GREEDY ALTERNATIVE FOR ROOM GEOMETRY ESTIMATION FROM ACOUSTIC ECHOES: A SUBSPACE-BASED METHOD

Mario Coutino[1], Martin Bo Møller[2,3], Jesper Kjær Nielsen[2,3], Richard Heusdens[1]
[1]Delft University of Technology, The Netherlands
[2]Bang & Olufsen A/S, Denmark
[3]Aalborg University, Denmark

***Abstract*— In this paper, we present a greedy subspace method for the acoustic echoes labeling problem, which occurs in applications such as source localization and room geometry estimation. The orthogonal projection into the null space of the microphones position matrix is used to filter and sort all possible combinations of echoes. A greedy strategy, based on the rank constraint of Euclidean distance matrices (EDMs), is used on the sorted subset of echo combinations to extract the feasible combinations. Numerical simulations using room impulse responses (RIRs) from shoe-box shaped rooms show that the method provides improvements in terms of computational complexity and the number of required measurements with respect to a recently published graph-based method.**

***Index Terms*— acoustic echoes, room geometry, sorting reflections, greedy algorithm, source localization**

I. INTRODUCTION

In the past years there has been an increasing interest in mapping the shape of a room using acoustic echos [1]-[3]. Knowledge of the room shape can benefit a large number of applications. For example, in the creation of personal sound zones [4][5] one needs to know the room impulse response (RIR) in different locations, which could be modeled if the geometry information of the room is available. In autonomous navigation, knowledge of the enclosure boundary aids collision avoidance. For speech enhancement, knowledge of walls reflections is desirable to compensate for reverberation.

Echoes generated by sound reflected from the room walls carry information about the geometry of the enclosure. By modeling room reflections using virtual sources [6], it is possible to exploit the geometric duality of this representation to estimate the room boundaries. For this purpose, several methods have been proposed to estimate the room geometry with high accuracy. Most of these methods assume knowledge of the RIRs. In [7], the shape in the 2D case is estimated by a single RIR. Antonacci et al. [8] solve the 2D problem assuming multiple sources and microphones.

In instances where multiple microphones, randomly placed in the room, are used to detect the acoustic echos in the RIRs, ambiguities arise at the moment of labeling the echoes according to the wall which produces them. This problem is illustrated in Fig. 1. In order to deal with this issue, Dokmanić et al. in [9] exploits the properties of EDMs to find the room geometry in the general 3D case. More recently, a newly proposed method [2] by Jager et al. has been shown to provide the same accuracy as Dokmanić's method but at a much lower computational complexity. This approach recasts the labeling problem of the acoustic echoes problem into a graph problem, which can be solved in reasonable time for instances with a small number of microphones. However, both [9] and [2] become intractable for an increasing number of microphones.

In this paper, we aim to build on previous work to further improve the current state-of-the-art solution for the acoustic echo labeling problem. We propose a subspace-based filtering to reduce the computational complexity of the graph-based approach of [2]. Furthermore, we devise a greedy strategy which attains comparable performance to the graph-based method at a reduced computational cost. In addition, the proposed method only requires measurements from a single source, in contrast to the current state-of-the-art method that requires more than one source. In this work, we restrict ourselves to shoe-box shaped rooms as they are commonly found in typical audio reproduction scenarios. However, the proposed method can be extended to other (convex) room geometries.

This work was partially supported by CONACYT

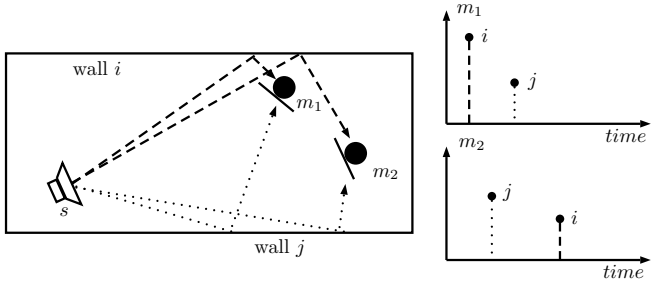

Fig. 1: Ambiguity in the echoes labels due to different order of arrival of wall reflections

II. DATA MODEL

First, let us consider an arbitrary set $\mathcal{M}$ of M microphones located at random positions. That is, $\mathcal{M} = \{\mathbf{r}_m = [x_m, y_m, z_m]^T \in \mathbb{R}^3\}_{m=1}^M$. These locations are assumed known up to a non-singular transformation. Furthermore, consider the set $\mathcal{S} = \{\mathbf{s}_n = [X_n, Y_n, Z_n]^T \in \mathbb{R}^3\}_{n=1}^N$ of N image sources. The squared distances $\mathcal{D} = \{d_{m,n}\} \forall\, (m,n) \in [1,\dots,M] \times [1,\dots,N]$ between the image sources $\mathcal{S}$ and receivers $\mathcal{M}$ can be measured, i.e., the time-of-arrival (TOA) of the reflections can be estimated at the microphones. Hence, the squared distance $d_{m,n}$ for the (m,n)-th pair can be written as

$$(x_m - X_n)^2 + (y_m - Y_n)^2 + (z_m - Z_n)^2 = d_{m,n} \tag{1}$$

This can be expressed as an inner product as [10]

$$\mathbf{R}_m^T \mathbf{S}_n = d_{m,n} \tag{2}$$

where the two vectors $\mathbf{R}_m$ and $\mathbf{S}_n$ are given by

$$\mathbf{R}_m = [\mathbf{r}_m^T \mathbf{r}_m \; -2x_m \; -2y_m \; -2z_m \; 1]^T \in \mathbb{R}^{5\times 1}, \tag{3}$$

$$\mathbf{S}_n = [1 \; X_n \; Y_n \; Z_n \; \mathbf{s}_n^T \mathbf{s}_n]^T \in \mathbb{R}^{5\times 1} \tag{4}$$

Collecting all the squared distances $d_{m,n}$ for the pairs (m,n) leads to the distance matrix $\mathbf{D} \in \mathbb{R}^{M\times N}$, and the model can be written in matrix form as

$$\mathbf{R}^T \mathbf{S} = \mathbf{D} \in \mathbb{R}^{M\times N} \tag{5}$$

where $\mathbf{R} = [\mathbf{R}_1, \ldots, \mathbf{R}_M]$ and $\mathbf{S} = [\mathbf{S}_1, \ldots, \mathbf{S}_N]$ are known microphone and unknown image source position matrices, respectively. Even when the positions of the microphones are known up to an arbitrary non-singular matrix $\mathbf{H} \in \mathbb{R}^{5\times5}$, and the transformed microphones positions matrix $\hat{\mathbf{R}}^T = \mathbf{R}^T\mathbf{H}$ is known instead of $\mathbf{R}$, the model in (5) still holds as

$$\hat{\mathbf{R}}^T\mathbf{H}^{-1}\mathbf{H}\hat{\mathbf{S}} = \mathbf{D}. \tag{6}$$

where $\hat{\mathbf{S}} = \mathbf{H}^{-1}\mathbf{S}$ is the transformed matrix of sources positions.

III. LABELING ACOUSTIC ECHOES

From the model in (5), the unknown matrix $\mathbf{S}$ with the position of the sources can be estimated by means of least squares given that $\text{rank}(\mathbf{R}) \geq 5$ when the positions of the receivers and the distance matrix $\mathbf{D}$ are known. However, in most cases, the squared distances in $\mathcal{D}$ are not grouped according to the sources that originate them. That is, the subindex n from the elements in $\mathcal{D}$ is unknown to us. Therefore, an approach to generate $\mathbf{D}$ from the unlabeled set $\mathcal{D}$ is needed.

In this work, we consider the projection into the null space of $\mathbf{R}$, denoted by $\mathcal{N}(\mathbf{R})$, to filter and sort all possible combinations of echoes. This projection exploits the structure in the model (5) to estimate the matrix $\mathbf{D}$ from the unlabeled data $\mathcal{D}$. The goal of this approach is to deliver a complexity reduction similar to the one achieved in [2] allowing us to deal with larger instances of the problem generated either by a larger number of microphones and sources, or by uncertainties in the set $\mathcal{D}$.

When proper diversity in $\mathbb{R}^3$ is assumed for the microphone positions, i.e., non co-located locations for the receivers, the only constraint needed in the method to ensure the rank-5 property of the distance matrix $\mathbf{D}$ is $M \geq 5$ [10][11].

A. Subspace Filtering

Let the rank-5 economy-sized SVD of the known receivers position matrix $\mathbf{R}$ be given by

$$\mathbf{R} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T. \tag{7}$$

The orthogonal projection $\Pi_{\mathcal{N}(\mathbf{R})}$ onto $\mathcal{N}(\mathbf{R})$ can then be computed from the SVD in (7) as

$$\Pi_{\mathcal{N}(\mathbf{R})} = \mathbf{I}_M - \mathbf{V}\mathbf{V}^T \tag{8}$$

This projection can be shown to have the property

$$\Pi_{\mathcal{N}(\mathbf{R})}\mathbf{R}^T = \mathbf{0}, \tag{9}$$

hence from (5) it follows that

$$\Pi_{\mathcal{N}(\mathbf{R})}\mathbf{R}^T\mathbf{S} = \Pi_{\mathcal{N}(\mathbf{R})}\mathbf{D} = \mathbf{0}, \tag{10}$$

for $\mathbf{D}$-matrices with the correct sorting. In this work (10) is used to estimate $\mathbf{D}$ from $\mathcal{D}$. An interesting property of the complementary projection matrix is that

$$\|\Pi_{\mathcal{N}(\mathbf{R})}\|_2 = 1 \tag{11}$$

which implies that there is no amplification of errors, i.e.,

$$\begin{aligned} \|\Pi_{\mathcal{N}(\mathbf{R})}(\mathbf{D}_c + \mathbf{n})\|_2 &= \|\Pi_{\mathcal{N}(\mathbf{R})}(\mathbf{R}^T\mathbf{S}_c + \mathbf{n})\|_2 && (12)\\ &= \|\Pi_{\mathcal{N}(\mathbf{R})}\mathbf{n}\|_2 && (13)\\ &\leq \|\mathbf{n}\|_2 && (14) \end{aligned}$$

where $\mathbf{D}_c$ is the c-th column of the matrix $\mathbf{D}$. This makes the projection particularly useful in cases where the elements of $\mathcal{D}$ are perturbed with noise.

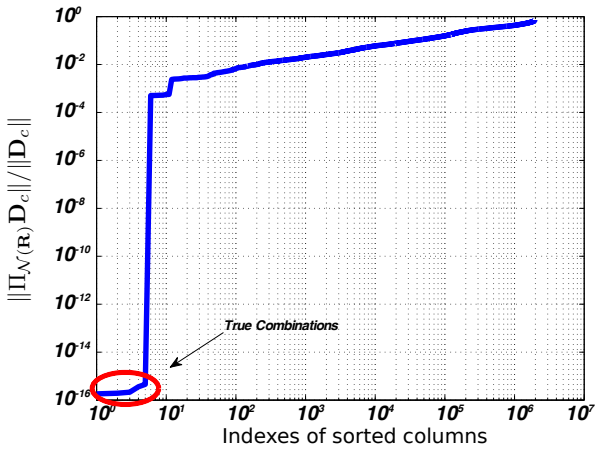

Fig. 2: Normalized functional (16) for the columns $\tilde{\mathbf{D}}$, in the noise free case, sorted in ascending order. In this example $M = 9$ and $N = 5$.

In order to apply the projection given in (8), we first consider the matrix $\tilde{\mathbf{D}}$ defined as the distance matrix generated by all the possible combinations of the elements in $\mathcal{D}$, e.g.,

$$\tilde{\mathbf{D}} = \begin{bmatrix} d_{1,1} & \cdots & d_{1,2} & \cdots & d_{1,N} \\ d_{2,1} & \cdots & d_{2,1} & \cdots & d_{2,N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ d_{M,1} & \cdots & d_{M,2} & \cdots & d_{M,N} \end{bmatrix} \in \mathbb{R}^{M\times N^M} \tag{15}$$

In the ideal case, i.e., measurements without any kind of noise, the results are straightforward. By defining the functional

$$f(c) = \|\Pi_{\mathcal{N}(\mathbf{R})}\tilde{\mathbf{D}}_c\|_2^2 \;\forall\; c \;\in\; [1, \ldots, N^M] \tag{16}$$

we can select the subset of feasible columns as

$$\mathcal{C} = \{c \,:\, f(c) = 0\}, \tag{17}$$

and provide an estimate of the feasible distance matrix given by

$$\hat{\mathbf{D}} = \tilde{\mathbf{D}}_{\mathcal{C}} \in \mathbb{R}^{M\times N} \tag{18}$$

where $\tilde{\mathbf{D}}_{\mathcal{C}}$ represents the trimmed distance matrix, which only retains the columns specified by the set $\mathcal{C}$. The functional is illustrated in Fig. 2 for a problem instance with $M = 9$ microphones and $N = 5$ (image) sources.

However, in real applications there is no guarantee that the true distances $\mathcal{D}$ are measured perfectly, hence the set in (17) will, most likely, turn out empty. In order to deal with noisy measurements, we provide a column-dependent upper bound for the proposed functional which considers the effect of perturbations.

Consider that the measured squared distance $\hat{d}_{m,n}$ can be expanded as

$$\hat{d}_{m,n} \triangleq (\sqrt{d_{m,n}} + w_{m,n})^2 = d_{m,n} + 2\sqrt{d_{m,n}}w_{m,n} + w_{m,n}^2 \tag{19}$$

where $w_{m,n}$ is the noise in the (m,n)-pair measurement with power σ_w^2. After the projection is applied to a stacked version of (19), the following residual is obtained

$$\Pi_{\mathcal{N}(\mathbf{R})}\hat{\mathbf{D}}_c = \Pi_{\mathcal{N}(\mathbf{R})}\left[2\text{diag}(\mathbf{w}_c)\mathbf{D}_c^{\circ\frac{1}{2}} + \text{diag}(\mathbf{w}_c)\mathbf{w}_c\right] \in \mathbb{R}^{M\times1} \tag{20}$$

where $\mathbf{A}^{\circ p}$ denotes the p-th Hadamard power of the matrix $\mathbf{A}$ and $\mathbf{w}_c$ is the measurement noise vector associated with the c-th column of $\mathbf{D}$. Therefore, it is possible to provide a selection rule similar

to (17) by upper bounding the square norm of the expression in (20). An appropriate upper bound for the residual norm can be given by

$$\begin{aligned} \|\Pi_{\mathcal{N}(\mathbf{R})}\hat{\mathbf{D}}_c\|_2^2 &= \|\Pi_{\mathcal{N}(\mathbf{R})}[\text{diag}(\mathbf{w}_c)(2\mathbf{D}_c^{\circ\frac{1}{2}} + \mathbf{w}_c)]\|_2^2 \quad (21)\\ &\leq \|\Pi_{\mathcal{N}(\mathbf{R})}\|_2^2\|\text{diag}(\mathbf{w}_c)\|_2^2\|2\mathbf{D}_c^{\circ\frac{1}{2}} + \mathbf{w}_c\|_2^2 \quad (22)\\ &\leq 4\max^2(\mathbf{w}_c)\|\mathbf{D}_c^{\circ\frac{1}{2}} + 0.5\mathbf{w}_c\|_2^2 \quad (23)\\ &= \kappa_c \quad (24)\end{aligned}$$

where $\max(\mathbf{a})$ denotes the maximum absolute value of the vector $\mathbf{a}$, and the fact that $\|\Pi_{\mathcal{N}(\mathbf{R})}\|_2 = 1$ has been used. Using the derived upper bound, we can build the subset of columns for the distance matrix as

$$\mathcal{C} = \{c : f(c) \leq \kappa_c\} \quad (25)$$

and estimate the distance matrix using expression (18). Although the bound always holds, κ_c is not directly available from the measurements. As in practice, we deal with realizations of the measurement process, in order to use the bound in (24) we introduce

$$\kappa_c^{(i)} = 4\gamma^i\sigma_w^2\|\hat{\mathbf{D}}_c^{\circ\frac{1}{2}}\|_2^2, \;\; \gamma \geq 1 \quad (26)$$

as a surrogate to provide a practical iterative threshold for the functional. The power of the noise σ_w^2 can be assumed known as it is considered that the accuracy of the method employed for obtaining the TOA estimates is known. For simplicity, we consider that all columns are subject to the same noise level σ_w^2. This assumption affects the performance of the bound as sources located at different positions have different accuracy levels. However, this can be considered a reasonable assumption as the ordering of echoes is unknown. In simulations it has been observed that $\kappa_c^{(0)}$ is sufficient for the method to deliver adequate results so our algorithm fixes κ_c^i to κ_c^0.

B. Avoiding the Graph Problem

For real measurements, $|\mathcal{C}| \gg N$. Therefore, further processing is required to only select feasible columns. For this step two possible strategies can be applied: (i) the recently proposed graph-based method from [2], where the problem is recast as a maximum independent set problem, or (ii) a greedy approach that sequentially selects feasible combinations. In the following, we solely focus on (ii) as we want to avoid solving the NP-hard problem of listing all maximal independents sets.

To avoid the graph-problem, we first make the observation that when using the functional $f(c)$ for sorting the columns of $\tilde{\mathbf{D}}$, the columns of the lowest normalized functional value, meeting the rank constraint for EDMs, most likely belong to the true distance matrix. For this, consider the matrix $\mathbf{E} \in \mathbb{R}^{M\times M}$ as the EDM constructed from the relative distances between receivers. The matrix $\tilde{\mathbf{E}}_c$ denotes an augmented EDM built by adding the distances from the vector $\hat{\mathbf{D}}_c$. The rank of $\tilde{\mathbf{E}}_c$ is larger than five, if $\mathbf{E}$ is augmented with distances to different sources. As suggested in [2], the ϵ-rank defined as [12]

$$\text{rank}(\tilde{\mathbf{E}}_c, \epsilon) = \min_{\|\tilde{\mathbf{E}}_c - \mathbf{X}\|_2 \leq \epsilon} \text{rank}(\mathbf{X}) \quad (27)$$

can be employed to sequentially exclude echoes combinations that, approximately, violate the rank constraint. As the threshold ϵ is unknown a priori, an iterative approach is employed to obtain the suitable candidate for ϵ.

Secondly, as pointed out in [2], the columns in $\hat{\mathbf{D}}$ are unlikely to share elements, so in addition to the sequential exclusion of columns by the ϵ-rank constraint, columns sharing elements with already selected feasible columns are rejected. The sub-optimal algorithm combining these two observations is presented in Algorithm 1. The $\eta > 1$ parameter controls the growth of the rank constraint. This allows the solution to only include the best ranked columns.

Algorithm 1 Subspace-based Greedy Acoustic Echoes Sorting

Input: $\mathcal{D}$, $\Pi_{\mathcal{N}(\mathbf{R})}$, $\mathbf{E}$, ϵ_0, N, $\sigma_{\mathbf{w}}$
Output: $\mathbf{D}$

Initialization: Generate $\tilde{\mathbf{D}}$ and $\boldsymbol{\kappa}^{(0)}$, $\mathbf{D} = \{\}$, $\epsilon = \epsilon_0$
$\mathcal{C} = \{c : f(c) \leq \boldsymbol{\kappa}_c^{(0)}\}$
$\mathcal{C}_s = \text{sort}(\mathcal{C}, f(c)/\|\tilde{\mathbf{D}}_c\|_2^2, \text{"ascending"})$
$\hat{\mathbf{D}} = \tilde{\mathbf{D}}_{\mathcal{C}_s}$
while numCols$(\mathbf{D}) < N$ **do**
  for $c = 1$ **to** $|\mathcal{C}_s|$ **do**
    $\tilde{\mathbf{E}}_c = \begin{bmatrix} \mathbf{E} & \hat{\mathbf{D}}_c \\ \hat{\mathbf{D}}_c^T & 0 \end{bmatrix}$
    if rank$(\tilde{\mathbf{E}}_c, \epsilon) \leq 5$ and $\hat{\mathbf{D}}_c \cap \mathbf{D} == \emptyset$ **then**
      $\mathbf{D} = [\mathbf{D}, \hat{\mathbf{D}}_c]$
    end if
  end for
  if numCols$(\mathbf{D}) < N$ **then**
    $\epsilon = \eta\epsilon$
  end if
end while

Finally, after the matrix $\mathbf{D}$ is estimated by the greedy approach, the least squares solution for the estimates of the source locations, for $M \geq 5$, can be directly obtained by

$$\hat{\mathbf{S}} = (\mathbf{R}^T)^\dagger\mathbf{D}. \quad (28)$$

Contrary to (i), where more than one maximum independent set can be found in the graph, (ii) provides a unique solution. The unique solution allows the echoes to be sorted even when the constraint imposed by Polleyfey's method [10] used in [2] is not met.

Notice that if measurements from Q acoustic sources are available, i.e.,

$$\mathbf{D}_{\text{Tot}} = \left[\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_Q\right] = \mathbf{R}^T\left[\mathbf{S}_1, \ldots, \mathbf{S}_Q\right]. \quad (29)$$

A combination of Pollefeys' method, using the SVD of $\mathbf{D}_{\text{Tot}}$, and Procrustes analysis can be performed to estimate the image source positions instead of using (28). This approach could lead to better reconstruction results for cases in which the pseudo-inverse of $\mathbf{R}^T$ is not well conditioned.

IV. NUMERICAL SIMULATIONS

In this section results from numerical simulations comparing the proposed greedy method and a modified version of [2] are presented. First, to evaluate the proposed method we generated a set of 500 Monte Carlo simulations for different uncertainties in the measured distances. The simulation illustrates the acoustic echo labeling problem from multiple room reflections, i.e., $N = 6$ for a 3D shoe-box shaped room. The number of microphones considered is $M = 9$. The noise-free distances from the reflections of the walls are taken from the peaks in the simulated impulse responses (RIRs) generated by the acoustics simulation software [13]. As the graph-based method requires multiple sources to provide an estimate of the source positions, a version with an oracle is used instead, i.e., if more than one maximal independent set is found, the closest set (in the least square sense) with respect to the noisy distance matrix is considered as the solution of the method. To provide a speed up to the method, the subspace filtering step is added in this modified version. The addition of the subspace filtering to the method shows

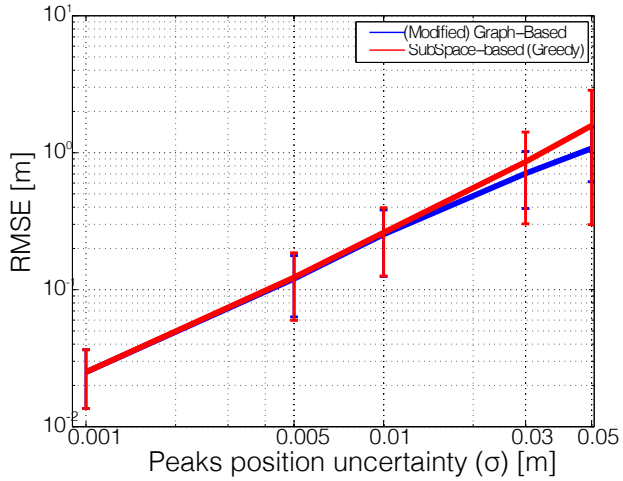

Fig. 3: Estimation error comparison for $M = 9$ and $N = 6$. Error bars show results within one standard deviation.

that it is possible to deliver a feasible implementation of the graph-based approach for a relatively large number of microphones, contrary to the intractability statement given in [14].

In Fig. 3 the estimation error of both methods is compared. The error is computed as the norm of the Euclidean distance between the true $\mathbf{s}_n$ and estimated position $\hat{\mathbf{s}}_n$ of each source, i.e., $\|\mathbf{e}\|_2 = \| \left[\text{dist}_E(\mathbf{s}_1, \hat{\mathbf{s}}_1), \ldots, \text{dist}_E(\mathbf{s}_N, \hat{\mathbf{s}}_N)\right]^T \|_2$, where the estimated positions are found by (28) assuming $\mathbf{R}$ known (up to a non-singular transform). Notice how the accuracy of the estimation decreases as the uncertainty in the distances increases. For low uncertainties, i.e., $\sigma < 0.01$m, the accuracy of both methods is identical. However, as the uncertainty increases, the results in Fig. 3 show higher degradation of the greedy approach due to its sub-optimality.

A comparison of the relative running time of each method with respect the baseline case of $M = 5$ using the graph-based approach is shown in Fig. 4. For this comparison 500 Monte Carlo simulations were made using different number of microphones. The time consumed by the methods with subspace filtering is considerably lower than the graph-based approach. This result shows that it is possible to find tractable solutions for larger instances of the graph problem by adding subspace filtering. By pre-

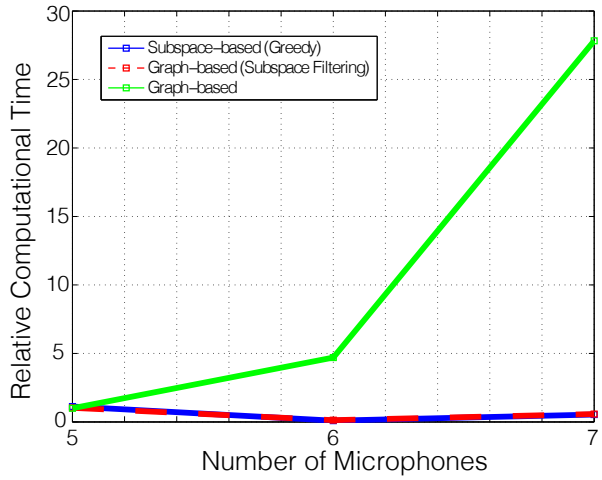

Fig. 4: Comparison of computation time between the graph-based methods and greedy approach for number of microphones.

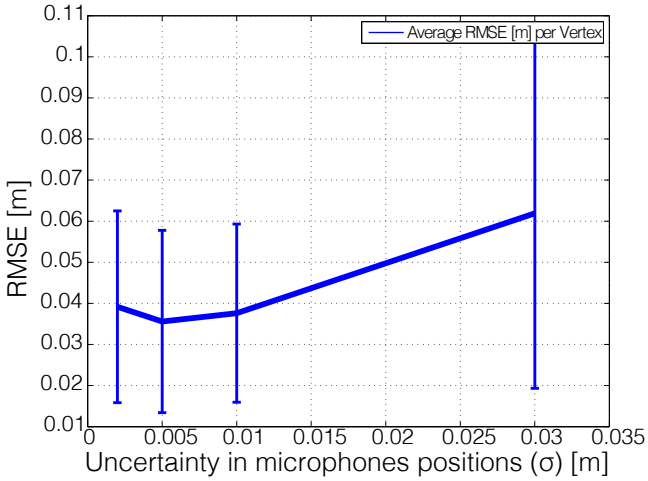

Fig. 5: Average RMSE per vertex of the room for the proposed greedy strategy for different uncertainties in the locations of the microphones. Error bars show results within one standard deviation.

filtering the combinations using the proposed functional, the number of computed SVDs reduces drastically. In addition, by removing combinations that might not be rejected by the rank constraint, the number of nodes in the graph, used to find the maximum independent sets, is reduced. Hence, the method gains an additional speed up. The reduction in time when the number of microphones increases from $M = 5$ to $M = 6$ for the methods with subspace filtering is explained by the selectivity of the kernel of $\mathbf{R}$. As more combinations of echoes are rejected by the subspace filtering, less ϵ-rank checks are performed to obtain a feasible distance matrix $\mathbf{D}$.

Finally, Fig. 5 illustrates the performance of the proposed method for different uncertainties in the locations of the microphones. For this experiment, 500 simulated measurements were produced from four different acoustic sources. The reconstruction of the image sources positions is done using the noisy locations of the microphones and Pollefey's method [10]. In this experiment, it is assumed that distances between each microphone-image source pair contain additive white Gaussian noise with standard deviation $\sigma_{RIR} = 1$cm. Notice how even in the presence of noise, in both RIRs peaks and microphones locations, the method provides vertex estimates with average RMSE close to 5cm. The high dependency on the accuracy of the positions of the microphones is seen in the increased standard deviation of the RMSE and its mean value.

All numerical simulations were run on a Macbook Air (Mid 2013) 1.7 GHz Inter Core i7 using non-optimized MATLAB code.

V. CONCLUSIONS

In this paper we proposed an alternative approach for the acoustic echoes labeling problem. Using a complementary orthogonal projection related to the receiver locations, it is possible to construct a filtering and sorting criteria for the columns of the distance matrix built from all possible combinations of available echoes. It is shown, that in the noise free case perfect identification of the true columns can be achieved. Furthermore, for the noisy case, a greedy alternative is proposed to avoid the solution of the NP-hard problem of listing all maximal independent sets in a graph. Numerical simulations show the applicability of the method and the benefits of applying the subspace filtering to the original graph-based method. In addition, effects of uncertainties in the distance measurements, not discussed in literature before, were shown through numerical experiments.

REFERENCES

[1] M. Kreković, I. Dokmanić, and M. Vetterli. "EchoSLAM: Simultaneous localization and mapping with acoustic echoes." 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2016.

[2] I. Jager, R. Heusdens, and N.D. Gaubitch. "Room geometry estimation from acoustic echoes using graph-based echo labeling." 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2016.

[3] I. Dokmanić, L. Daudet, and M. Vetterli. "From acoustic room reconstruction to slam." 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2016.

[4] T. Betlehem, et al. "Personal Sound Zones: Delivering interface-free audio to multiple listeners." IEEE Signal Processing Magazine 32.2 (2015): 81-91.

[5] F. Jacobsen, et al. "A comparison of two strategies for generating sound zones in a room." 18th International Congress on Sound and Vibration. 2011.

[6] J.B. Allen, and D.A. Berkley. "Image method for efficiently simulating small room acoustics." The Journal of the Acoustical Society of America 65.4 (1979): 943-950.

[7] I. Dokmanić, Y.M. Lu, and M. Vetterli. "Can one hear the shape of a room: The 2-D polygonal case." 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2011.

[8] F. Antonacci, et al. "Inference of room geometry from acoustic impulse responses." IEEE Transactions on Audio, Speech, and Language Processing 20.10 (2012): 2683-2695.

[9] I. Dokmanić, et al. "Acoustic echoes reveal room shape." Proceedings of the National Academy of Sciences 110.30 (2013): 12186-12191.

[10] M. Pollefeys, and D. Nister. "Direct computation of sound and microphone locations from time-difference-of-arrival data." ICASSP, pp. 2445-2448. 2008.

[11] J.C. Gower. "Properties of Euclidean and non-Euclidean distance matrices." Linear Algebra and its Applications 67 (1985): 81-97.

[12] G.H. Golub and C.F. Van Loan, Matrix Computations, North Oxford Academic, Oxford, third edition, 1983

[13] E. A. Habets, Room impulse response generator, Technische Universiteit Eindhoven, Tech. Rep 2, 1 (2006).

[14] I. Jager. "Room Shape Estimation from Acoustic Echoes using Graph-based Echo Labeling." MSc. Thesis., TU Delft, Delft University of Technology, 2015.