3D ROOM GEOMETRY ESTIMATION FROM MEASURED IMPULSE RESPONSES

Sakari Tervo and Timo Tossavainen

Aalto University School of Science
Department of Media Technology
P.O. Box 15400, FI00076 Aalto

ABSTRACT

Estimation of the room geometry from spatial room impulse responses is studied. An algorithm for estimating the geometry is presented. The algorithm does not require any a priori information on the room shape, number of walls, or order of the reflections, but deduces the set of planes that explain the measured source and image-source locations and covariances iteratively. The algorithm is demonstrated with real data experiments.

Index Terms— Room geometry estimation, room impulse response, reflection

1. INTRODUCTION

The geometry of the room is one of the most essential parts of room acoustic modeling. Besides the prediction of the acoustics of rooms, the room acoustic models can be used for example to enhance source localization performance [1].

Estimation of the room geometry can be divided into three subtopics, localization of reflections, i.e. the image-sources, estimation of the surface parameters, i.e plane points and normals, and the estimation of room geometry. In principle, any general localization method can be used to localize the reflections. As an example, in [2] reflections are localized using sound intensity vectors and time of arrival (TOA).

The locations of the reflections can be used together with the estimated or a priori known source location to deduce the surface parameters. This requires the knowledge of the order of the reflections. Plane parameters are estimated in [3] by rotating a B-format microphone around a loudspeaker, directed towards the microphone. The estimation is based on the TOA and the direction of arrival (DOA) of the first arriving reflection in each direction. The TOA and DOA measurements are grouped using hierarchical clustering to avoid estimating the same plane multiple times. Moreover, in [4] the plane parameters are estimated with a common tangent algorithm. The same approach is applied in [5] and several other publications for the estimation of plane parameters.

The actual room geometry estimation algorithms combine the locations of reflections and source as well as the orders of the reflections. One such algorithm, which uses only one room impulse response, has been proposed in [6]. The algorithm requires the knowledge of the order of the first and second order reflections and of their arrival times. Moreover, in [7] a constrained room model and l_1-regularized least-squares method is applied to fit 3-D shoebox model to a set of measured impulse responses. The number of walls is assumed to be known a priori. In addition, in [3] the clustering of the TOA and DOA measurements constitutes as the room geometry estimation algorithm. The geometry estimation presented in [3], as well as in [4] and [5], use the assumption that all the detected reflections are of first order. To the understanding of the present authors all the previous approaches use a priori information either on the number of the walls, shape of the enclosure, or on the order of the reflections. Especially the a priori assumption on the order of the reflections is not feasible, since in most of the practical situations the earliest arriving second order reflection arrives before the latest first order reflection.

Here a room geometry estimation algorithm is proposed that is able to deduct the room geometry without any of the above listed a priori information. The algorithm deduces iteratively the set of planes that has produced a set of estimated reflection locations and covariances. Rest of the article is organized as follows. Section 2, presents the estimation of the reflection locations and of their covariance matrices from the spatial room impulse responses. In Section 3, the geometry that explains the estimated locations and covariances of the reflections is estimated with an iterative maximum likelihood algorithm. Experiments are conducted with real data in Section 4. Section 5 discusses the results and concludes the article.

2. ESTIMATION OF SOURCE AND REFLECTION LOCATIONS

2.1. Reflection signal model

In this paper, a room impulse response measured with a microphone at location $\mathbf{r}_n$ and a loudspeaker at location $\mathbf{x}$ is considered as a sum of the direct sound and individual reflections:

$$\begin{aligned} h_n(t) &\triangleq h(\mathbf{r}_n, \mathbf{x}; t) = \left[\sum_{k=0}^{K} h_{k,n}(t)\right] + w_n(t) \\ &= \left[\sum_{k=0}^{K} \left(\int_{-\infty}^{\infty} H_{k,n}(\omega) e^{j\omega t} d\omega\right)\right] + w_n(t), \quad (1) \end{aligned}$$

where t is time, ω is angular frequency, n is the index for microphone, $k = 0$ is the direct sound, $k = 1, \ldots, K$ are the reflections, $w_n(t)$ is measurement noise independent for each microphone and of the signal and distributed according to normal distribution for each microphone. Moreover, $h_{k,n}(t)$ and $H_{k,n}(\omega)$ are the time and the frequency domain presentation of the direct sound and of the reflections.

The applied microphone array is assumed to have a small aperture size compared to the dimensions of the room. Then the impulse responses can be divided into short time windows which each include only one reflection. In realistic situations this is true for the

This work was supported by ERC grant agreement no. [203636], HECSE, and Nokia Foundation

first arriving reflections up to some time instant. Therefore, here this windowed reflection signal is considered as

$$H_{k,n}(\omega) = A_{k,n}(\omega)S(\mathbf{n}_k,\omega)\exp\{-j\omega t_{k,n}(\mathbf{x}_k)\} + W_n(\omega) \quad (2)$$

where $W_n(\omega) \in \mathbb{C}$ is the frequency domain presentation of the noise signal $w_n(t)$, and $S(\mathbf{n}_k,\omega) \in \mathbb{C}$ is the impulse response of the loudspeaker to the direction $\mathbf{n}_k$ of the reflection location $\mathbf{x}_k$, $A_{k,n}(\omega)$ is the gain of the reflection or of the direct sound, and the time of arrival (TOA) is given for a reflection location $\mathbf{x}_k$ as

$$t_{k,n}(\mathbf{x}_k) \triangleq t(\mathbf{r}_n;\mathbf{x}_k) = c^{-1}\|\mathbf{r}_n - \mathbf{x}_k\|, \quad (3)$$

where c is the speed of sound. The gain factor is dependent at least on the properties of the surfaces $s = 1,\ldots,S$ which the sound wave has encountered, frequency dependent air absorption, which is dependent on the distance of the reflection path distance, the attenuation according to the $1/r$-law by the distance of the reflection path, and the directivity of the microphone. Here ideal specular reflections are assumed, the microphones are omni-directional and other phenomena affecting the gain factor are assumed to be linear. Then the gain factor affects only the amplitude of the reflection, i.e. $A_{k,n}(\omega) \in \mathbb{R}$ and does not contribute to the delay of the reflection.

2.2. Maximum likelihood estimation of time of arrival

Time delay estimation framework presented in [8] is applied to estimate the TOA of the reflection and of the direct sound. The cross correlation function between the measured reflection signal $h_{k,n}(t)$ and the a priori measured source signal $s(\mathbf{n}_k,t)$ is calculated via the generalized correlation function [8]

$$\mathrm{R}_{s,h_{k,n}}(t) = \mathcal{F}^{-1}\{\mathcal{W}^{\mathrm{ML}}_{s,h_{k,n}}(\omega)G_{s,h_{k,n}}(\omega)\}. \quad (4)$$

The maximum likelihood weighting for cross correlation is given as [8]

$$\mathcal{W}^{\mathrm{ML}}_{s,h_{k,n}}(\omega) = \frac{1}{|G_{s,h_{k,n}}(\omega)|}\frac{C_{s,h_{k,n}}(\omega)}{[1-C_{s,h_{k,n}}(\omega)]}. \quad (5)$$

$C_{s,h_{k,n}}(\omega) = \|G_{s,h_{k,n}}(\omega)\|^2/[G_{s,s}(\omega)G_{h_{k,n},h_{k,n}}(\omega)]$ is the magnitude squared coherence and the spectral densities $G(\cdot)$ are written based on the assumed reflection signal model as $G_{s,s}(\omega) = \|S(\omega)\|^2$, $G_{s,h_{k,n}}(\omega) = A_{k,n}(\omega)\|S(\omega)\|^2\exp\{-j\omega t_{k,n}(\mathbf{x}_k)\}$, and $G_{h_{k,n},h_{k,n}}(\omega) = \|A_{k,n}(\omega)S(\omega)\|^2 + \|W_n(\omega)\|^2$, since the source signal and noise are independent. Then, the ML weighting for the auto-correlation method is given as

$$\mathcal{W}^{\mathrm{ML}}_{s,h_{k,n}}(\omega) = \cdots = \frac{A_{k,n}(\omega)}{\|W_n(\omega)\|^2}. \quad (6)$$

In practical situations an estimate of the noise $\|\widehat{W}_n(\omega)\|^2$ is measured from the beginning of the impulse response where there is no signal and the gain is then estimated as $\widehat{A}_{k,n}(\omega) = \sqrt{\left(\|\widehat{H}_{k,n}(\omega)\|^2 - \|\widehat{W}_n(\omega)\|^2\right)/\|\widehat{S}(\omega)\|^2}$ when $\|\widehat{H}_{k,n}(\omega)\|^2 > \|\widehat{W}_n(\omega)\|^2$ and 0 otherwise. Here $\widehat{\cdot}$ denotes a measurement. The maximum argument of the cross correlation is then the TOA estimate, i.e $\hat{t}_{k,n} = \arg\max_t\{\mathrm{R}_{s,h_{k,n}}(t)\}$.

2.3. Maximum likelihood estimation of the source and reflection positions

The TOA estimation errors are assumed to be uncorrelated and have normally distributed errors with variance $\sigma^2_{k,n}$. Then, the maximum likelihood estimation function for location with TOA estimates is given as [9, Ch. 7]

$$p(\mathbf{x}|\hat{\boldsymbol{t}}_k,\Sigma_k) = \frac{\exp(-\frac{1}{2}[\hat{\boldsymbol{t}}_k - \boldsymbol{t}_k(\mathbf{x})]^{\mathrm{T}}\Sigma_k^{-1}[\hat{\boldsymbol{t}}_k - \boldsymbol{t}_k(\mathbf{x})])}{(2\pi)^{(N)/2}\sqrt{\det(\Sigma_k)}}, \quad (7)$$

where N is the number of microphones, $\hat{\boldsymbol{t}}_k = [\hat{t}_{k,1},\hat{t}_{k,2},\ldots,\hat{t}_{k,N}]^{\mathrm{T}}$ are the TOA estimates, $\boldsymbol{t}_k(\mathbf{x}) = [t_{k,1}(\mathbf{x}),t_{k,2}(\mathbf{x}),\ldots,t_{k,N}(\mathbf{x})]^{\mathrm{T}}$ are the true TOAs, and $\Sigma_k = \mathrm{diag}\left(\sigma^2_{k,1},\sigma^2_{k,2},\ldots,\sigma^2_{k,N}\right)$ is the TOA error covariance matrix with individual variances σ^2 on the diagonal. The maximum argument of this MLE-TOA function is the ML estimate for the reflection location $\hat{\mathbf{x}}_k = \arg\max_{\mathbf{x}}\{p(\mathbf{x}|\hat{\boldsymbol{t}}_k,\Sigma_k)\}$. Since the problem is non-convex the solution can typically be found only using non-convex optimization methods and here the maximum is searched using Levenberg-Marquardt-algorithm.

2.4. Estimation of the error covariances

The minimum covariance of the localization error, i.e. the Cramér-Rao lower bound (CRLB) is given by the inverse of the Fisher information [9, Ch. 3]. In the TOA-based localization the CRLB is given by

$$\boldsymbol{\Sigma}(\mathbf{x}) \geq \mathbf{J}^{-1}(\mathbf{x}) = \left[\left[\frac{\partial}{\partial\mathbf{x}}\boldsymbol{t}(\mathbf{x})\right]^{\mathrm{T}}\Sigma^{-1}\left[\frac{\partial}{\partial\mathbf{x}}\boldsymbol{t}(\mathbf{x})\right]\right]^{-1}. \quad (8)$$

The minimum variance of TOA estimation is given by the inverse of its Fisher information, and by using the knowledge of the spectral densities the Fisher information simplifies to

$$J(t) = \frac{T}{\pi}\int_0^\infty \omega^2\frac{C_{s,h}(\omega)}{1-C_{s,h}(\omega)}d\omega = \frac{T}{\pi}\int_0^\infty \omega^2\mathrm{SNR}(\omega)d\omega, \quad (9)$$

where $\mathrm{SNR}(\omega) = \frac{\|A(\omega)S(\omega)\|^2}{\|W(\omega)\|^2}$ is the signal-to-noise ratio. Since the source signal and the noise term are assumed uncorrelated the SNR of the current reflection signal can be estimated as

$$\widehat{\mathrm{SNR}}_{k,n}(\omega) = \left(\|\widehat{H}_{k,n}(\omega)\|^2 - \|\widehat{W}_n(\omega)\|^2\right)/\|\widehat{W}_n(\omega)\|^2. \quad (10)$$

Then the inverse of the error covariance matrix for TOA estimation is estimated as

$$\hat{\Sigma}_k^{-1} = \mathrm{diag}\left(\hat{\sigma}^{-2}_{k,1},\ldots,\hat{\sigma}^{-2}_{k,N}\right) = \mathrm{diag}\left(\hat{J}(t_{k,1}),\ldots,\hat{J}(t_{k,n})\right), \quad (11)$$

where $\hat{J}(t_{k,n})$ for n^{th} microphone and k^{th} reflection is calculated using Eq. (10) as the SNR estimate in Eq. (9). This TOA error covariance matrix is then used in Eq. (8) to estimate the covariance matrix of the localization error.

3. 3D GEOMETRY ESTIMATION

Room geometry estimation is based on the estimated image sources, $\mathbf{x}_0,\mathbf{x}_1,\ldots,\mathbf{x}_k$ in order of distance from the microphone array at the origin with corresponding estimated error covariances $\boldsymbol{\Sigma}_k \triangleq \boldsymbol{\Sigma}(\mathbf{x}_k)$. The reflection order of a given reflection is unknown, but it is assumed that $\mathbf{x}_0$ is the source. As is well-known, in the image source model, image sources are generated from the original source by reflecting it with respect to surfaces. A plane $\pi : \mathbf{x}\cdot\mathbf{n} + a = 0$, where $\|\mathbf{n}\| = 1$ is the plane's normal, generates the first order image source $\mathbf{x} - 2(a + \mathbf{x}\cdot\mathbf{n})\mathbf{n}$ that can be reflected again in another plane to produce a higher order image source. Assuming the theory valid

and neglecting visibility, the process of iterating reflections of the source away from the microphone array generates all the observed image sources.

Supposing a convex space bounded by a few dominant planes, the geometry estimation method seeks to find the planes and the reflection paths generating the observed image sources. A heuristic greedy method is used for computational efficiency. The method starts with an empty set of planes and backtraces a reflection path from each $\mathbf{x}_i$ to $\mathbf{x}_0$ adding new planes when existing planes can not account for $\mathbf{x}_i$. The backtracing is searching multiple order paths through existing planes that have been generated in the earlier steps. The image sources are processed in order of increasing distance from the source. When backtracing from $\mathbf{x}_i$, the method tries to find a low-order reflection path from $\mathbf{x}_i$ to $\mathbf{x}_0$. When a plausible path does not exist using the present planes, a new plane is added to connect $\mathbf{x}_i$ to some $\mathbf{x}_j$ with $j < i$. The process continues until all image sources have a reflection path from the source. The result is a set of planes giving the estimated room geometry and a reflection path from each $\mathbf{x}_i$ to $\mathbf{x}_0$.

The algorithm considers planes between two image sources, that is, planes π_{ij} with normals given by $\mathbf{n}_{ij} = \|\mathbf{x}_i - \mathbf{x}_j\|^{-1}(\mathbf{x}_i - \mathbf{x}_j)$ and constants by $a_{ij} = -\|\mathbf{x}_i - \mathbf{x}_j\|^{-1}(\mathbf{x}_i + \mathbf{x}_j)/2$. Such a plane is uncertain, because it is formed between two uncertain image sources; its uncertainty can be expressed using a covariance matrix Σ_π for the parameters $(\mathbf{n}, a)$. Forming the Jacobian matrices $\mathbf{J}_i$ and $\mathbf{J}_j$ from $\mathbf{x}_i$ and $\mathbf{x}_j$ to $(\mathbf{n}, a)$ and using standard covariance propagation gives the plane's covariance

$$\Sigma_\pi = \mathbf{J}_i \Sigma_i \mathbf{J}_i^T + \mathbf{J}_j \Sigma_j \mathbf{J}_j^T.$$

Note that general plane parameters are homogeneous, but forming the planes with unit normals causes variance to vanish in directions that change $\|\mathbf{n}\|$ in Σ_π giving a meaningful covariance for the plane. Also, reflection is simpler to deal with using unit normals. If π reflects $\mathbf{x}$ to $\mathbf{x}'$, the covariances of π and $\mathbf{x}$ are Σ_π and Σ_x, and the Jacobian matrices of $\mathbf{x}'$ with respect to π and $\mathbf{x}$ are $\mathbf{J}_\pi$ and $\mathbf{J}_x$, respectively, then covariance propagation again gives the reflection's covariance $\Sigma_{x'} = \mathbf{J}_\pi \Sigma_\pi \mathbf{J}_\pi^T + \mathbf{J}_x \Sigma_x \mathbf{J}_x^T$. A reflection from $\mathbf{x}$ to another point $\mathbf{y}$ is ruled out if the Mahalanobis distance

$$D^2 = (\mathbf{x}' - \mathbf{y})^T (\Sigma_{x'} + \Sigma_y)^{-1} (\mathbf{x}' - \mathbf{y})$$

is large. Considering $\mathbf{x}'$ and $\mathbf{y}$ multivariate normal, the distance D^2 follows a χ^2 distribution with 3 degrees of freedom. If D^2 exceeds a confidence region's upper bound obtained from the inverse cumulative χ^2 distribution, then the corresponding confidence ellipsoid for $\mathbf{x}' - \mathbf{y}$ does not include $\mathbf{0}$ and $\mathbf{x}' \neq \mathbf{y}$ with the chosen degree of confidence, assuming that the linearizations are valid. A reflection between image sources in a plane is considered plausible if reflections both ways have small Mahalanobis distances and are the best reflections for both sources. Initially the algorithm finds all plausible reflection pairs for all planes π_{ij}.

While backtracing the reflection path for an image source $\mathbf{x}_i$, the algorithm first considers single reflections from image sources with paths to $\mathbf{x}_0$ using the current set of planes and chooses a plausible reflection that minimizes the reflection order of $\mathbf{x}_i$ and breaks ties with D^2. If no plausible single reflection is found, the algorithm considers two reflections using the current planes using covariance propagation to deal with uncertainty over multiple reflections and chooses similarly the option with smallest reflection order and breaks ties with D^2. If a plausible path is not found even then, the algorithm adds a new plane from the set of predefined planes, choosing a plane that connects $\mathbf{x}_i$ to some $\mathbf{x}_j$ with $j < i$. The plane is chosen to have least some support in the data set (i.e. more than one plausible reflection), to minimize the reflection order of $\mathbf{x}_i$ and breaks ties with the number of plausible reflections in the data set with respect to that plane.

Table 1: Hand measured dimensions and normals of the planes in the experiment as well as microphone array dimensions. The spacing between each axis is equal to 100 mm.

	Microphone #					
Dimension #	1	2	3	4	5	6
x [mm]	50	-50	0	0	0	0
y [mm]	0	0	50	-50	0	0
z [mm]	0	0	0	0	50	-50
	Plane parameters in the experiments			
Refl. #	$n_x[\cdot]$	$n_y[\cdot]$	$n_z[\cdot]$	d [m]
1	-1.00	0.00	0.00	3.54
2	0.00	-1.00	0.00	4.68
3	0.00	0.00	-1.00	1.88
4	1.00	0.00	0.00	3.55
5	0.00	1.00	0.00	4.67
6	0.00	0.00	1.00	1.88

4. REAL ROOM EXPERIMENTS

Real data experiments were conducted in a shoebox-shaped class room (7.09m×9.35m×3.76m) stripped of chairs and tables. Skeleton model of the room is shown in Fig. 1 and the hand-measured values for the plane parameters are given in Table 1. As illustrated in Fig. 1, there was a closet on the west wall, an extrusion, a window, and a door on the south wall, and a whiteboard and a door on the east wall. The walls of the room are of painted sheet rock and the floor is concrete covered with a plastic mat. These materials have a reasonably low absorption coefficient and it is expected that they produce clearly identifiable reflections to the impulse responses. However, the ceiling of the room has been treated with absorptive panels, which reduces the amplitude of the reflections. Moreover, lamps, ventilation, and other equipment typical for a modern class room are hanging from the ceiling.

Impulse responses were measured using sine-sweep signals with Genelec 1029A and G.R.A.S vector intensity probe VI50 with the geometry presented in Table 1. The loudspeaker was located in the back of the room and the microphone array in front. The experiment was repeated four times with different locations for the loudspeaker and the microphone array. The height of the loudspeaker and array was from 1m to 1.5 m. In addition, the sampling frequency was set to 48 kHz and the speed of sound was estimated to be 345.2 m/s based on temperature and humidity.

The sparse impulse response technique proposed in [2] was used for measuring the impulse responses. Impulse responses to directions from 0 to 360 degrees between 10 degrees around the z-axis were measured. The impulse responses were then divided into short time windows of size 1.5 ms with an overlap of 95 %. The maximum direction that produces the highest absolute pressure on average in the microphone array was selected to represent the impulse response during that time window, similarly as described in [2]. The use of this technique ensures that the source signal is similar to all directions and also allows better temporal and spatial separation between the reflections. In addition, only one impulse response of the loudspeaker $S(\mathbf{n}_0, \omega)$ to the direction $\mathbf{n}_0$ that produces the highest

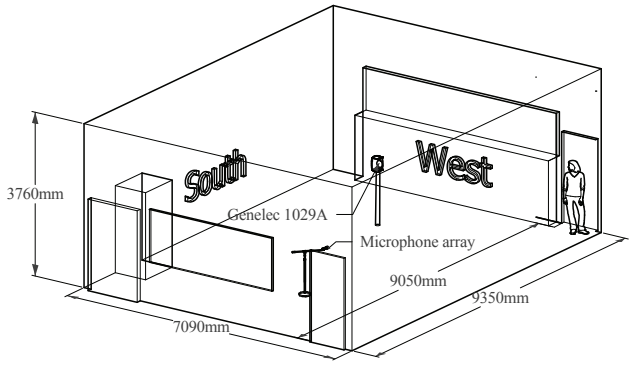
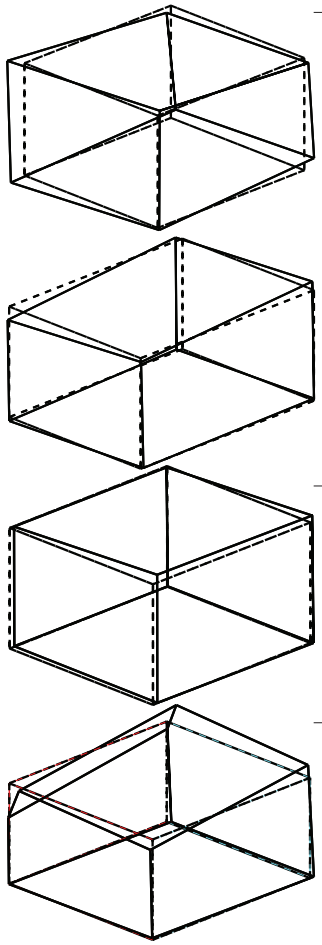


Fig. 1: Experimental setup.

pressure is required in the TOA estimation.

The location of the direct sound and of those reflections which are local maxima in absolute pressure-wise in the compound sparse response (described in [2]), and have a SNR higher than 30 dB is calculated as stated in Section 2. On average about 27 reflections were found by using this criterion. The room geometry is then estimated individually for all four loudspeaker-microphone array location pairs using the technique described in Section 3. The limiting parameter for the χ^2 distribution is set to 0.9 in the experiments.

The results of the experiments are shown in Fig. 2, where the estimated and true room geometries are illustrated with lines and the plane parameters (normal $\mathbf{n}$ and distance d) with respect to the center of the room are given in a table. In the first three cases six planes are found and the geometry is estimated with a good accuracy. In the last case seven planes are found and the ceiling is not estimated well. This is due to the fact that the absorptive panels attenuate the first order ceiling reflection in this setup so much that it can not be found. In the other setups the absorptive panels are not directly on the reflection path.



Experiment #1

#	n_x	n_y	n_z	d
1	-1.00	-0.02	0.00	3.53
2	-0.01	-0.99	-0.12	5.35
3	0.03	-0.05	-1.00	1.89
4	1.00	-0.00	0.05	4.18
5	-0.03	1.00	-0.00	4.67
6	-0.03	-0.03	1.00	1.99

Experiment #2

#	n_x	n_y	n_z	d
1	-1.00	-0.03	-0.02	3.54
2	0.02	-1.00	-0.01	4.82
3	0.03	0.03	-1.00	1.72
4	1.00	0.05	0.01	3.52
5	-0.03	1.00	0.02	4.67
6	-0.01	0.00	1.00	1.90

Experiment #3

#	n_x	n_y	n_z	d
1	-1.00	-0.00	-0.00	3.57
2	0.00	-1.00	-0.02	4.82
3	0.04	-0.01	-1.00	1.96
4	1.00	0.01	0.00	3.57
5	0.01	1.00	-0.00	4.43
6	-0.00	0.01	1.00	1.87

Experiment #4

#	n_x	n_y	n_z	d
1	-1.00	0.01	-0.03	3.47
2	0.00	-1.00	-0.05	4.78
3	-0.05	0.10	-0.99	1.88
4	1.00	-0.01	0.01	3.55
5	-0.01	1.00	0.00	4.66
6	-0.01	0.00	1.00	1.88

Fig. 2: Estimated and true room geometries presented with solid and dashed line, respectively. Also shown are the estimated parameters.

5. CONCLUSIONS

Room geometry is required in room acoustic prediction and can benefit, for example, source localization. This article studied the estimation of room geometry from measured room impulse responses. A method for deducing the room geometry was presented. Unlike previous approaches, the proposed method does not require a priori knowledge of the number of walls, room shape, or the order of the reflections. The method is based on iteratively searching planes that explain the measured source and image-source locations and their covariances. The results from real room experiments show that the methods works well in realistic conditions. The use of higher order reflections to enhance the room geometry estimation is in the future work of the present authors.

6. REFERENCES

[1] T. Korhonen, *Acoustic Source Localization Utilizing Reflective Surfaces*, Ph.D. thesis, Tampere Uni. of Tech., 2010.

[2] S. Tervo, J. Pätynen, and T. Lokki, "Acoustic reflection path tracing using a highly directional loudspeaker," in *IEEE WASPAA*, 2009, pp. 245–248.

[3] B. Gunel, "Room shape and size estimation using directional impulse response measurements," in *Forum Acusticum*, 2002, pp. 1–7.

[4] F. Antonacci, A. Sarti, and S. Tubaro, "Geometric reconstruction of the environment from its response to multiple acoustic emissions," in *IEEE ICASSP*, 2010, pp. 2822–2825.

[5] J. Filos, E.A.P. Habets, and P.A. Naylor, "A two-step approach to blindly infer room geometries," in *IWAENC*, 2010.

[6] I. Dokmanic, Y.M. Lu, and M. Vetterli, "Can one hear the shape of a room: the 2–d polygonal case," in *IEEE ICASSP*, 2011, pp. 321–324.

[7] D. Ba, F. Ribeiro, C. Zhang, and D. Florêncio, "L1 regularized room modeling with compact microphone arrays," in *IEEE ICASSP*, 2010, pp. 157–160.

[8] C. Knapp and G. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech and Signal Proc.*, vol. 24, no. 4, pp. 320–327, 1976.

[9] S.M. Kay, *Fundamentals of Statistical signal processing, Volume 1: Estimation theory*, Prentice-Hall, New Jersey, USA, 1998.