



Hearing the shape of a room

Plumbley, MD

For additional information about this publication click this link.

<http://qmro.qmul.ac.uk/jspui/handle/123456789/5262>

Information about this research object was correct at the time of download; we occasionally make corrections to records, please therefore check the published record when citing. For more information contact scholarlycommunications@qmul.ac.uk

Hearing the shape of a room

Mark D. Plumbley

Queen Mary University of London

Seeing the shape of a room is not difficult for most people. Using a combination of stereo vision and parallax as we move around the room, we can resolve the position and angle of walls to determine the size and shape of the room we are in. However, suppose now that you are blind, or in a windowless room with no lights. It is still possible to get a sense of the size of the room by using sound. Clap your hands and the echoes will typically tell you if you are in a small office, a medium sized classroom, or a large concert hall. But is it possible to tell the shape of a room using sound alone? This is the question addressed by Dokmanic et al (1) in this issue.

Human hearing is particularly sensitive to the sound of objects and shapes. In the age of steam railways, “wheel-tappers” would check for cracks in railway carriage wheels by tapping with a hammer and listening to the echoes (2). In fiction, the character “Daredevil” (Marvel Comics 1964, movie release 2003) used acoustic “radar” to navigate the world. In reality, a small number of blind people are able to use echolocation to find their way around and locate objects, by producing mouth clicks and listening to the returned echoes (3). The findings of Thaler et al (4) suggest that brain regions normally used for vision can be adopted by such echolocation experts to process these click echoes. However, while these results suggest that sounds can be used to determine differences between shapes, they do not confirm whether or not it is possible to *uniquely* determine the shape of a room using sound alone.

We can approach the problem of finding the location of reflective walls in a room by measuring the time between the sound being emitted from the loudspeaker and being picked up by the microphone, after reflection from one of the walls. In Figure 1(A) we see a geometrical view of a 2d example, where we see the sound paths from the sound source s to the microphones r_1 and r_2 after reflection from the north wall (N). Much as we would see light images in a mirror, we can think of the reflections as creating an “image” s^N of the sound source s , and indeed images r_1^N and r_2^N of the microphones r_1 and r_2 . By measuring the time delays, and hence the distances, from s to r_1 and r_2 via the wall reflection, solving for the points that match the measured distances will find the possible locations of the images, at the points where the circles cross. In Figure 1(A) we see that one of these points where the circles cross is the true source image s^N reflected in the north wall: with only 2 microphones there are two points where this happens, so an additional 3rd (non-colinear) microphone will be needed to resolve this completely.

With additional walls, we would like to apply the same technique to find the images of the sound source in the other walls, and so find the remaining walls (Figure 1(B)). However, with more than one wall, the situation is considerably more complex, because the reflections from different walls are not labelled to indicate which wall the echo has been reflected from. The microphones will pick up a sequence of echoes, but we do not know which echo has been reflected from which wall. This

type of ambiguity has an analogy in stereo vision, where a repeating pattern such as a grid or picket fence can be locally “fused” to give illusions where the stereo depth is closer or farther away than the true stereo depth (5).

If the arrangement of microphones is small compared to the room dimensions, such as a microphone array with small diameter, then the echoes will cluster together in time. The echoes from the source to all microphones reflected via the closest wall (e.g. wall N) will arrive before all the echoes reflected via the next-closest wall (e.g. wall W), and so on. The echoes can then be uniquely labelled and the shape of the room resolved (6).

However, in the general case the echoes from different walls may be intermingled, and this simple time-clustering approach is not possible. If we are unable to label the echoes we may get completely illusory (“ghost”) source images, and hence false wall locations. To illustrate, Figure 1(C) shows a situation where the reflection from s to r_1 via the north wall (N) has been mistakenly labelled together with the reflection from s to r_2 via the west wall (W). Here the apparent solution gives two false source images labelled s^2 , neither of which correspond to a valid source images (s^N or s^W). These two “ghost” source images suggest two entirely false walls, as shown by the green dashed lines.

Dokmanic et al (2013) tackle this problem. Their approach is based on an interesting property of Euclidean distance matrices (EDMs), the matrix of pairwise Euclidean distances $\|r_i - r_j\|^2$ between the microphones, as well as the distances $\|s^\alpha - r_i\|^2$ between loudspeaker source images and each microphone. Using the fact that an EDM for a point set in n -d space has rank at most $n+2$, they are able to reject the type of false echo labellings that we saw in Figure 1(C). Specifically, for a 3d room with a loudspeaker and at least 4 microphones, where the microphones are placed at random inside a region where they will pick up all first-order reflections from the loudspeaker, they show that the unlabelled echoes determine the room shape with probability 1. They also develop practical algorithms to find the room shape, and demonstrate it on finding the shape of a classroom, as well as stretching the model by attempting to find the shape of a more complex room (a cathedral portal) that does not satisfy the modelling assumptions. These promising results indicate that it is possible to find the shape of a room from a loudspeaker and a small number of microphones in an almost arbitrary arrangement, without the need for a special microphone array or soundfield microphone.

The authors concentrate on first-order reflections from the walls. They are able to detect and eliminate second- and higher-order reflections from their calculation: these are not needed to estimate the wall locations. But by extending this work to estimate additional reflections, it may be possible to measure the so-called plenacoustic function (7), the room impulse response function between any two points in a room, perhaps using methods based on compressed sensing (8) or other sparsity-based techniques.

While the present paper relies on control and knowledge of the loudspeaker sound source, with fixed microphones, it will be interesting to see if the technique can be extended to handle estimated sound sources and a small number of mobile microphones. We could speculate that, not only could the room shape be estimated from someone moving around talking into their mobile phone, as the authors suggest, but it may be possible to “crowdsource” the shape of a room from the microphones on the many smartphones which are now carried around.

As a final note, the authors of (Dokmanic et al, 2013) have released the code and data to reproduce the results of the paper in their Reproducible Research Repository (<http://rr.epfl.ch/>). Research in signal and image processing is often based on algorithms implemented in software with many hidden complexities and adjustable parameters, such that can be very difficult for other researchers to follow precisely what has been done just from the published paper alone. Together with the group of Donoho and colleagues (Donoho et al, 2009) at Stanford, the EPFL group have been one of the leading actors promoting reproducible research in this field, setting an example for other researchers to follow.

References

- [1] Dokmanic I, Parhizkar R, Walther A, Lu YM, Vetterli M (2013) Acoustic echoes reveal room shape. *Proc Natl Acad Sci USA* 110:12186-12191.
- [2] Keppens VM, Maynard JD, Migliori (2010) A Listening to materials: From auto safety to reducing the nuclear arsenal. *Acoustics Today*, 6(2):6-13
- [3] Downey G (2011) Getting around by sound: Human echolocation. *PLoS Blogs: Neuroanthropology*. URL <http://blogs.plos.org/neuroanthropology/2011/06/14/getting-around-by-sound-human-echolocation/>. Posted 14 June 2011. Last accessed 16 June 2013
- [4] Thaler L, Arnott SR, Goodale MA (2011) Neural correlates of natural human echolocation in early and late blind echolocation experts. *PLoS ONE*, 6(5): e20162 (16 pages)
- [5] Marr D, Poggio T (1979) A computational theory of human stereo vision. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, 204(1156):301-328
- [6] Tervo S, Tossavainen T (2012) 3d room geometry estimation from measured impulse responses. In: *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2012)*, Kyoto, Japan, pp. 513-516, 25-30 March 2012
- [7] Ajdler T, Sbaiz L, Vetterli M (2006) The plenacoustic function and its sampling, *IEEE Transaction on Signal Processing* 54:3790-3804
- [8] Mignot R, Daudet L, Ollivier F (2012) Interpolation of room impulse responses in 3d using compressed sensing. In: *Proceedings of the Acoustics 2012 Nantes Conference*, 23-27 April 2012, Nantes, France, pp. 2943-2948
- [9] Donoho DL, Maleki A, Rahman IU, Shahram M, Stodden V (2009) Reproducible Research in Computational Harmonic Analysis. *Computing in Science & Engineering* 11(1):8-18

Caption for Figure 1:

(A) Sound path from source \mathbf{s} to microphones \mathbf{r}_1 and \mathbf{r}_2 reflected from the north wall (N). Source image \mathbf{s}^N is one of two points which have the correct distances $\|\mathbf{s} - \mathbf{r}_1\|$ and $\|\mathbf{s} - \mathbf{r}_2\|$. This is also shown in (B) for the west wall (W). If the echoes from different walls are labelled together by mistake (C), we will get incorrect "ghost" source images ($\mathbf{s}^?$) suggesting false walls that do not exist.

