KU LEUVEN

STADIUS
Center for Dynamical Systems, Signal Processing and Data Analytics

Citation/Reference	Authors: Kathleen MacWilliam, Filip Elvander, Toon van Waterschoot Title: Simultaneous Acoustic Echo Sorting and 3-D Room Geometry Inference Conference: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) Vol: Pages:
Archived version	☐ Submitted version ☐ Accepted (Author manuscript: the content is identical to the content of the published paper, but without the final typesetting by the publisher) ■ Published ☐ Published (Open Access)
Acknowledgements	In this manuscript, the following funding is acknowledged: ■ EU H22020 ☐ EU Horizon Europe ■ FWO ☐ KULeuven Internal Funds ☐ VLAIO ☐ AI ☐ Other
Published version	https://ieeexplore.ieee.org/document/10096005
Copyright statement	©2023 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works
Author contact	kathleen.macwilliam@esat.kuleuven.be

ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) | 978-1-7281-6327-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICASSP49357.2023.10096005

SIMULTANEOUS ACOUSTIC ECHO SORTING AND 3-D ROOM GEOMETRY INFERENCE

Kathleen MacWilliam, Filip Elvander†, Toon van Waterschoot**

*Department of Electrical Engineering (ESAT-STADIUS), KU Leuven, Belgium
†Department of Signal Processing and Acoustics, Aalto University, Finland
emails: *firstname.lastname@esat.kuleuven.be , †filip.elvander@aalto.fi

ABSTRACT

Room geometry estimation from multiple acoustic room impulse responses (RIRs) relies on being able to correctly identify first-order echoes corresponding to the same reflector. This can be done by exploiting the properties of various mathematical concepts but often results in needing to solve large combinatorial problems owing to the increased number of measurements these tools require for efficacy. A low-complexity iterative method is proposed for determining the boundaries in convex polygonal rooms using common tangent planes to sets of ellipsoids. Prior knowledge of the order of echoes is unnecessary and only three RIRs are required, which is one less than the minimum number typically needed to find a unique reflector in 3-D, allowing for simpler common tangent plane estimation and reducing possible echo combinations. Candidate partial rooms are built by checking whether expected second-order echoes appear in the RIRs, adding a reflector at each iterate until a room is found. The proposed method is validated by means of computer simulations.

Index Terms— Room inference, echo labeling, acoustic room impulse responses

1. INTRODUCTION

Room boundaries are generally the dominant acoustic reflectors in everyday environments and prior knowledge of them is a key component in applications such as room acoustic modeling and source localization [1, 2]. Estimating the shape of a room using acoustic echoes is advantageous because it makes use of information already acquired from Room Impulse Response (RIR) measurements instead of having to use separate devices to define boundaries relative to source and microphone configurations [3, 4]. The time of arrival (TOA) of echoes retrieved from an acoustic RIR can be converted into total path lengths traveled by an emitted sound wave to reach a receiver. If the Image Source Method (ISM) [5] is employed as the underlying room acoustic model, these path lengths can be equated to the distance between "image sources" and the receiver under consideration. Assuming the relative positions of the source and microphone are known, a prolate spheroid (hereafter referred to as an ellipsoid) corresponding to each first-order image source can be formed with the locations of the source and microphone as the foci. If multiple RIRs are available and the echo TOAs corresponding to the same first-order image source can correctly be identified, a common tangent plane to the respective ellipsoids will align with the reflector under consideration [6]. At least d+1 RIRs, where d is the dimension of the domain, would usually be required in order to get a unique reflector position using this approach. One would also need to be able to identify first-order echoes and their correct groupings across the RIRs. The latter is referred to as the echo labeling problem (ELP) in line with [7].

Some methods resolve the ELP, or at least reduce its size, by imposing geometric constraints. For example, the works [6, 8, 9] place sources close to reflectors under consideration, and in [10] a compact microphone array is used to ensure echoes arrive in the same order. Other methods require that first-order echoes have already been identified [11]. In [12], the ELP is circumvented altogether by using a single co-located source and microphone, however, it requires that first- and second-order echoes have previously been distinguished and is not robust to noise. More recently, convolutional neural networks have been used to estimate room acoustical parameters using a single room impulse response in [13]. Other relevant previous works that solve the ELP without imposing specific geometric constraints or assuming any prior knowledge of the echo order are [14], which uses elliptical constraints, and [7, 15, 16, 17], which use the properties of Euclidean Distance Matrices (EDMs). However, these methods require at least six RIRs in 3-D in order to use the tools they propose. Furthermore, although these methods solve the ELP in a mathematically rigorous manner, they often require a computation time that would not be suited to real-time applications [16].

In this paper we propose a method able to jointly perform echo-order estimation, echo labeling, and room geometry estimation. In contrast to similar previous works, the proposed method only requires access to three RIRs, which is one less than the minimum four RIRs required by other methods to find a unique reflector in 3-D. In this case, there will typically be two (or more) common tangent planes that could potentially correspond to the desired reflector. This scenario has generally been over-looked because there will be more than one room that could generate the same set of first-order echoes. However, this problem can be overcome by taking into account resulting expected second-order image source information, as is proposed herein. By doing this, the dimension of the combinatorial problem of labeling the echoes is substantially reduced.

To the best of the authors' knowledge, there are no widely available analytical competitor methods directly applicable to the scenario considered here, i.e., the case of handling three RIRs in which the correct labeling (both order and grouping) of echoes is unknown. Specifically, most earlier works either require more microphones or rely on first-order TOAs already having been identified. Simulations for a box-shaped room are included in this paper, but the method can be extended to other room shapes in which the ISM is applicable [18].

This work was supported in part by the Research Foundation – Flanders (FWO, grant 12ZD622N) and the European Research Council under the European Union's Horizon 2020 research and innovation program / ERC Consolidator Grant: SONORA (no. 773268). This paper reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on October 03,2024 at 10:14:35 UTC from IEEE Xplore. Restrictions apply.

2. PROBLEM FORMULATION AND SIGNAL MODEL

In this work, we consider a room with a single source placed at position $\mathbf{s} \in \mathbb{R}^3$ and microphones at positions $\mathbf{r}_m \in \mathbb{R}^3$, for $m = 1, 2, 3$. By translation, we may without loss of generality assume that $\mathbf{s} = \mathbf{0}$. A polygonal room can be seen as the intersection of a set of half-spaces defined by infinite planes that contain each room boundary. Given that boundary planes, also referred to as reflectors, will not pass through the origin, the equation

$$l_1 x + l_2 y + l_3 z + 1 = 0 \tag{1}$$

describes such a plane, where $\mathbf{n} = [l_1, l_2, l_3]^T$ is the corresponding normal vector. Then, a room with K boundaries, i.e., walls, ceiling, and floor, is completely described by a corresponding set of normal vectors $\mathcal{N} = \{\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_K\}$. With this, assume that each microphone records an RIR, and let $\boldsymbol{\tau}_m$, for $m = 1, 2, 3$, be the sets of TOAs estimated[1] from these RIRs. As these sets of TOAs encode information about the room geometry, and in particular its reflection paths, the aim of the proposed method is to estimate the reflector set $\mathcal{N}$ given $\boldsymbol{\tau}_m$, for $m = 1, 2, 3$.

We here rely on using the ISM to parameterize TOAs in terms of first- and second-order image sources. In the ISM of [5], a (physical) source is reflected across every plane $\mathbf{n}_k$, defining a set of first-order image source positions, $\mathbf{q}_k \in \mathbb{R}^3$ for $k = 1, \ldots, K$. For the special case of the source being placed at the origin, these coordinates are given by $\mathbf{q}_k = -2\mathbf{n}_k / \|\mathbf{n}_k\|^2$. The corresponding first-order TOAs $\tau^{(1)}_{k,m}$ observed at each microphone m can be written as a function of these image sources according to

$$\tau^{(1)}_{k,m}(\mathbf{q}_k) = \frac{\|\mathbf{r}_m - \mathbf{q}_k\|_2}{c}, \tag{2}$$

where $c = 343$ m/s is the speed of sound. Higher-order image sources may then be found iteratively by reflecting previous-order image sources over the planes containing the room boundaries. Now, each pair $(\mathbf{r}_m, \tau^{(i)}_{k,m})$ of microphone position and first-order TOA, together with the source position $\mathbf{s}$, define an ellipsoid to which the corresponding reflector, i.e., boundary surface $\mathbf{n}_k$, is tangent. An illustration of this in the analogous 2-D case is shown in Fig. 1(a). It may be noted here that this ellipsoid consists of all points $\mathbf{p}_e \in \mathbb{R}^3$ satisfying

$$\frac{\|\mathbf{p}_e - \mathbf{s}\|_2 + \|\mathbf{p}_e - \mathbf{r}_m\|_2}{c} = \tau^{(1)}_{k,m}, \tag{3}$$

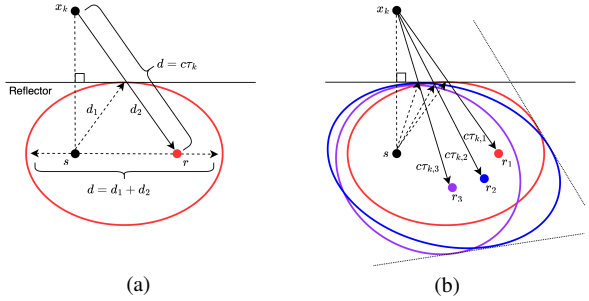
which can equivalently be expressed using the homogeneous equation in matrix form,

$$\mathbf{p}^T \mathbf{Q} \mathbf{p} = 0 \tag{4}$$

where $\mathbf{p} = \begin{bmatrix} \mathbf{p}_e^T & 1 \end{bmatrix}^T$ and the entries of $\mathbf{Q} \in \mathbb{R}^{4\times 4}$ are the quadric parameters found by expanding (3). With this, a normal vector $\mathbf{n}_k$ will be tangent to the ellipsoid, and therefore constitute a potential reflector for the corresponding TOA, if and only if $\begin{bmatrix} \mathbf{n}_k^T & 1 \end{bmatrix} \mathbf{Q}^{-1} \begin{bmatrix} \mathbf{n}_k^T & 1 \end{bmatrix}^T = 0$ (see, e.g., [22]). For the case where measurements from M microphones are available, this implies that a reflector $\mathbf{n}_k$ should satisfy

$$\sum_{m=1}^{M} \begin{bmatrix} \mathbf{n}_k^T & 1 \end{bmatrix} \mathbf{Q}_m^{-1} \begin{bmatrix} \mathbf{n}_k^T & 1 \end{bmatrix}^T = 0, \tag{5}$$

[1]The TOA estimation problem will not be considered in this paper, as several methods capable of identifying both first and second-order TOAs are available in the literature (see, e.g., [19], [20], [21]). We will, however, take into account potential TOA estimation errors, assuming that individual echoes can be resolved.



(a) (b)

Fig. 1. a) Ellipse defined by microphone position $\mathbf{r}$, source position $\mathbf{s}$, and TOA τ_k, together with tangent reflector and, b) Ellipses defined by triplets $(\mathbf{s}, \mathbf{r}_m, \tau_{k,m})$, for $m = 1, 2, 3$, corresponding to a common first-order image source. Note that grouping only two ellipses would result in there being more than one common tangent.

where $\mathbf{Q}_m$ are the ellipsoidal parameters for the pair $(\mathbf{r}_m, \tau^{(i)}_{k,m})$. In previous works (see, e.g., [6]), $M > 3$, meaning that the non-linear equation (5) can only be solved in an approximate, e.g., least-squares, sense. In contrast, we herein consider the case $M = 3$, allowing for exact roots of the equation to be found. This, however, comes at the price of potentially non-unique solutions $\mathbf{n}_k$ (c.f., Fig. 1(b)). In the proposed method, we remedy this issue by also exploiting the information from second-order reflections.

3. PROPOSED METHOD

Following the model outlined in the previous section, two main difficulties are encountered when estimating the room geometry. Firstly, one needs to be able to correctly group ellipsoids belonging to the same first-order image source, which is a combinatorial problem. Secondly, as only three microphones are used, the first-order information is not sufficient for uniquely identifying the location of the first-order image source, or equivalently the corresponding reflector. In particular, two consistent tangent planes may exist for any set of ellipsoids in the general set-up considered.

The proposed method addresses these two challenges jointly by exploiting the information contained in second-order echoes. In particular, the proposed method identifies potential groupings of first-order echoes by checking if the different RIRs contain TOAs that are consistent with second-order reflections. Recalling the ISM, second-order image sources are found by reflecting a first-order image source over any first-order plane $\mathbf{n}_k$, except the one from which it was generated. Specifically, it may be readily verified that two first-order image sources $\mathbf{q}_i$ and $\mathbf{q}_j$ uniquely define a second-order image source position $\mathbf{q}_{ij} \in \mathbb{R}^3$, $i \neq j$, according to

$$\mathbf{q}_{ij} = \mathbf{q}_i + \mathbf{q}_j - 2\frac{\mathbf{q}_i^T \mathbf{q}_j}{\|\mathbf{q}_j\|_2^2}\mathbf{q}_j \tag{6}$$

with corresponding second-order TOAs $\tau^{(2)}_{ij,m}$ given by

$$\tau^{(2)}_{ij,m}(\mathbf{q}_i, \mathbf{q}_j) = \frac{\|\mathbf{r}_m - \mathbf{q}_{ij}\|_2}{c}. \tag{7}$$

Thus, for our sets of TOAs $\boldsymbol{\tau}_1$, $\boldsymbol{\tau}_2$, and $\boldsymbol{\tau}_3$, two estimated first-order image sources are said to constitute a feasible pair only if there exists a consistent valid second-order image source, i.e., whose TOA (7) is found in $\boldsymbol{\tau}_m$, $m = 1, 2, 3$. In the following, we outline the algorithm

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on October 03,2024 at 10:14:35 UTC from IEEE Xplore. Restrictions apply.

that uses this information to iteratively build a room from identified reflectors. The progression of the algorithm relies on checking a set of conditions, detailed as follows.

1. **(TOA Condition)**: A pair of TOAs from separate RIRs corresponding to the same image source must lie within a time range of each other determined by the inter-microphone distance and the triangle inequality.
2. **(Combination Condition)**: A triplet of ellipsoids generated from respective TOAs must have at least one point of intersection in order to have a solution to (5).
3. **(Reflector Pair Condition)**: The combination of two *potential* reflectors denoted by $\tilde{\mathcal{N}} = \{\tilde{\mathfrak{n}}_1, \tilde{\mathfrak{n}}_2\}$ is said to constitute a feasible pair only if their corresponding potential second-order TOAs $\tilde{\tau}^{(2)}_{ij,m}$ appear in the respective RIRs within some temporal error tolerance η.
4. **(Room Condition)**: A set of *potential* reflectors $\tilde{\mathcal{N}} = \{\tilde{\mathfrak{n}}_1, \ldots, \tilde{\mathfrak{n}}_k\}$, $k \geq 2$, forms part of a feasible room if all pairs of reflectors in $\tilde{\mathcal{N}}$ meet the **Reflector Pair Condition**.

The **Combination Condition** can be checked without explicitly having to calculate the ellipsoid parameters using a simple EDM test similar to the method of [7] but using [23] to check the distance to the closest EDM in case of noise. This is not detailed here owing to space limitations. The hyper-parameter η in the **Reflector Pair Condition** can be selected based on prior knowledge of the statistics of the TOA estimation errors.

3.1. Algorithm overview

Without loss of generality, we let the RIR corresponding to microphone $m = 1$ contain the earliest arriving echo. The proposed method aims to iteratively identify and group potential first-order TOAs from $\boldsymbol{\tau}_1, \boldsymbol{\tau}_2$, and $\boldsymbol{\tau}_3$, in order to narrow down a set of reflectors $\hat{\mathcal{N}} = \{\hat{\mathfrak{n}}_1, \hat{\mathfrak{n}}_2, \ldots, \hat{\mathfrak{n}}_K\}$, corresponding to the room boundaries, where K is the number of boundaries. It does so by, in a greedy fashion, moving along the reference $\boldsymbol{\tau}_1$ and finding the most feasible combinations of TOAs across the remaining $\boldsymbol{\tau}_2$ and $\boldsymbol{\tau}_3$. Here, for an identified first-order TOA $\tau^{(1)}_{k,1} \in \boldsymbol{\tau}_1$, a set of candidate reflectors, denoted $\mathcal{S}_k$, is found using **Algorithm 2**. Each reflector contained in the set constitutes a candidate first-order image source consistent with all three RIRs in the sense of satisfying the **TOA & Combination Condition**. The reflectors in $\mathcal{S}_k$ are then appended to previously verified sets of reflectors, denoted $\mathcal{N}_p$, if they are consistent with $\mathcal{N}_p$ in terms of the **Room Condition** being satisfied. Thus, the sets $\mathcal{N}_p$ are expanded in such a way that all implied first- and second-order TOAs can be found in the RIRs $\boldsymbol{\tau}_m$, $m = 1, 2, 3$, and thereby constitute valid (partial) rooms. The collection of all such partial rooms, called a "House" set $\mathcal{H}^{(i)}$, is the main quantity being updated by the room geometry estimation algorithm, where i is the iteration index. If a particular partial room $\mathcal{N}_p$ cannot be updated, this implies that the room cannot be completed while being consistent with the TOAs. Consequently, such partial rooms are discarded. As the algorithm progresses, fewer and fewer partial rooms can be updated, leading to $\mathcal{H}^{(i)}$ becoming a singleton consisting of a single (complete) room with K reflectors and the algorithm terminates. The process is outlined in **Algorithm 1**. It may here be noted that the number of reflectors, K, is an input parameter. However, this is actually not an absolute requirement and was implemented for ease of describing the method. The algorithm could be allowed to iterate until either the next candidate reflector lies outside the enclosed convex space of the previous room, or a reflector that satisfies the requirements can no longer be found. Note also that for notational brevity,

Algorithm 1 Room geometry estimation and echo sorting

input $\mathcal{H}^{(1)}$, $\boldsymbol{\tau}_1$, $\boldsymbol{\tau}_2$, $\boldsymbol{\tau}_3$, K
output Estimated room $\hat{\mathcal{N}} = \{\hat{\mathfrak{n}}_1, \hat{\mathfrak{n}}_2, \ldots, \hat{\mathfrak{n}}_k\}$ for $2 \leq k \leq K$

for $i = 1, \ldots, (K-2)$ **do**
  $\mathcal{H}^{(i+1)} \leftarrow \emptyset$
  for $p = 1, \ldots, |\mathcal{H}^{(i)}|$ **do**
    $\hat{\boldsymbol{\tau}}_m \leftarrow \psi(\boldsymbol{\tau}_m, \mathcal{N}_p^{(i)})$, $m = 1, 2, 3$, with $\mathcal{N}_p^{(i)} \in \mathcal{H}^{(i)}$.
    Let $\hat{\tau}^{(1)}_{k,1}$ be the first TOA in $\hat{\boldsymbol{\tau}}_\mathbf{1}$.
    Compute $\mathcal{S}_p$ using **Algorithm 2**.
    for $j = 1, \ldots, |\mathcal{S}_p|$ **do**
      $\tilde{\mathcal{N}} \leftarrow \mathcal{N}_p^{(i)} \cup \{\mathfrak{n}_j\}$, where $\mathfrak{n}_j \in \mathcal{S}_p$.
      if $\tilde{\mathcal{N}}$ satisfies **Room Condition then**
        $\mathcal{H}^{(i+1)} \leftarrow \mathcal{H}^{(i+1)} \cup \left\{\tilde{\mathcal{N}}\right\}$
      end if
    end for
  end for
  if $\mathcal{H}^{(i+1)} = \emptyset$ **then**
    $\mathcal{H}^{(i+1)} \leftarrow \mathcal{H}^{(i)}$ and exit for-loop.
  end if
end for
if $|\mathcal{H}^{(i+1)}| = 1$ **then**
  $\hat{\mathcal{N}} \leftarrow \mathcal{H}^{(i+1)}$
else
  print "More than one feasible output room"
end if

Algorithm 2 Candidate reflector identification

Input TOA $\tau^{(1)}_{k,1}$, TOA sets $\hat{\boldsymbol{\tau}}_2$ and $\hat{\boldsymbol{\tau}}_3$
Output Reflector set S_k

Identify subsets $\tilde{\boldsymbol{\tau}}_2 \subset \hat{\boldsymbol{\tau}}_2$ and $\tilde{\boldsymbol{\tau}}_3 \subset \hat{\boldsymbol{\tau}}_3$ compatible with $\tau^{(1)}_{k,1}$ via the **TOA Condition**.
Construct the set $\mathcal{T} = \left\{(\tau^{(1)}_{k,1}, \tau_2, \tau_3) \mid (\tau_2, \tau_2) \in \tilde{\boldsymbol{\tau}}_2 \times \tilde{\boldsymbol{\tau}}_3\right\}$.
Discard all elements of $\mathcal{T}$ not satisfying the **Combination Condition**.
$S_k \leftarrow \emptyset$.
for each element of $\mathcal{T}$ **do**
  Compute the ellipsoid parameters $\mathbf{Q}_m$, for $m = 1, 2, 3$.
  Find all solutions $\mathfrak{n}$ to (5) and append solutions to $\mathcal{S}_k$.
end for
Return S_k.

we have introduced a pruning function $\psi : (\boldsymbol{\tau}_m, \mathcal{N}_p) \mapsto \hat{\boldsymbol{\tau}}_m$ where $\boldsymbol{\tau}_m$ is the full set of TOAs for microphone m, and $\mathcal{N}_p$ is a set of reflectors. Specifically, $\psi(\boldsymbol{\tau}_m, \mathcal{N}_p)$ returns the set of delays in $\boldsymbol{\tau}_m$ that do not correspond to first- or second-order TOAs as defined by the reflector set $\mathcal{N}_p$ according to equations (2) and (7), i.e., the TOAs already accounted for by the current estimate of the room. This also ensures that the next TOA in $\boldsymbol{\tau}_1$ can be assumed to be first-order.

As to initialize **Algorithm 1**, an input "House" set $\mathcal{H}^{(1)}$ needs to be found. This is done by using the fact that the first two echo TOAs in any RIR are first-order because at least two first-order echoes must exist before higher-order ones can appear. Therefore, we start by finding two sets of reflectors $\mathcal{S}_1$ and $\mathcal{S}_2$ using Algorithm 2. Taking combinations of the reflectors in each set, we obtain pairs of potential reflectors. Those that satisfy the **Room Condition** are stored as partial rooms $\mathcal{N}_p^{(1)}$ in $\mathcal{H}^{(1)}$.

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on October 03,2024 at 10:14:35 UTC from IEEE Xplore. Restrictions apply.

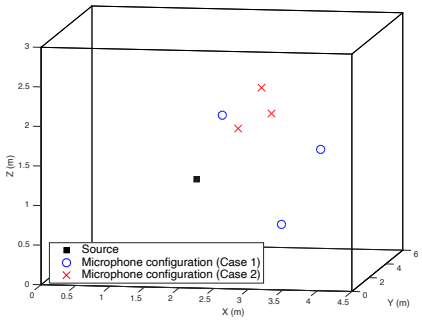


Fig. 2. Set-up for the numerical simulations illustrating the two microphone configurations corresponding to Case 1 and Case 2.

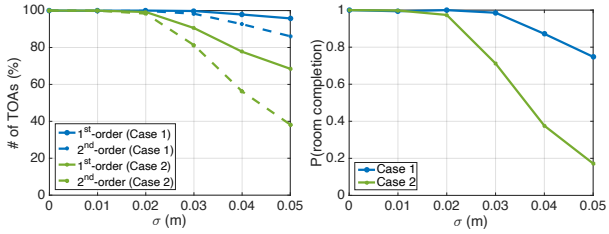
4. EXPERIMENTAL RESULTS

In this section, we demonstrate the performance of the proposed method using Monte Carlo simulations. In particular, we demonstrate the robustness of the method to the presence of errors on the estimated TOAs. Throughout, we consider a box-shaped room ($K = 6$) of dimensions 4.5 m x 6 m x 3 m, with the source placed centrally in the room at $\mathbf{s} = [2, 2, 1.2]^T$. The positions of the three microphones were chosen randomly in the region of the room bounded by $[2 : 4.5, 2 : 6, 0.5 : 2.5]$ m, conditioned on minimal and maximal inter-microphone distances according to two scenarios. In the first scenario (Case 1), the distance between any microphone pair is in the interval [1.5, 2] m, whereas in the second (Case 2), the distance range is reduced to the interval [0.55, 0.75] m. The resulting set-ups are illustrated in Fig. 2. Note that this is prior to all relative positions being shifted such that the source is at the origin. In each simulation, we generate sets of TOAs up to the second-order using the ISM [5] and perturb the distances from the underlying image sources by independent Gaussian noise with variance σ^2. That is, each estimated (implied) distance from an image source is given by $\tilde{d} = d + \xi$ where d is the nominal (noise-free) distance and $\xi \sim \mathcal{N}(0, \sigma^2)$. Varying σ on the interval [0,0.05], we for each considered value of σ perform 500 Monte Carlo simulations. The hyper-parameter for the second-order tolerance was selected as $\eta = \eta(\sigma) = 2\sigma + c/f_s$ where $f_s = 44.1$ kHz is the sampling frequency considered.

The percentage of correctly identified and grouped first- and second-order echoes, as a function of the noise standard deviation, is shown in Fig.3 (a). Fig.3 (b) displays the empirical probability of the proposed method converging to a room with all six reflectors having been correctly identified. As can be seen, the probability of correctly labeling echoes is essentially 1 for small standard deviations on the TOA estimates, and deteriorates more slowly in the case where microphones are well-separated (Case 1 vs. Case 2). This is also reflected in the probability of obtaining a complete room. Furthermore, conditioned on a room being correctly identified, Fig. 4 displays the average distance and angular errors for the estimated reflectors, corresponding to how well the positions of the room boundaries are estimated. These errors are defined as

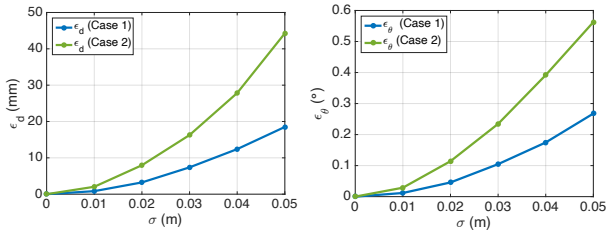
$$\epsilon_d(\mathbf{n}_k, \hat{\mathbf{n}}_k) = |d_k - \hat{d}_k| \quad , \quad \epsilon_\theta(\mathbf{n}_k, \hat{\mathbf{n}}_k) = \text{atan2}\left(\frac{\|\mathbf{n}_k \times \hat{\mathbf{n}}_k\|}{\mathbf{n}_k^T \hat{\mathbf{n}}_k}\right),$$

where d_k and $\hat{d}_k$ are the orthogonal distances between the source $\mathbf{s}$ and respectively, the actual reflector $\mathbf{n}_k$ and estimated reflector $\hat{\mathbf{n}}_k$. Here, atan2 is the four-quadrant inverse tangent function with range limited from $[0, 180°]$. In terms of computational speed, the greedy



(a) Echo labeling (b) Room completion

Fig. 3. (a) Percentage of total number of echoes that have been correctly labeled, as a function of the standard deviation σ. (b) Probability of converging to one correct complete room (6 reflectors) as a function of the standard deviation σ.



(a) Average distance error ϵ_d (b) Average angular error ϵ_θ

Fig. 4. Reflector localization error as a function of the noise standard deviation σ

structure of the proposed method efficiently narrows down the number of possible room configurations, resulting in rapidly being able to reconstruct a room. For Case 1, the average run-time was 7 seconds, whereas for Case 2, it was 1 second when implemented on a 2,6 GHz 6-Core Intel Core i7 processor running Matlab 2015b. As noted earlier, there are no widely available reference methods applicable to the scenario considered herein, i.e., three microphones and unknown echo orders. However, for the method in [14] which requires more microphones, the reconstruction of a similar size room needed around 200 seconds. One of the main reasons our method breaks down in the case of larger TOA errors is that, in the process of removing higher-order image sources to reduce the number of echo combinations at each iterate and ensure that the next echo in the reference is first-order, a first-order echo itself may accidentally be removed. This can somewhat be mitigated by placing microphones further apart (see Case 1). In future work, the algorithm will be supplemented with TOA error correction so as to reduce the chance of erroneously removing echoes.

5. CONCLUSIONS

In this work, we have proposed a method for room geometry estimation in 3-D based on the availability of measured acoustic RIRs. The method requires only three RIRs, and does not rely on prior knowledge of the echoes' order. Instead, it solves the problem of echo labeling jointly with the room reconstruction. As shown in numerical simulations, the method displays robustness to TOA estimation errors, particularly for cases in which the microphones are well-separated. The proposed method is also attractive from a computational point of view as it avoids the combinatorial explosion of the ELP problem by a greedy approach to room reconstruction. However, further experimental validation using real RIR measurements is required, and the method still needs to be adapted in order to take into account real-world problems such as missing TOAs.

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on October 03,2024 at 10:14:35 UTC from IEEE Xplore. Restrictions apply.

6. REFERENCES

[1] Lauri Savioja and U Peter Svensson, "Overview of geometrical room acoustic modeling techniques," *The Journal of the Acoustical Society of America*, vol. 138, no. 2, pp. 708–730, 2015.

[2] Luca Remaggi, Philip JB Jackson, Philip Coleman, and Wenwu Wang, "Room boundary estimation from acoustic room impulse responses," in *2014 Sensor Signal Processing for Defence (SSPD)*. IEEE, 2014, pp. 1–5.

[3] Michael Lovedee-Turner and Damian Murphy, "Three-dimensional reflector localisation and room geometry estimation using a spherical microphone array," *The Journal of the Acoustical Society of America*, vol. 146, no. 5, pp. 3339–3352, 2019.

[4] Tilak Rajapaksha, Xiaojun Qiu, Eva Cheng, and Ian Burnett, "Geometrical room geometry estimation from room impulse responses," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 331–335.

[5] Jont B Allen and David A Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.

[6] Fabio Antonacci, Jason Filos, Mark RP Thomas, Emanuël AP Habets, Augusto Sarti, Patrick A Naylor, and Stefano Tubaro, "Inference of room geometry from acoustic impulse responses," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 10, pp. 2683–2695, 2012.

[7] Ivan Dokmanić, Reza Parhizkar, Andreas Walther, Yue M Lu, and Martin Vetterli, "Acoustic echoes reveal room shape," *Proceedings of the National Academy of Sciences*, vol. 110, no. 30, pp. 12186–12191, 2013.

[8] E Nastasia, Fabio Antonacci, Augusto Sarti, and Stefano Tubaro, "Localization of planar acoustic reflectors through emission of controlled stimuli," in *2011 19th European Signal Processing Conference*. IEEE, 2011, pp. 156–160.

[9] Luca Remaggi, Hansung Kim, Philip JB Jackson, Filippo Maria Fazi, and Adrian Hilton, "Acoustic reflector localization and classification," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 201–205.

[10] Sakari Tervo and Timo Tossavainen, "3d room geometry estimation from measured impulse responses," in *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2012, pp. 513–516.

[11] Jason Filos, Emanuël AP Habets, and Patrick A Naylor, "A two-step approach to blindly infer room geometries," in *Proc. Intl. Workshop Acoust. Echo Noise Control (IWAENC), Tel Aviv, Israel*. Citeseer, 2010.

[12] Ivan Dokmanić, Yue M. Lu, and Martin Vetterli, "Can one hear the shape of a room: The 2-d polygonal case," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011, pp. 321–324.

[13] Wangyang Yu and W Bastiaan Kleijn, "Room acoustical parameter estimation from room impulse responses using deep neural networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 436–447, 2020.

[14] Sooyeon Park and Jung-Woo Choi, "Iterative echo labeling algorithm with convex hull expansion for room geometry estimation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 1463–1478, 2021.

[15] Ingmar Jager, Richard Heusdens, and Nikolay D Gaubitch, "Room geometry estimation from acoustic echoes using graph-based echo labeling," in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2016, pp. 1–5.

[16] Mario Coutino, Martin Bo Møller, Jesper Kjær Nielsen, and Richard Heusdens, "Greedy alternative for room geometry estimation from acoustic echoes: A subspace-based method," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017, pp. 366–370.

[17] Marco Crocco, Andrea Trucco, and Alessio Del Bue, "Uncalibrated 3d room geometry estimation from sound impulse responses," *Journal of the Franklin Institute*, vol. 354, no. 18, pp. 8678–8709, 2017.

[18] Joseph B Keller, "The scope of the image method," *Communications on pure and applied mathematics*, vol. 6, no. 4, pp. 505–512, 1953.

[19] Ian J Kelly and Francis M Boland, "Detecting arrivals in room impulse responses with dynamic time warping," *IEEE/ACM transactions on audio, speech, and language processing*, vol. 22, no. 7, pp. 1139–1147, 2014.

[20] Jingdong Chen, Jacob Benesty, and Yiteng Huang, "Time delay estimation in room acoustic environments: An overview," *EURASIP Journal on Advances in Signal Processing*, vol. 2006, pp. 1–19, 2006.

[21] Hannes Rosseel and Toon van Waterschoot, "Improved acoustic source localization by time delay estimation with subsample accuracy," in *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*. IEEE, 2021, pp. 1–8.

[22] Richard Hartley and Andrew Zisserman, *Multiple view geometry in computer vision*, Cambridge university press, 2003.

[23] Norbert Gaffke and Rudolf Mathar, "A cyclic projection algorithm via duality," *Metrika*, vol. 36, no. 1, pp. 29–54, 1989.

Authorized licensed use limited to: KU Leuven Libraries. Downloaded on October 03,2024 at 10:14:35 UTC from IEEE Xplore. Restrictions apply.