



Audio Engineering Society Convention e-Brief

Presented at the 139th Convention
2015 Oct 29–Nov 1, New York, USA

This Engineering Brief was selected on the basis of a submitted synopsis. The author is solely responsible for its presentation, and the AES takes no responsibility for the contents. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Audio Engineering Society.

Visualization of Compact Microphone Array Room Impulse Responses

Luca Remaggi¹, Philip J. B. Jackson¹, Philip Coleman¹, and Jon Francombe²

¹Centre for Vision, Speech, and Signal Processing, University of Surrey, Guildford, GU2 7XH, UK

²Institute of Sound Recording, University of Surrey, Guildford, GU2 7XH, UK

Correspondence should be addressed to Luca Remaggi (l.remaggi@surrey.ac.uk)

ABSTRACT

For many audio applications, availability of recorded multi-channel room impulse responses (MC-RIRs) is fundamental. They enable development and testing of acoustic systems for reflective rooms. We present multiple MC-RIR datasets recorded in diverse rooms, using up to 60 loudspeaker positions and various uniform compact microphone arrays. These datasets complement existing RIR libraries and have dense spatial sampling of a listening position. To reveal the encapsulated spatial information, several state of the art room visualization methods are presented. Results confirm the measurement fidelity and graphically depict the geometry of the recorded rooms. Further investigation of these recordings and visualization methods will facilitate object-based RIR encoding, integration of audio with other forms of spatial information, and meaningful extrapolation and manipulation of recorded compact microphone array RIRs.

1 Introduction

The room impulse response (RIR) is an audio signal characterizing the acoustic of the room where it is recorded. Hence, it underpins many audio signal processing research areas (e.g. spatial audio, source separation, source tracking, audio reverberation and dereverberation) since it provides information about microphone and loudspeaker positions, room geometry and room size. Multi-channel RIRs (MC-RIRs) are usually used, providing the opportunity to apply algorithms to extract important parameters such as directions or times of arrival (DOAs or TOAs, respectively).

In the literature, no model yet exists to generate RIRs that can exactly simulate the real acoustic properties of a room. Therefore, many researchers use databases of recorded RIRs. One of the first publicly-available

datasets was the one presented in [1], where, like in [2], a high number of loudspeaker positions were measured. However, the use of a uniform linear array (ULA) of microphones limited its applicability because models to analyse room acoustics in 3D must exploit at least a 2D configuration of microphones. In [3] and [4], one of the main contributions was given by binaural recordings. B-format datasets were provided by [5], together with omnidirectional recordings, with microphones being placed in a grid which covered almost the entire plan of the rooms. However, the microphones were spatially too sparse to apply algorithms assuming the far-field. In [6] the reflection positions were visualized using directional MC-RIRs recorded with an Eigenmike.

Recently, several algorithms for MC-RIR visualization have been implemented, designed to show different acoustic properties of the rooms. In [7], plane

wave decomposition (PWD) was applied to visualize the amount of energy arriving over time to a uniform circular array (UCA) from any DOA. Reflector positions can also be estimated and graphically shown as planes, localizing the image sources directly [8, 9], or constructing geometric surfaces in the space [10]. In [11], the spatio-temporal response was visualized to analyse concert hall acoustics. Image source locations relative to a spatial RIR were presented in [12].

In this article we present datasets and visualizations of MC-RIRs recorded using a compact UCA in two rooms at the University of Surrey and at Emmanuel Church in Guildford. Other datasets recorded using a compact uniform rectangular array (URA) are also available. The array’s compactness allows us to use algorithms assuming the far-field, and offers a listener’s perspective on the recorded rooms. To demonstrate the information contained in the MC-RIRs and provide a graphical representation of the rooms involved, room visualization methods are applied. Sec. 2 introduces the visualization algorithms; Sec. 3 presents the datasets and shows the visualizations; and Sec. 4 concludes.

2 Room visualization techniques

In this section we describe the visualization techniques and the room characteristics that they demonstrate.

Raw data and DOA-time energy analysis. One useful technique to understand the room acoustics is to visualize the DOA of acoustic energy over time. Here, a visualization similar to [7] was achieved by steering a superdirective beamformer (the superdirective array, SDA [13]) in each azimuth direction with a resolution of one degree. The energy arriving from each direction is visualized after calculating the short-term power average by sliding a 0.37 ms Hann window along the steered RIRs. This representation can be considered as an evolution of [14], where the author presented a visualization of MC-RIRs generated by plotting the raw signals adjacent to one another in the time domain.

Reflection and reflector localization. These techniques aim to visualize the reflections and reflecting surfaces. A first model is based on image sources. To localize the image sources, the TOAs and DOAs are utilized. The dynamic programming projected phase-slope algorithm (DYPSA), modified to be used with RIRs [15], is used to extract the TOAs. Based on the TOAs, the RIRs are segmented. The segmented

Table 1: Properties of the three rooms presented.

Dataset	Dimension (m)	RT60 (ms) 500Hz–2kHz
MainChurch	$19.68 \times 24.32 \times 5.97$	1500–1200
Studio1	$14.55 \times 17.08 \times 6.50$	1400–1100
AudioBooth	$4.12 \times 4.98 \times 2.10$	413–115

signals are used to extract the DOA parameters for the early reflections using a 3D delay-and-sum beamformer (DSB) [15]. Finally, the reflector is drawn as the plane perpendicular to the line generated by the image source and the loudspeaker and passing through its mid-point. The position of the reflection is given by the intersection of the reflector and the line between the microphone array centre and the image source.

A second model uses ellipsoids to estimate the reflector positions. A set of ellipsoids are generated, having foci on the microphone-source combinations and major axis of the reflection’s path length. The random sample consensus (RANSAC)-based technique is used to find the estimated reflector location, i.e. the common tangent plane to all the ellipsoids generated [10].

3 Recorded dataset visualization

In this section we present the MC-RIR datasets recorded using a compact microphone array in various rooms. We then show the visualizations applied to these measurements, and comment on the room acoustic features highlighted.

3.1 Recorded MC-RIR datasets

Several sets of RIRs are available. Here, we describe the three datasets that are used for the visualizations presented in this paper (summarized in Tab. 1), and further sets are available online. Countryman B3 omni lavalier microphones were used for each dataset.

AudioBooth. The “AudioBooth” is an acoustically treated room at the University of Surrey. A 17 channel loudspeaker array was mounted on a truncated geodesic sphere with the equator at 1.02 m elevation. The array comprised nine Genelec 8020B loudspeakers around the equator at 1.68 m radius and $0, \pm 30, \pm 70, \pm 110$ and ± 155 degrees in azimuth relative to the centre channel. At ± 30 and ± 110 degrees, further loudspeakers were placed at ± 30 degrees elevation. The microphone array, positioned at the centre of the loudspeaker array, was a 48 channel double concentric UCA having 24

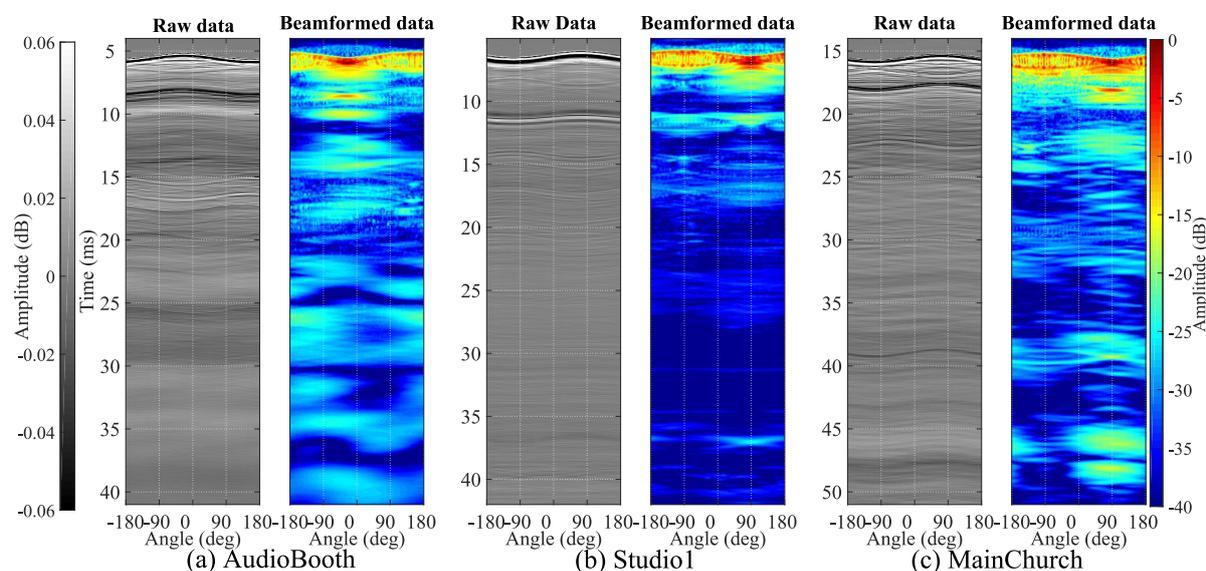


Fig. 1: Raw MC-RIR data visualization and superdirective DOA-time energy analysis (here titled beamformed data), for AudioBooth (a), Studio1 (b) and MainChurch (c).

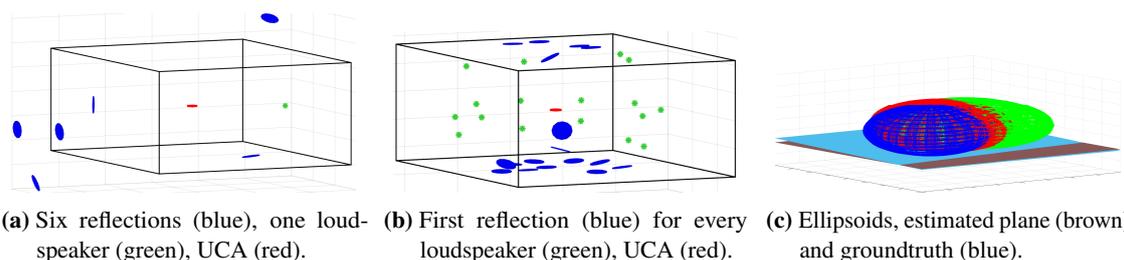


Fig. 2: AudioBooth reflection and reflector estimation, showing the first six reflections due to (a) a single loudspeaker; (b) first reflection of multiple loudspeakers simultaneously; (c) the resulting reflector estimation.

microphones evenly spaced around radii of 0.083 m and 0.104 m. A sound-field microphone was also positioned at the centre of the double UCA. RIRs were recorded at a sampling frequency of 48 kHz by the log sine sweep method.

Studio1. RIRs were also recorded in “Studio1”, a large recording studio at the University of Surrey. A total of 15 loudspeakers positioned 2.0–4.0 m from the array at heights of 0.3–1.50 m were used. As before, RIRs were recorded at 48 kHz by the log sine sweep method. The loudspeakers were Genelec 1032B and the same 48 channel double UCA, with 1.5 m height, was used as for the AudioBooth.

Emmanuel Church. RIRs were recorded in two rooms at Emmanuel Church: the “MainChurch” and

the “OldChurch”. Visualizations of the MainChurch are given here, and the OldChurch data and documentation is available online. The MainChurch MC-RIRs were recorded using Genelec 8030A loudspeakers positioned at $0, \pm 30$ and ± 110 degrees in azimuth and 0 and 30 degrees elevation at a radius of 5 m, giving a total of 10 positions. The 48 channel dual UCA and Soundfield microphones were used as for the AudioBooth.

Further datasets. Further datasets recorded in “Studio2” and the “Vislab” are available, in each case having 60 loudspeakers equally spaced around a radius of 1.68 m, and with various positions of a 48 channel uniform rectangular microphone array combined to make a grid of measurement positions. In Studio2 864

different microphones positions were measured, and in Vislab 384 positions were measured. In each case the maximum length sequence (MLS) technique was used at sampling frequency 48 kHz. These measurements are available from <http://cvssp.org/soundzone/resource>; DOI <http://dx.doi.org/10.15126/surreydata.00808179>.

3.2 MC-RIRs visualization

The MC-RIR visualization techniques presented in Sec. 2 were applied to the recorded data. As shown in Fig. 1, the raw MC-RIR data representation allows visualization of the sound waves arriving to the microphones. In addition, the superdirective DOA-time energy analysis emphasises the DOAs of the early reflections. There are clearly differences among the room acoustics. In particular, the early reflections much stronger for AudioBooth and MainChurch compared to Studio1. Studio1 is essentially a shoebox, with some cluttering introduced by the measurement setup (see [16]). Conversely, MainChurch has a complex shape and many other objects near the propagation path, and the later reflections in AudioBooth are diffuse and due to low frequency modes. The AudioBooth reflections are also shown in Fig. 2, where the blue spots show the estimated reflection locations. The three sub-figures show how it is possible to extract first-order (represented inside the shoebox) and higher-order reflections using one loudspeaker (Fig. 2a), and the possibility of selecting just the first reflection from each loudspeaker (Fig. 2b), to localize the reflector (Fig. 2c).

4 Conclusion

A new database of RIR measurements was recorded using a compact microphone array. Visualization methods were applied to them, highlighting the detail inherent in compact array perspective of the room acoustics. The presented datasets, formatted following the Spatially Oriented Format for Acoustics (SOFA)[17], are available for download from <http://cvssp.org/data/s3a>; DOI <http://dx.doi.org/10.15126/surreydata.00808465>.

5 Acknowledgments

This work was supported by the EPSRC Grant S3A: Future Spatial Audio for an Immersive Listener Experience at Home (EP/L000539/1), and the BBC as part of the Audio Research Partnership. This work was also supported by the EPSRC Grant EP/K014307/1, and the MOD University Defence Research Collaboration in Signal Processing.

References

- [1] Wen, J. Y. C., Gaubitch, N. D., Habets, E. A. P., Myatt, T., and A., N. P., "Evaluation of speech dereverberation algorithms using the MARDY database," in *Proc. of the IWAENC*, 2006.
- [2] Hadad, E., Heese, F., Vary, P., and Gannot, S., "Multichannel audio database in various acoustic environments," in *Proc. of the IWAENC*, 2014.
- [3] Kayser, H., Ewert, S. D., Anemüller, J., Rohdenburg, T., Hohmann, V., and Kollmeier, B., "Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses," *EURASIP J. on ASP*, (6), 2009.
- [4] Erbes, V., Geier, M., Weinzierl, S., and Spors, S., "Database of single-channel and binaural room impulse responses of a 64-channel loudspeaker array," in *Proc. of the 138th AES Convention*, 2015.
- [5] Stewart, R. and Sandler, M., "Database of omnidirectional and B-format room impulse responses," in *Proc. of the ICASSP*, 2010.
- [6] Farina, A., Amendola, A., Capra, A., and Varani, C., "Spatial analysis of room impulse responses captured with a 32-capsules microphone array," in *Proc. of the 130th AES Convention*, 2011.
- [7] Melchior, F., Sladeczek, C., Partzsch, A., and Brix, S., "Design and implementation of an interactive room simulation for wave field synthesis," in *Proc. of the 40th AES Conference*, 2010.
- [8] Tervo, S. and Tossavainen, T., "3D room geometry estimation from measured impulse responses," in *Proc. of the ICASSP*, 2012.
- [9] Dokmanić, I., Parhizkar, R., Walther, A., Lu, Y. M., and Vetterli, M., "Acoustic echoes reveal room shape," *PNAS*, 110(30), p. 12186–12191, 2013.
- [10] Remaggi, L., Jackson, P. J. B., Wang, W., and Chambers, J. A., "A 3D model for room boundary estimation," in *Proc. of the ICASSP*, 2015.
- [11] Pätynen, J., Tervo, S., and Lokki, T., "Analysis of concert hall acoustics via visualizations of time-frequency and spatiotemporal responses," *J. ASA*, 133(2), pp. 842–857, 2013.
- [12] Tervo, S., Pätynen, J., Kuusinen, A., and Lokki, T., "Spatial decomposition method for room impulse responses," *J. AES*, 61(1/2), pp. 17–28, 2013.
- [13] Bai, M. R. and Chen, C.-C., "Application of convex optimization to acoustical array signal processing," *J. Sound Vib.*, 332(25), pp. 6596–6616, 2013.
- [14] Hulsebos, E., *Auralization using wave field synthesis*, Ph.D. thesis, Technische Universiteit Delft, 2004.
- [15] Remaggi, L., Jackson, P. J. B., and Coleman, P., "Estimation of room reflection parameters for a reverberant spatial audio object," in *Proc. of the 138th AES Convention*, 2015.
- [16] Francombe, J., Brookes, T., Mason, R., Flindt, R., Coleman, P., Liu, Q., and Jackson, P. J. B., "Production and Reproduction of Program Material for a Variety of Spatial Audio Formats," in *Proc. of the 138th AES Convention*, 2015.
- [17] AES69, "AES standard for file exchange - Spatial acoustic data file format," 2015.