Geometric calibration of distributed microphone arrays from acoustic source correspondences”, Proc. IEEE Int. Workshop on Multimedia Signal Processing, 2010

[3] M. Hennecke, et. al.,“A Hierarchical Approach to unsupervised shape calibration of microphone array networks”, Proc. IEEE/SP Workshop on Statistical Signal Processing, Sep. 2009

[4] J. Kemper, M. Walter, H. Linde, “Human-Assisted Calibration of an Angulation based Indoor Location System”, Proc. Int. Conference on Sensor Technologies and Applications, pp.196-201, 2008.

[5] A. Redondi, M. Tagliasacchi, F. Antonacci, A. Sarti, “Geometric calibration of distributed microphone arrays”, IEEE Int. Workshop on Multimedia Signal Processing, Oct. 2009

[6] I. McCowan, M. Lincoln, I. Himawan, “Microphone Array Shape Calibration in Diffuse Noise Fields”, IEEE Trans. On Audio, Speech, and Language Processing, vol. 16, no. 3, mar. 2008

[7] E. Warsitz, R. Haeb-Umbach, “Acoustic filter-and-sum beamforming by adaptive principal component analysis”, Proc. IEEE Int. Conf. on Acoustic, Speech, and Signal Processing, Mar. 2005

[8] C. Knapp, G. Carter, “The generalized correlation method for estimation of time delay”, IEEE Trans. Acoust., Speech, Signal Process., vol. ASSP-24, no. 4, pp. 320-327, Aug. 1976.

[9] M. Fischler, R. Bolles, “Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography”, Magazine Comm. of the ACM, vol. 24, issue 6, 1981

[10] J. B. Allen, D. A. Berkley, “Image method for efficiently simulating small-room acoustics”, Journal Acoust. Soc. Amer., vol. 65, no. 4, pp. 943-950, Apr. 1979.