# Shape From Echoes

Mireille Boutin (Purdue University) and
Gregor Kemper (Technical University of Munich)

March 20-27, 2022

## 1 Problem Statement

Assume we have a room, i.e., an arrangement of planar walls, which may include ceilings, floors, and sloping walls. An omnidirectional loudspeaker and some omnidirectional microphones are in the room. The loudspeaker, modeled as a point source, emits a short duration sound at a frequency high enough so that the ray acoustics approximation is valid. The microphones receive several delayed responses corresponding to the sound bouncing back from each wall. These are the first-order echoes. This project centers around the problem of reconstructing the shape of the room from the measured delay times of the first-order echoes. The difficulty of this lies in the fact that in order to determine the wall positions, echoes received by different microphones but coming from the same wall need to be matched. This has to be accomplished even though the delay times detected by each microphone come as an unlabeled set. So the problem is to figure out under which circumstances, and how, one can find out the correct echo-wall matching and reconstruct the wall positions.

## 2 Overview of the Field

For a long time, researchers have been trying to understand the relation between the geometry of a room its acoustic. For examples, a considerable amount of work was done in the later part of last century to model the source-to-receiver acoustic impulse response so to improve the design of concert halls. For example, the two-point impulse response within rectangular rooms was analyzed in [1]. The results were later extended to arbitrarily polyhedral rooms in [3]. In the last 20 years, there has been a strong interest in developing methods to reconstruct the geometry of a room using various arrangements of microphones and sound sources. Some reconstruction methods assume that the room is two-dimensional (e.g., [7],[2]], [[9]]) while others consider three-dimensional, but still constrained, geometry (e.g., [14], [8]). In some cases, the wall points are reconstructed directly (e.g., [6], [11], [19], [9], [14]) while in other cases, the reflection of a loudspeaker with respect to a wall plane, called a ”mirror point,” is reconstructed (e.g., [22],[13],[8],[18], [10]). In some setups, the clocks of the microphones are synchronized (e.g., [22], [8], [12], [17]); in other setups, they are not (e.g., [20], [16], [2]). The microphones are sometimes placed on a vehicle carrying the microphones (e.g. [15], [4]), for example a cell phones (e.g, [21], [23]).

By and large, previous work has been mostly numerical and experimental. A theoretical standpoint was recently taken by Dokmanić et al. [8], who proved the following result:

**Theorem 1.** *(Dokmanić et al. [8]). Consider a room with a loudspeaker and at least four microphones placed at random positions. Then with probability 1, the first-order echoes of a sound from the loudspeaker uniquely specify the room.*

Our research project started when we got interested in this result and tried to verify it by using methods from computational commutative algebra. Using the modeling done in [5], this turned out to be not very hard, so we started wondering if Theorem 1 can be strengthened. In the theorem, all four (or more) microphones are placed independently at random positions, so the microphone placement requires a configuration space of dimension (at least) 12. For applications, it would be useful if four microphones could be mounted on the body of some vehicle, say a drone, so their relative positions are fixed, but the vehicle is moving in the room. For a vehicle such as a drone, this means that the microphone placement requires only a 6-dimensional configuration space. For many applications it would be even more useful if the loudspeaker, too, could be mounted on the vehicle. Again using methods from computational commutative algebra and computer algebra, we were able to prove the following results:

**Theorem 2.** *(Boutin and Kemper [4]). Consider a given arrangement of walls and a loudspeaker. Also consider a drone that carries four non-coplanar microphones at fixed locations on its body. Within the 6-dimensional configuration space of possible drone positions, those where an incorrect detection of a wall may occur lie in a subspace of dimension 5.*

**Theorem 3.** *(Boutin and Kemper [4]). The same is true if the loudspeaker is carried by the drone.*

We also gave an algorithm for the wall detection. In this, the Cayley-Menger determinants are used for solving the problem of matching echoes, heard by different microphones, that are coming from the same wall. It is easy to see that this algorithm detects every wall from which an echo is heard by all four microphones. So in the above theorems, incorrect detection means that walls are detected which do not actually exist (ghost walls). There can be no method that completely rules this out, so finding that the occurrence of ghost walls is unlikely is the best possible result. Theorem 3 is much harder to prove than Theorem 2. In fact, the proof required a lot of experimenting with different computational methods and some preconditioning of the problem by choosing suitable coordinate systems.

# 3 Scientific Progress Made

During the workshop at Banff we continued the above line of investigation by considering a situation with even fewer degrees of freedom: we considered the case where the microphones are mounted on a ground-based vehicle. It may be counter-intuitive that fewer degrees of freedom make the problem harder, but this is because less freedom of movement make it harder to get out of a position where an incorrect wall detection occurs (a **bad position**). And indeed we found wall configurations in a three-dimensional space where a ground-based vehicle cannot get out of a bad position by an infinitesimal movement. So fewer degrees of freedom do make a difference. In fact, we managed to classify the wall arrangements where such a phenomenon occurs, and characterized them by the catchphrase of "an unlucky stack of mirror points." For the exact definition, we refer to our preprint [5]. Perhaps surprisingly, the very same exceptional wall configurations hold for the situation of a "hovering drone", which has four degrees of freedom.

In summary, we have proved the following result:

**Theorem 4.** *(Boutin and Kemper [5]) Consider a given arrangement of walls and a loudspeaker, and assume that this does not have an unlucky stack of mirror points. Also consider a ground-based vehicle or a hovering drone that carries four non-coplanar microphones at fixed locations on its body. Within the configuration space of possible positions, those where an incorrect detection of a wall may occur lie in a subspace of lower dimension.*

As mentioned above, a result as general as Theorem 2 cannot be obtained in this situation, and Theorem 4 is n this sense the best possible. The theorem considers the physically (and technically) relevant situation of a vehicle that is restricted to two dimensions in a 3-dimensional world. We also considered the truly 2-dimensional case, which is also relevant in some potential applications, and obtained the following smoother result.

**Theorem 5.** *(Boutin and Kemper [5]) Consider a given two-dimensional scene with finitely many walls. Also consider a vehicle in the scene which carries three microphones that do not lie on a common line. A loudspeaker is either placed at a fixed location or mounted on the vehicle. Then within the configuration* *space of possible vehicle positions, those where an incorrect detection of a wall may occur lie in a subspace of lower dimension.*

# 4 Outcome of the Meeting

A direct outcome of the meeting is the new preprint [5], which has already been submitted. But a week at BIRS also provided an optimal forum for holding brainstorming sessions and thinking about new research projects. During the research stay, we initiated the following projects:

- How can our methods take care of inaccurate measurements?
- Can a vehicle work out its own position from the echoes of a sound event?
- Is echo matching also possible if the time of sound emission is unknown?

We are optimistic that tangible results will come out of these projects.

# References

[1] JB Alien and DA Berkley. Image method for efficiently simulating small-room acoustics. *The Journal of the Acoustical Society of America*, 60(S1):S9–S9, 1976.

[2] 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*, 20(10):2683–2695, 2012.

[3] Jeffrey Borish. Extension of the image model to arbitrary polyhedra. *The Journal of the Acoustical Society of America*, 75(6):1827–1836, 1984.

[4] Mireille Boutin and Gregor Kemper. A drone can hear the shape of a room. *SIAM J. Appl. Algebra Geometry*, 4:123–140, 2020.

[5] Mireille Boutin and Gregor Kemper. Can a ground-based vehicle hear the shape of a room? *ArXiv preprint* `https://arxiv.org/abs/2204.00244`, 2022.

[6] Antonio Canclini, Fabio Antonacci, Mark RP Thomas, Jason Filos, Augusto Sarti, Patrick A Naylor, and Stefano Tubaro. Exact localization of acoustic reflectors from quadratic constraints. In *2011 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 17–20. IEEE, 2011.

[7] 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)*, pages 321–324. IEEE, 2011.

[8] Ivan Dokmanić, Reza Parhizkar, Andreas Walther, Yue M. Lu, and Martin Vetterli. Acoustic echoes reveal room shape. *Proceedings of the National Academy of Sciences*, 110, 2013.

[9] Youssef El Baba, Andreas Walther, and Emanuël AP Habets. Reflector localization based on multiple reflection points. In *2016 24th European Signal Processing Conference (EUSIPCO)*, pages 1458–1462. IEEE, 2016.

[10] Youssef El Baba, Andreas Walther, and Emanuël AP Habets. 3d room geometry inference based on room impulse response stacks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 26(5):857–872, 2017.

[11] Jason Filos, Antonio Canclini, Mark RP Thomas, Fabio Antonacci, Augusto Sarti, and Patrick A Naylor. Robust inference of room geometry from acoustic measurements using the hough transform. In *2011 19th European Signal Processing Conference*, pages 161–165. IEEE, 2011.

[12] 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)*, pages 1–5. IEEE, 2016.

[13] Edwin Mabande, Konrad Kowalczyk, Haohai Sun, and Walter Kellermann. Room geometry inference based on spherical microphone array eigenbeam processing. *The Journal of the Acoustical Society of America*, 134(4):2773–2789, 2013.

[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*, 29:1463–1478, 2021.

[15] Fangrong Peng, Tiexing Wang, and Biao Chen. Room shape reconstruction with a single mobile acoustic sensor. In *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, pages 1116–1120. IEEE, 2015.

[16] Marc Pollefeys and David Nister. Direct computation of sound and microphone locations from time-difference-of-arrival data. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 2445–2448. IEEE, 2008.

[17] 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)*, pages 331–335. IEEE, 2016.

[18] Luca Remaggi, Philip JB Jackson, Philip Coleman, and Wenwu Wang. Acoustic reflector localization: Novel image source reversion and direct localization methods. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 25(2):296–309, 2016.

[19] Luca Remaggi, Philip JB Jackson, Wenwu Wang, and Jonathon A Chambers. A 3d model for room boundary estimation. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 514–518. IEEE, 2015.

[20] Jan Scheuing and Bin Yang. Disambiguation of tdoa estimation for multiple sources in reverberant environments. *IEEE Transactions on Audio, Speech, and Language Processing*, 16(8):1479–1489, 2008.

[21] Oliver Shih and Anthony Rowe. Can a phone hear the shape of a room? In *2019 18th ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*, pages 277–288. IEEE, 2019.

[22] 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)*, pages 513–516. IEEE, 2012.

[23] Bing Zhou, Mohammed Elbadry, Ruipeng Gao, and Fan Ye. Batmapper: Acoustic sensing based indoor floor plan construction using smartphones. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, pages 42–55, 2017.