# Acoustic echoes reveal room shape

Ivan Dokmanić[a,1], Reza Parhizkar[a], Andreas Walther[a], Yue M. Lu[b], and Martin Vetterli[a]

[a]Audiovisual Communications Laboratory (LCAV), School of Computer and Communication Sciences, Ecole Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland; and [b]Signals, Information, and Networks Group (SING), School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138

Edited by David L. Donoho, Stanford University, Stanford, CA, and approved May 17, 2013 (received for review January 4, 2013)

**Imagine that you are blindfolded inside an unknown room. You snap your fingers and listen to the room's response. Can you hear the shape of the room? Some people can do it naturally, but can we design computer algorithms that hear rooms? We show how to compute the shape of a convex polyhedral room from its response to a known sound, recorded by a few microphones. Geometric relationships between the arrival times of echoes enable us to "blindfoldedly" estimate the room geometry. This is achieved by exploiting the properties of Euclidean distance matrices. Furthermore, we show that under mild conditions, first-order echoes provide a unique description of convex polyhedral rooms. Our algorithm starts from the recorded impulse responses and proceeds by learning the correct assignment of echoes to walls. In contrast to earlier methods, the proposed algorithm reconstructs the full 3D geometry of the room from a single sound emission, and with an arbitrary geometry of the microphone array. As long as the microphones can hear the echoes, we can position them as we want. Besides answering a basic question about the inverse problem of room acoustics, our results find applications in areas such as architectural acoustics, indoor localization, virtual reality, and audio forensics.**

room geometry | geometry reconstruction | echo sorting | image sources

In a famous paper (1), Mark Kac asks the question "Can one hear the shape of a drum?" More concretely, he asks whether two membranes of different shapes necessarily resonate at different frequencies.* This problem is related to a question in astrophysics (2), and the answer turns out to be negative: Using tools from group representation theory, Gordon et al. (3, 4) presented several elegantly constructed counterexamples, including the two polygonal drum shapes shown in Fig. 1. Although geometrically distinct, the two drums have the same resonant frequencies.†

In this work, we ask a similar question about rooms. Assume you are blindfolded inside a room; you snap your fingers and listen to echoes. Can you hear the shape of the room? Intuitively, and for simple room shapes, we know that this is possible. A shoebox room, for example, has well-defined modes, from which we can derive its size. However, the question is challenging in more general cases, even if we presume that the room impulse response (RIR) contains an arbitrarily long set of echoes (assuming an ideal, noiseless measurement) that should specify the room geometry.

It might appear that Kac's problem and the question we pose are equivalent. This is not the case, for the sound of a drum depends on more than its set of resonant frequencies (eigenvalues)—it also depends on its resonant modes (eigenvectors). In the paper "Drums that sound the same" (5), Chapman explains how to construct drums of different shapes with matching resonant frequencies. Still, these drums would hardly sound the same if hit with a drumstick. They share the resonant frequencies, but the impulse responses are different. Even a single drum struck at different points sounds differently. Fig. 1 shows this clearly.

Certain animals can indeed "hear" their environment. Bats, dolphins, and some birds probe the environment by emitting sounds and then use echoes to navigate. It is remarkable to note that there are people that can do the same, or better. Daniel Kish produces clicks with his mouth, and uses echoes to learn the shape, distance, and density of objects around him (6). The main cues for human echolocators are early reflections. Our computer algorithms also use early reflections to calculate shapes of rooms.

Many applications benefit from knowing the room geometry. Indoor sound-source localization is usually considered difficult, because the reflections are difficult to predict and they masquerade as sources. However, in rooms one can do localization more accurately than in free-field if the room geometry (7–10) is known. In teleconferencing, auralization, and virtual reality, one often needs to compensate the room influence or create an illusion of a specific room. The success of these tasks largely depends on the accurate modeling of the early reflections (11), which in turn requires the knowledge of the wall locations.

We show how to reconstruct a convex polyhedral room from a few impulse responses. Our method relies on learning from which wall a particular echo originates. There are two challenges with this approach: First, it is difficult to extract echoes from RIRs; and second, the microphones receive echoes from walls in different orders. Our main contribution is an algorithm that selects the "correct" combinations of echoes, specifically those that actually correspond to walls. The need for assigning echoes to walls arises from the omnidirectionality of the source and the receivers.

There have been several attempts in estimating the room geometry from RIRs (12–14). In (13), the problem is formulated in 2D, and the authors take advantage of multiple source locations to estimate the geometry. In (14) the authors address the problem by $\ell_1$-regularized template matching with a premeasured dictionary of impulse responses. Their approach requires measuring a very large matrix of impulse responses for a fixed-source–receiver geometry. The authors in (15) propose a 3D room reconstruction method by assuming that the array is small enough so that there is no need to assign echoes to walls. They use sparse RIRs obtained by directing the loudspeaker to many orientations and processing the obtained responses. In contrast, our method works with arbitrary measurement geometries. Furthermore, we prove that the first-order echoes provide a unique description of the room for almost all setups. A subspace-based formulation allows us to use the minimal number of microphones (four microphones in 3D). It is impossible to further reduce the number of microphones, unless we consider higher-order echoes, as attempted in (12). However, the arrival times of higher-order echoes are often challenging to obtain and delicate to use, both for theoretical and practical reasons. Therefore, in the proposed method, we choose to use more than one microphone, avoiding the need for higher-order echoes.

In addition to theoretical analysis, we validate the results experimentally by hearing rooms on Ecole Polytechnique Fédérale

Author contributions: I.D. and M.V. designed research; I.D., R.P., A.W., and Y.M.L. performed research; I.D. and R.P. analyzed data; and I.D. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

[1]To whom correspondence should be addressed. E-mail: ivan.dokmanic@epfl.ch.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1221464110/-/DCSupplemental.

*Resonant frequencies correspond to the eigenvalues of a Laplacian on a 2D domain.

†More details about this counterexample are given in the *SI Text*.

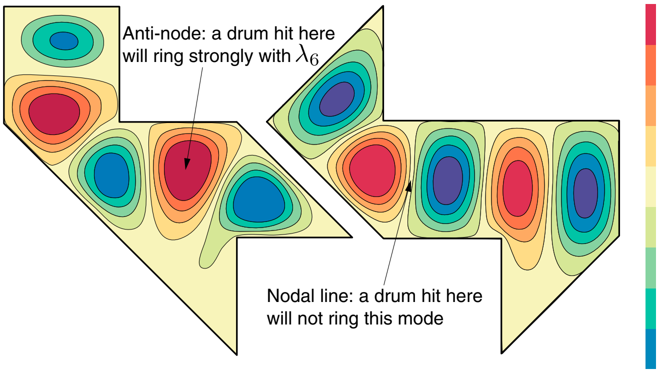


**Fig. 1.** Figure shows two isospectral drums (16). Although geometrically distinct, these drums have the same resonant frequencies. The standing waves corresponding to the eigenvalue $\lambda_6$ are shown for both drums. It is clear that this mode will be excited with different amplitudes, depending on where we hit the drum. Extremes are nodes and anti-nodes.

de Lausanne (EPFL) campus. Moreover, by running it in a portal of the Lausanne cathedral, we show that the algorithm still gives useful output even when the room thoroughly violates the assumptions of it being a convex polyhedron.

## Modeling

We consider the room to be a $K$-faced convex polyhedron. We work in 3D, but the results extend to arbitrary dimensionalities (2D is interesting for some applications). Sound propagation in a room is described by a family of RIRs. An RIR models the channel between a fixed source and a fixed receiver. It contains the direct path and the reflections. Ideally, it is a train of pulses, each corresponding to an echo. For the $m$th microphone it is given by

$$h_m(t) = \sum_i \alpha_{m,i}\delta(t - \tau_{m,i}). \quad [1]$$

Microphones hear the convolution of the emitted sound with the corresponding RIR, $y_m = x * h_m = \int x(s)h_m(\cdot - s)ds$. By measuring the impulse responses we access the propagation times $\tau_{m,i}$, and these can be linked to the room geometry by the image source (IS) model (17, 18). According to the IS model, we can replace reflections by virtual sources. As illustrated in Fig. 2, virtual sources are mirror images of the true sources across the corresponding reflecting walls. From the figure, the image $\tilde{\mathbf{s}}_i$ of the source $\mathbf{s}$ with respect to the $i$th wall is computed as

$$\tilde{\mathbf{s}}_i = \mathbf{s} + 2\langle \mathbf{p}_i - \mathbf{s}, \mathbf{n}_i \rangle \mathbf{n}_i, \quad [2]$$

where $\mathbf{n}_i$ is the unit normal, and $\mathbf{p}_i$ any point belonging to the $i$th wall. The time of arrival (TOA) of the echo from the $i$th wall is $t_i = \|\tilde{\mathbf{s}}_i - \mathbf{r}\|/c$, where $c$ is the speed of sound.

In a convex room with a known source, knowing the image sources is equivalent to knowing the walls—we can search for points instead of searching for walls. The challenge is that the distances are unlabeled: It might happen that the $k$th peak in the RIR from microphone 1 and the $k$th peak in the RIR from microphone 2 come from different walls. This is illustrated in Figs. 3 and 4. Thus, we have to address the problem of echo labeling. The loudspeaker position need not be known. We can estimate it from the direct sound using either TOA measurements, or differences of TOAs if the loudspeaker is not synchronized with the microphones (19–21).

In practice, having a method to find good combinations of echoes is far more important than only sorting correctly selected echoes. Impulse responses contain peaks that do not correspond to any wall. These spurious peaks can be introduced by noise, nonlinearities, and other imperfections in the measurement system. We find that a good strategy is to select a number of peaks greater than the number of walls and then to prune the selection. Furthermore, some second-order echoes might arrive before some first-order ones. The image sources corresponding to second-order or higher-order echoes (e.g., Fig. 2) will be estimated as any other image source. However, because we can express a second-order image source in terms of the first-order ones as

$$\tilde{\mathbf{s}}_{ij} = \tilde{\mathbf{s}}_i + 2\left\langle \mathbf{p}_j - \tilde{\mathbf{s}}_i, \mathbf{n}_j \right\rangle \mathbf{n}_j, \quad [3]$$

and

$$\left\| \mathbf{s} - \tilde{\mathbf{s}}_{ij} \right\| = \left\| \tilde{\mathbf{s}}_i - \tilde{\mathbf{s}}_j \right\|, \quad [4]$$

we can eliminate it during postprocessing by testing the above two expressions.

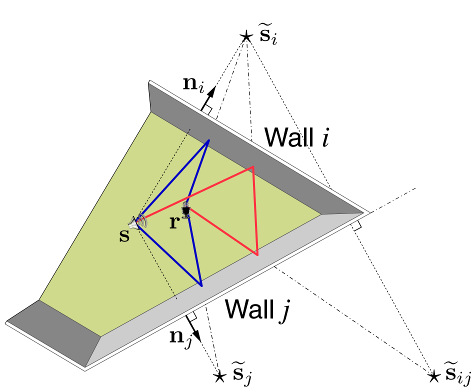


**Fig. 2.** Illustration of the image source model for first- and second-order echoes. Vector $\mathbf{n}_i$ is the outward-pointing unit normal associated with the $i$th wall. Stars denote the image sources, and $\tilde{\mathbf{s}}_{ij}$ is the image source corresponding to the second-order echo. Sound rays corresponding to first reflections are shown in blue, and the ray corresponding to the second-order reflection is shown in red.

## Echo Labeling

The purpose of echo labeling is twofold. First, it serves to remove the "ghost" echoes (that do not correspond to walls) detected at the peak-picking stage. Second, it determines the correct assignment between the remaining echoes and the walls. We propose two methods for recognizing correct echo combinations. The first one is based on the properties of Euclidean distance matrices($\mathbb{EDM}$), and the second one on a simple linear subspace condition.

**EDM-Based Approach.** Consider a room with a loudspeaker and an array of $M$ microphones positioned so that they hear the first-order echoes (we typically use $M = 5$). Denote the receiver positions by $\mathbf{r}_1, \ldots, \mathbf{r}_M$, $\mathbf{r}_m \in \mathbb{R}^3$ and the source position by $\mathbf{s} \in \mathbb{R}^3$. The described setup is illustrated in Fig. 5. We explain the $\mathbb{EDM}$-based echo sorting with reference to this figure. Let $\mathbf{D} \in \mathbb{R}^{M \times M}$ be a matrix whose entries are squared distances between microphones, $\mathbf{D}[i,j] = \|\mathbf{r}_i - \mathbf{r}_j\|_2^2$, $1 \le i,j \le M$. Here, $\mathbf{D}$ is an $\mathbb{EDM}$ corresponding to the microphone setup. It is symmetric with a zero diagonal and positive off-diagonal entries.

If the loudspeaker emits a sound, each microphone receives the direct sound and $K$ first-order echoes corresponding to the $K$ walls. The arrival times of the received echoes are proportional to the distances between image sources and microphones. As

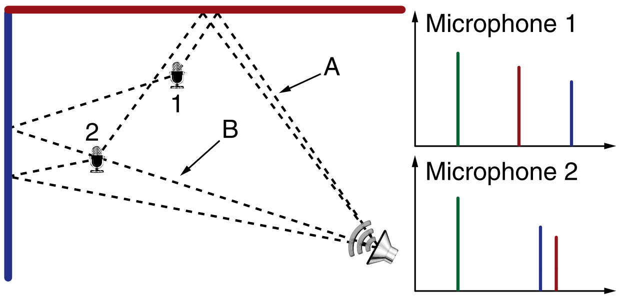


**Fig. 3.** Illustration of echo swapping. Microphone 1 hears the echo from the red wall before hearing the echo from the blue wall, because path A is shorter than path B. The opposite happens for microphone 2.

already discussed, we face a labeling problem as we do not know which wall generated which echo. This problem is illustrated in Fig. 3 for two walls and in Fig. 4 for the whole room. Simple heuristics, such as grouping the closest pulses or using the ordinal number of a pulse, have limited applicability, especially with larger distances between microphones. That these criteria fail is evident from Fig. 4.

We propose a solution based on the properties of $\mathbb{EDM}$s. The loudspeaker and the microphones are—to a good approximation—points in space, so their pairwise distances form an $\mathbb{EDM}$. We can exploit the rank property: An $\mathbb{EDM}$ corresponding to a point set in $\mathbb{R}^n$ has rank at most $(n+2)$ (22). Thus, in 3D, its rank is at most 5. We start from a known point set (the microphones) and want to add another point—an image source. This requires adding a row and a column to $\mathbf{D}$, listing squared distances between the microphones and the image source. We extract the list of candidate distances from the RIRs, but some of them might not correspond to an image source; and for those that do correspond, we do not know to which one. Consider again the setup in Fig. 5. Microphone 1 hears echoes from all of the walls, and we augment $\mathbf{D}$ by choosing different echo combinations. Two possible augmentations are shown. Here, $\mathbf{D}_{\mathbf{aug},1}$ is a plausible augmentation of $\mathbf{D}$ because all of the distances correspond to a single image source, and they appear in the correct order. This matrix passes the rank test, or more specifically, it is an $\mathbb{EDM}$. The second matrix, $\mathbf{D}_{\mathbf{aug},2}$, is a result of an incorrect echo assignment, as it contains entries coming from different walls. A priori, we cannot tell whether the red echo comes from wall 1 or from wall 2. It is simply an unlabeled peak in the RIR recorded by microphone 1. However, the augmented matrix $\mathbf{D}_{\mathbf{aug},2}$ does not pass the rank test, so we conclude that the corresponding combination of echoes is not correct.

To summarize, wrong assignments lead to augmentations of $\mathbf{D}$ that are not $\mathbb{EDM}$s. In particular, these augmentations do not have the correct rank. As it is very unlikely (as will be made precise later) for incorrect combinations of echoes to form an $\mathbb{EDM}$, we have designed a tool to detect correct echo combinations.

More formally, let $\mathbf{e}_m$ list the candidate distances computed from the RIR recorded by the $m$th microphone. We proceed by augmenting the matrix $\mathbf{D}$ with a combination of $M$ unlabeled squared distances $\mathbf{d}_{(i_1,\ldots,i_M)}$ to get $\mathbf{D}_{\mathbf{aug}}$,

$$\mathbf{D}_{\mathrm{aug}}\left(\mathbf{d}_{(i_1,\ldots,i_M)}\right)=\begin{bmatrix}\mathbf{D} & \mathbf{d}_{(i_1,\ldots,i_M)}\\ \mathbf{d}^{\top}_{(i_1,\ldots,i_M)} & 0\end{bmatrix}. \qquad [5]$$

The column vector $\mathbf{d}_{(i_1,\ldots,i_M)}$ is constructed as

$$\mathbf{d}_{(i_1,\ldots,i_M)}[m]=\mathbf{e}_m^2[i_m], \qquad [6]$$

with $i_m\in\{1,\ldots,\mathrm{length}(\mathbf{e}_m)\}$. In words, we construct a candidate combination of echoes $\mathbf{d}$ by selecting one echo from each microphone. Note that $\mathrm{length}(\mathbf{e}_m)\neq\mathrm{length}(\mathbf{e}_n)$ for $m\neq n$ in general. That is, we can pick a different number of echoes from different microphones. We interpret $\mathbf{D}_{\mathbf{aug}}$ as an object encoding a particular selection of echoes $\mathbf{d}$.

One might think of $\mathbb{EDM}$ as a mold. It is very much like Cinderella's glass slipper: If you can snugly fit a tuple of echoes in it, then they must be the right echoes. This is the key observation: If $\mathrm{rank}(\mathbf{D}_{\mathrm{aug}})<6$ or more specifically $\mathbf{D}_{\mathrm{aug}}$ verifies the $\mathbb{EDM}$ property, then the selected combination of echoes corresponds to an image source, or equivalently to a wall. Even if this approach requires testing all of the echo combinations, in practical cases the number of combinations is small enough that this does not present a problem.

**Subspace-Based Approach.** An alternative method to obtain correct echo combinations is based on a simple linear condition. Note that we can always choose the origin of the coordinate system so that

$$\sum_{m=1}^{M}\mathbf{r}_m=0. \qquad [7]$$

Let $\tilde{\mathbf{s}}_k$ be the location vector of the image source with respect to wall $k$. Then, up to a permutation, we receive at the $m$th microphone the squared distance information,

$$y_{k,m}\overset{\mathrm{def}}{=}\|\tilde{\mathbf{s}}_k-\mathbf{r}_m\|^2=\|\tilde{\mathbf{s}}_k\|^2-2\,\tilde{\mathbf{s}}_k^{\top}\mathbf{r}_m+\|\mathbf{r}_m\|^2. \qquad [8]$$

Define further $\tilde{y}_{k,m}\overset{\mathrm{def}}{=}-\frac{1}{2}(y_{k,m}-\|\mathbf{r}_m\|^2)=\mathbf{r}_m^{\top}\tilde{\mathbf{s}}_k-\frac{1}{2}\|\tilde{\mathbf{s}}_k\|^2$. We have in vector form

$$\begin{bmatrix}\tilde{y}_{k,1}\\ \tilde{y}_{k,2}\\ \vdots\\ \tilde{y}_{k,M}\end{bmatrix}=\begin{bmatrix}\mathbf{r}_1^{\top} & -\frac{1}{2}\\ \mathbf{r}_2^{\top} & -\frac{1}{2}\\ \vdots & \vdots\\ \mathbf{r}_M^{\top} & -\frac{1}{2}\end{bmatrix}\begin{bmatrix}\tilde{\mathbf{s}}_k\\ \|\tilde{\mathbf{s}}_k\|^2\end{bmatrix},\quad\text{or}\quad\tilde{\mathbf{y}}_k=\mathbf{R}\tilde{\mathbf{u}}_k. \qquad [9]$$

Thanks to the condition **7**, we have that

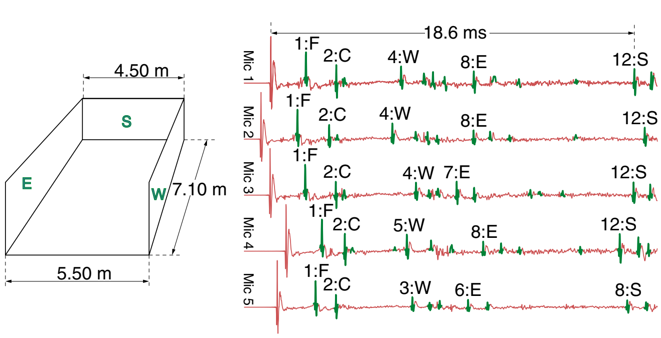


**Fig. 4.** Actual room impulses responses acquired in a room sketched on the *Left* (see experiments for more details). First peak corresponds to direct propagation. Detected echoes are highlighted in green. Annotations above the peaks indicate the ordinal number of the peak, and the wall to which it corresponds (south, north, east, west, floor, and ceiling). We can see that the ordinal number of the W-peak changes from one impulse response to another (similarly for E and S). For larger microphone arrays this effect becomes more dramatic. We also see that some peaks do not correspond to walls. Our algorithm successfully groups peaks corresponding to the same wall, and disregards irrelevant peaks.

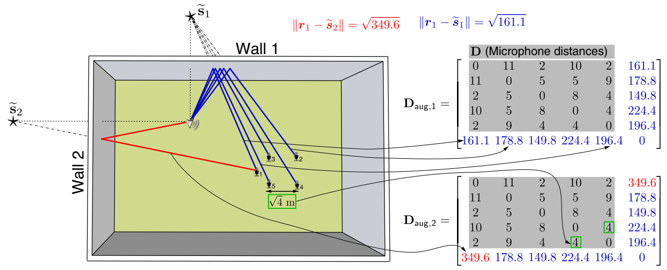


**Fig. 5.** Illustration of $\mathbb{EDM}$-based echo sorting. Microphones receive the echoes from all of the walls, and we aim to identify echoes coming from a single wall. We select one echo from each microphone and use these echoes to augment the $\mathbb{EDM}$ of the microphone setup, **D**. If all of the selected echoes come from the same wall, the augmented matrix is an $\mathbb{EDM}$ as well. In the figure, $\mathbf{D}_{\text{aug},1}$ is an $\mathbb{EDM}$ because it contains the distances to a single point $\tilde{\mathbf{s}}_1$; $\mathbf{D}_{\text{aug},2}$ contains a wrong distance (shown in red) for microphone 1, so it is not an $\mathbb{EDM}$. For aesthetic reasons the distances are specified to a single decimal place. Full precision entries are given in the *SI Text*.

$$1^\top \tilde{\mathbf{y}}_k = -\frac{M}{2}\|\tilde{\mathbf{s}}_k\|^2 \quad \text{or} \quad \|\tilde{\mathbf{s}}_k\|^2 = -\frac{2}{M}\sum_{m=1}^{M}\tilde{y}_{k,m}. \qquad [10]$$

The image source is found as

$$\tilde{\mathbf{s}}_k = \mathbf{S}\tilde{\mathbf{y}}_k, \qquad [11]$$

where **S** is a matrix satisfying

$$\mathbf{SR} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}. \qquad [12]$$

These two conditions characterize the distance information. In practice, it is sufficient to verify the linear constraint

$$\tilde{\mathbf{y}}_k \in \text{range}(\mathbf{R}), \qquad [13]$$

where range$(\mathbf{R})$ is a proper subspace when $M \geq 5$. However, note that we can use the nonlinear condition **10** even if $M = 4$.

**Uniqueness.** Can we guarantee that only one room corresponds to the collected first-order echoes? To answer this, we first define the set of "good" rooms in which our algorithm can be applied. The algorithm relies on the knowledge of first-order echoes, so we require that the microphones hear them. This defines a good room, which is in fact a combination of the room geometry and the microphone array/loudspeaker location.

***Definition 1:*** (Feasibility). Given a room $\mathcal{R}$ and a loudspeaker position **s**, we say that the point $\mathbf{x} \in \mathcal{R}$ is feasible if a microphone placed at **x** receives all the first-order echoes of a pulse emitted from **s**.

Our argument is probabilistic: The set of vectors **d** such that rank $\mathbf{D}_{\text{aug}} = 5$ has measure zero in $\mathbb{R}^5$. Analogously, in the subspace formulation, range$(\mathbf{R})$ is a proper subspace of $\mathbb{R}^5$ thus having measure zero. To use four microphones, observe that the same is true for the set of vectors satisfying **[10]** in $\mathbb{R}^4$. These observations, along with some technical details, enable us to state the uniqueness result.[‡]

**Theorem 1.** *Consider a room with a loudspeaker and $M \geq 4$ microphones placed uniformly at random inside the feasible region.* Then the unlabeled set of first-order echoes uniquely specifies the room with probability 1. In other words, almost surely exactly one assignment of first-order echoes to walls describes a room.

This means that we can reconstruct any convex polyhedral room if the microphones are in the feasible region. A similar result could be stated by randomizing the room instead of the microphone setup, but that would require us to go through the inconvenience of generating a random convex room. In the following, we concentrate on the $\mathbb{EDM}$ criterion, as it performs better in experiments.

## Practical Algorithm

In practice, we face different sources of uncertainty. One such source is the way we measure the distances between microphones. We can try to reduce this error by calibrating the array, but we find the proposed schemes to be very stable with respect to uncertainties in array calibration. Additional sources of error are the finite sampling rate and the limited precision of peak-picking algorithms. These are partly caused by unknown loudspeaker and microphone impulse responses, and general imperfections in RIR measurement. They can be mitigated with higher sampling frequencies and more sophisticated time-of-arrival estimation algorithms. At any rate, testing the rank of $\mathbf{D}_{\text{aug}}$ is not a way to go in the presence of measurement uncertainties. The solution is to measure how close $\mathbf{D}_{\text{aug}}$ is to an $\mathbb{EDM}$. We can consider different constructions:

(*i*) Heuristics based on the singular values of $\mathbf{D}_{\text{aug}}$;

(*ii*) distance of $\tilde{\mathbf{y}}_k$ from range$(\mathbf{R})$ (Eq. **13**);

(*iii*) nonlinear norm condition **10**; and

(*iv*) distance between $\mathbf{D}_{\text{aug}}$ and the closest $\mathbb{EDM}$.

The approach based on the singular values of $\mathbf{D}_{\text{aug}}$ captures only the rank requirement on the matrix. However, the requirement that $\mathbf{D}_{\text{aug}}$ be an $\mathbb{EDM}$ brings in many additional subtle dependencies between its elements. For instance, we have that (23)

$$\mathbf{D_{aug}} \in \mathbb{EDM} \Leftrightarrow \left(\mathbf{I} - \frac{1}{M+1}11^\top\right)\mathbf{D_{aug}}\left(\mathbf{I} - \frac{1}{M+1}11^\top\right) \preceq 0. \qquad [14]$$

Unfortunately **[14]**, does not allow us to specify the ambient dimension of the point set. Imposing this constraint leads to even more dependencies between the matrix elements, and the resulting set of matrices is no longer a cone (it is actually not convex anymore). Nevertheless, we can apply the family of algorithms used in multidimensional scaling (MDS) (23) to find the closest $\mathbb{EDM}$ between the points in a fixed ambient dimension.

**Multidimensional Scaling.** As discussed, in the presence of noise the rank test on $\mathbf{D}_{\text{aug}}$ is inadequate. A good way of dealing with this nuisance (as verified through experiments) is to measure how close $\mathbf{D}_{\text{aug}}$ is to an $\mathbb{EDM}$. To this end we use MDS to construct the point set in a given dimension (3D) that produces the $\mathbb{EDM}$ "closest" to $\mathbf{D}_{\text{aug}}$. MDS was originally proposed in psychometrics (24) for data visualization. Many adaptations of the method have been proposed for sensor localization. We use the so-called "s-stress" criterion (25). Given an observed noisy matrix $\tilde{\mathbf{D}}_{\text{aug}}$, s-stress$(\tilde{\mathbf{D}}_{\text{aug}})$ is the value of the following optimization program,

$$\min.\sum_{i,j}\left(\mathbf{D_{aug}}[i,j] - \tilde{\mathbf{D}}_{\mathbf{aug}}[i,j]\right)^2 \text{ s.t. } \mathbf{D_{aug}} \in \mathbb{EDM}^3. \qquad [15]$$

By $\mathbb{EDM}^3$ we denote the set of $\mathbb{EDM}$s generated by point sets in $\mathbb{R}^3$. We say that s-stress$(\tilde{\mathbf{D}}_{\text{aug}})$ is the score of the matrix $\tilde{\mathbf{D}}_{\text{aug}}$, and use it to assess the likelihood that a combination of echoes

[‡]Proof is given in *SI Text*.

corresponds to a wall. A method for solving **[15]** is described in the *SI Text*.

**Reconstruction Algorithm.** Combining the described ingredients, we design an algorithm for estimating the shape of a room. The algorithm takes as input the arrival times of echoes at different microphones (computed from RIRs). For every combination of echoes, it computes the score using the criterion of choice. We specialize to constructing the matrix $\mathbf{D}_{\text{aug}}$ as in (5) and computing the s-stress score. For the highest ranked combinations of echoes, it computes the image-source locations. We use an additional step to eliminate ghost echoes, second-order image sources, and other impossible solutions. Note that we do not discuss peak picking (selecting peaks from RIRs) in this work. The algorithm is summarized as

*i*) For every $\mathbf{d}_{(i_1,\ldots,i_M)}$, $\text{score}[\mathbf{d}_{(i_1,\ldots,i_M)}] \leftarrow \text{s-stress}(\mathbf{D}_{\text{aug}})$;

*ii*) sort the scores collected in score;

*iii*) compute the image source locations;

*iv*) remove image sources that do not correspond to walls (higher-order by using step *iii*, ghost sources by heuristics); and

*v*) reconstruct the room.

Step *iv* is described in more detail in the *SI Text*. It is not necessary to test all echo combinations. An echo from a fixed wall will arrive at all of the microphones within the time given by the largest intermicrophone distance. Therefore, it suffices to combine echoes within a temporal window corresponding to the array diameter. This substantially reduces the running time of the algorithm. As a consequence, we can be less conservative in the peak-picking stage. A discussion of the influence of errors in the image-source estimates on the estimated plane parameters is provided in (15).

## Experiments

We ran the experiments in two distinctly different environments. One set was conducted in a lecture room at EPFL, where our modeling assumptions are approximately satisfied. Another experiment was conducted in a portal of the Lausanne cathedral. The portal is nonconvex, with numerous nonplanar reflecting objects. It essentially violates the modeling assumptions, and the objective was to see whether the algorithm still gives useful information. In all experiments, microphones were arranged in an arbitrary geometry, and we measured the distances between the microphones approximately with a tape measure. We did not use any specialized equipment or microphone arrays. Nevertheless, the obtained results are remarkably accurate and robust.

The lecture room is depicted in Fig. 6*A*. Two walls are glass windows, and two are gypsum-board partitions. The room is equipped with a perforated metal-plate ceiling suspended below a concrete ceiling. To make the geometry of the room more interesting, we replaced one wall by a wall made of tables. Results are shown for two positions of the table wall and two different source types. We used an off-the-shelf directional loudspeaker, an omnidirectional loudspeaker, and five nonmatched omnidirectional microphones. RIRs were estimated by the sine sweep technique (26). In the first experiment, we used an omnidirectional loudspeaker to excite the room, and the algorithm reconstructed all six walls correctly, as shown in Fig. 6*B*. Note that the floor and the ceiling are estimated near perfectly. In the second experiment, we used a directional loudspeaker. As the power radiated to the rear by this loudspeaker is small, we placed it against the north wall, thus avoiding the need to reconstruct it. Surprisingly, even though the loudspeaker is directional, the proposed algorithm reconstructs all of the remaining walls accurately, including the floor and the ceiling.

Fig. 6 *D* and *F* shows a panoramic view and the floor plan of the portal of the Lausanne cathedral. The central part is a pit reached by two stairs. The side and back walls are closed by glass windows, with their lower parts in concrete. In front of each side wall, there are two columns, and the walls are joined by column rows indicated in the figure. The ceiling is a dome ∼9 m high. We used a directional loudspeaker placed at the point L in Fig. 6*F*. Microphones were placed around the center of the portal. Alas, in this case we do not have a way to remove unwanted image sources, as the portal is poorly approximated by a convex polyhedron. The glass front, numeral 1 in Fig. 6*F*, and the floor beneath the microphone array can be considered flat surfaces. For all of the other boundaries of the room, this assumption does not hold. The arched roof cannot be represented by a single height estimate. The side windows, numerals 2 and 3 in Fig. 6*F*, with pillars in front of them and erratic structural elements at the height of the microphones, the rear wall, and the angled corners with large pillars and large statues, all present irregular surfaces creating diffuse reflections. Despite the complex room structure with obstacles in front of the walls and numerous small objects resulting in many small-amplitude, temporally spread echoes, the proposed algorithm correctly groups the echoes corresponding to the three glass walls and the floor. This certifies the robustness of the method. More details about the experiments are given in the *SI Text*.

## Discussion

We presented an algorithm for reconstructing the 3D geometry of a convex polyhedral room from a few acoustic measurements. It requires a single sound emission and uses a minimal number of microphones. The proposed algorithm has essentially no constraints on the microphone setup. Thus, we can arbitrarily reposition the microphones, as long as we know their pairwise distances (in our experiments we did not "design" the geometry of the microphone



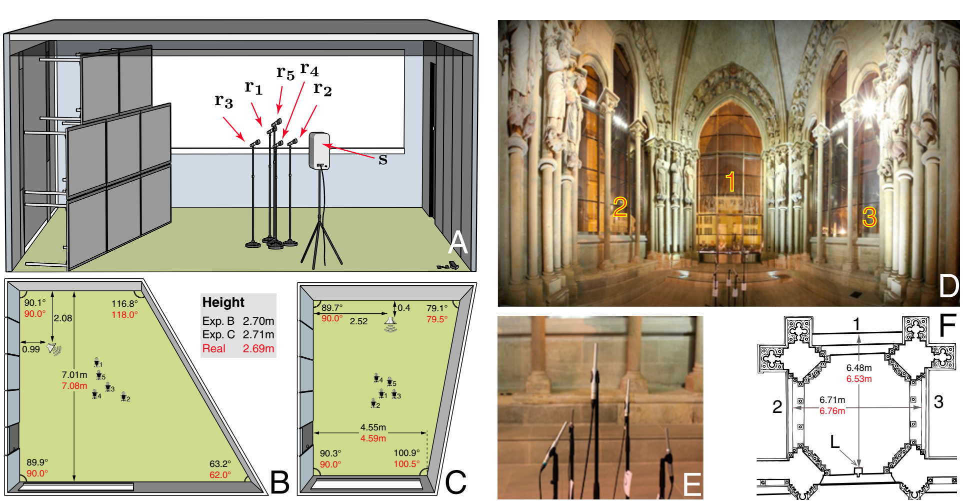


**Fig. 6.** (*A*) Illustration of the room used in the experiment with a movable wall. (*B* and *C*) Two reconstruction results. Real values are indicated in red, and the estimated values are indicated in black. (*D*) Panoramic photo of the portal of the Lausanne cathedral. (*E*) A close up of the microphone array used in cathedral experiments. (*F*) Floor plan of the portal.

setup). Further, we proved that the first-order echoes collected by a few microphones indeed describe a room uniquely. Taking the image source point of view enabled us to derive clean criteria for echo sorting.

Our algorithm opens the way for different applications in virtual reality, auralization, architectural acoustics, and audio forensics. For example, we can use it to design acoustic spaces with desired characteristics or to change the auditory perception of existing spaces. The proposed echo-sorting solution is useful beyond hearing rooms. Examples are omnidirectional radar, multiple-input–multiple-output channel estimation, and indoor localization to name a few. As an extension of our method, a person walking around the room and talking into a cellphone could enable us to both hear the room and find the person's location. Future research will aim at exploring these various applications.

Results presented in this article are reproducible. The code for echo sorting is available at http://rr.epfl.ch.

**ACKNOWLEDGMENTS.** This work was supported by a European Research Council Advanced Grant–Support for Frontier Research–Sparse Sampling: Theory, Algorithms and Applications (SPARSAM) 247006. A.W.'s research is funded by the Fraunhofer Institute for Integrated Circuits IIS, Erlangen, Germany.

1. Kac M (1966) Can one hear the shape of a drum? *Am Math Mon* 73:1–23.
2. Gordon C, Webb D (1996) You can't hear the shape of a drum. *Am Sci* 84:46–55.
3. Gordon C, Webb D, Wolpert S (1992) Isospectral plane domains and surfaces via Riemannian orbifolds. *Invent Math* 110:1–22.
4. Gordon C, Webb DL, Wolpert S (1992) One cannot hear the shape of a drum. *Bull Am Math Soc* 27:134–138.
5. Chapman SJ (1995) Drums that sound the same. *Am Math Mon* 102:124–138.
6. Rosenblum LD, Gordon MS, Jarquin L (2000) Echolocating distance by moving and stationary listeners. *Ecol Psychol* 12:181–206.
7. Antonacci F, Lonoce D, Motta M, Sarti A, Tubaro S (2005) Efficient source localization and tracking in reverberant environments using microphone arrays. *Proceedings of the IEEE International Conference of Acoustics, Speech, and Signal Processing* (IEEE Press, Los Alamitos, CA) 4:1061–1064.
8. Ribeiro F, Ba DE, Zhang C (2010) Turning enemies into friends: Using reflections to improve sound source localization. *Proceedings of the IEEE International Conference on Multimedia and Expo*, (IEEE Press, Los Alamitos, CA), pp 731–736.
9. Ribeiro F, Zhang C, Florencio DA, Ba DE (2010) Using reverberation to improve range and elevation discrimination for small array sound source localization. *IEEE Trans Acoust Speech Signal Process* 18:1781–1792.
10. Dokmanic I, Vetterli M (2012) Room helps: Acoustic localization with finite elements. *Proceedings of the IEEE International Conference of Acoustics, Speech, and Signal Processing* (IEEE Press, Los Alamitos, CA), pp 2617–2620.
11. Lokki T, Pulkki V (2002) Evaluation of geometry-based parametric auralization. *22nd International Conference: Virtual, Synthetic, and Entertainment Audio*.
12. Dokmanic I, Lu YM, Vetterli M (2011) Can one hear the shape of a room: The 2-D polygonal case. *Proceedings of the IEEE International Conference of Acoustics, Speech, and Signal Processing* (IEEE Press, Los Alamitos, CA), pp 321–324.
13. Antonacci F, et al. (2012) Inference of room geometry from acoustic impulse responses. *IEEE Trans Acoust Speech Signal Process* 20:2683–2695.
14. Ribeiro F, Florencio DA, Ba DE, Zhang C (2012) Geometrically constrained room modeling with compact microphone arrays. *IEEE Trans Acoust Speech Signal Process* 20:1449–1460.
15. Tervo S, Tossavainen T (2012) 3D room geometry estimation from measured impulse responses. *Proceedings of the IEEE International Conference of Acoustics, Speech, and Signal Processing* (IEEE Press, Los Alamitos, CA), pp 513–516.
16. Driscoll TA (1997) Eigenmodes of isospectral drums. *SIAM Rev* 39:1–17.
17. Allen JB, Berkley DA (1979) Image method for efficiently simulating small-room acoustics. *J Acoust Soc Am* 65:943–950.
18. Borish J (1984) Extension of the image model to arbitrary polyhedra. *J Acoust Soc Am* 75:1827–1836.
19. Beck A, Stoica P, Li J (2008) Exact and approximate solutions of source localization problems. *IEEE Trans Signal Process* 56:1770–1778.
20. Smith JO, Abel JS (1987) Closed-form least-squares source location estimation from range-difference measurements. *IEEE Trans Acoust Speech Signal Process* 35:1661–1669.
21. Larsson EG, Danev DP (2010) Accuracy comparison of LS and squared-range LS for source localization. *IEEE Trans Signal Process* 68:916–923.
22. Dattorro J (2011) *Convex Optimization & Euclidean Distance Geometry* (Meboo Publishing USA, Palo Alto, CA).
23. Boyd S, Vandenberghe L (2004) *Convex Optimization* (Cambridge University Press, New York).
24. Torgerson WS (1952) Multidimensional scaling: I. Theory and method. *Psychometrika* 17:401–419.
25. Takane Y, Young FW, Leeuw J (1977) Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. *Psychometrika* 42:7–67.
26. Farina A (2000) Simultaneous measurement of impulse response and distortion with a swept-sine technique. *108th Convention of the Audio Engineering Society*.