SurrRoom 1.0 Dataset: Spatial Room Capture with Controlled Acoustic and Optical Measurements

Craig Cieciura[1], Marco Volino[1], and Philip J. B. Jackson[1]

[1] *CVSSP, University of Surrey, UK*

Abstract

Room acoustics, and the perception thereof, are important considerations in research, engineering, architecture, creative expression, and many other areas of human activity, particularly indoors. Typical room datasets contain either disparate measurements of diverse spaces, e.g., the OpenAir dataset (Murphy and Shelley, 2010), or rich sets of measurements within a few rooms, e.g., CD4M (Stewart and Sandler, 2010). The development of techniques, such as the RSAO, with 6 degrees of freedom movement in media applications, requires testing distance-related effects. Hence, there is a need for consistency across room measurements in terms of source-room-receiver configuration, such as in (Lokki et al., 2011) but particularly for typical rooms. Those available, containing ARIRs, BRIRs, and with a consistent measurement procedure, tend to be limited in the range of rooms measured (Bacila and Lee, 2019). We designed an RIR dataset including seven rooms with typical reverberation times from 0.24 s to 1.00 s and volumes from 50 m3 to 1600 m3. 1O-ARIRs and BRIRs were captured with a regular measurement procedure. LiDAR scans and 3D pictures were also captured, providing precise geometries and visual references. Analysis based on metrics defined in ISO 3382-1:2009 is presented to describe each of the rooms in the dataset.

1 Introduction

Auditory perception of the local environment strongly influences our understanding of, and performance of activities within, the spaces we inhabit, and can influence our mood, experience, and sense of wellbeing. In creative industries, such as media production, spatial sound's ability to depict an acoustic environment bears a particular responsibility in forming the user's impression of presence and immersion. Room acoustics is also an important consideration in research, engineering, architecture, and many other areas of human activity. Perception of room acoustics is therefore an ever-popular topic of study.

This paper presents a dataset of consistent 3D measurements over a series of rooms, selected from across the University of Surrey Stag Hill campus in Guildford, UK, with mid-range reverberation times, calculated as RT60 T20s, of 0.24 s to 1.00 s. The dataset contains controlled acoustic measurements consisting of binaural (BRIR) and first-order ambisonic (1O-ARIR) room impulse responses, and optical measurements consisting of 3D photographic and Light Detection and Ranging (LiDAR) scans. This differs from existing available datasets which do not provide all of the features listed.

In the following sections we explain our motivation for producing this dataset; the process of creating it, including methods, diagrams of capture positions and images of the rooms; and describe their characteristics via plots of reverberation time and metrics, such as Early Decay Time (EDT), Clarity and Definition, which are defined in ISO 3382-1:2009 [1]. We also discuss potential applications and future work.

2 Motivation

Today's computational resources facilitate acoustical simulation of geometries; however, room simulations remain approximations to real rooms due to complex scattering and wave behaviour and limitations in modelling the surface geometry and materials of everyday spaces. In contrast, real-world acoustic impulse response measurement enables us to characterise any given room accurately under linearity and time-invariance assumptions.

Binaural RIRs (BRIRs) recorded using head and torso simulators can be used to reproduce the impression of a sound within a captured room over headphones. Ambisonic RIRs (ARIRs) are often used within 360-degree immersive technologies, such as virtual reality and augmented reality (VR and AR), supported by Unity, Unreal Engine, etc. For dynamic headphone reproduction, Ambisonic signals are rendered binaurally, although with various practical limitations.

One more flexible approach to spatial reverberation is the Reverberant Spatial Audio Object (RSAO) [2, 3, 4, 5, 6, 7] which parameterises the room response using principles of object-based production and reproduction from input signals including a 1O-ARIR input. For perceptual testing of binaural rendering systems, such as the RSAO, both ARIRs and BRIRs are required. Many existing datasets tend to provide either BRIRs, e.g., Erbes et al. [8], or ARIRs, e.g., CD4M [9], whereas ours offers both.

For examining distance-related effects, whilst assessing plausibility of room impression, there is a need for consistency across room measurements in terms of source-room-receiver configuration. Existing datasets either offer a regular set of measurements across a limited set of concert halls and other *acoustically interesting* rooms, such as CD4M [9] and those described by Lokki et al. [10] and Bacila and Lee [11], or disparate measurements across a wide variety of rooms, such as the OpenAir dataset [12]. For development of the RSAO, with applications within media production such as creating plausible environments for VR or adding reverb to objects added to a scene using AR, a new dataset including rooms with more typical acoustical conditions, i.e., low to mid-range reverberation times, measured with a consistent procedure, was therefore required.

RSAO parameters have also been generated from optical inputs [13, 14], hence our inclusion of optical measurements. These consisted of an arrangement of spherical 3D cameras, as described by Kim et al. [13], capturing at each IR measurement position, and LiDAR scans which provide precise geometry and could be used for generating simulations.

3 Acquisition

3.1 Audio

Seven rooms were selected based on their distribution of designs, architecturally and intended usage; dimensions, ranging from approximate volumes of 55 m^3 to 1600 m^3; and estimated reverb time and frequency response. The rooms are described as:

1. *Listening Room*: an ITU-R BS.1116 standard critical listening room.
2. *Pop Recording Studio*: a *dry* studio live room designed for recording pop and rock music;
3. *Living Room Lab*: an office space converted into a laboratory with furnishings installed to emulate a typical domestic living room environment, described further by Cieciura et al. [15].
4. *Large Lecture Theatre*: a lecture theatre with tiered seating for 280 persons.
5. *Large Classroom*: a large classroom with single-level seating for approximately 60 persons depending on seating arrangement.
6. *Small Lecture Theatre*: a lecture theatre with tiered seating for 160 persons.
7. *Classical Recording Studio*: a large studio live room designed for recording classical music ensembles. Heavy floor-to-ceiling stage curtains are installed in this room which can be closed to cover approximately 40% of the wall area. Separate measurements were taken with the curtains in fully opened (CO) and fully closed positions (CC).

Figure 1 displays diagrams of the rooms and capture positions, drawn to the same relative scale. Images of each room, extracted from the LiDAR

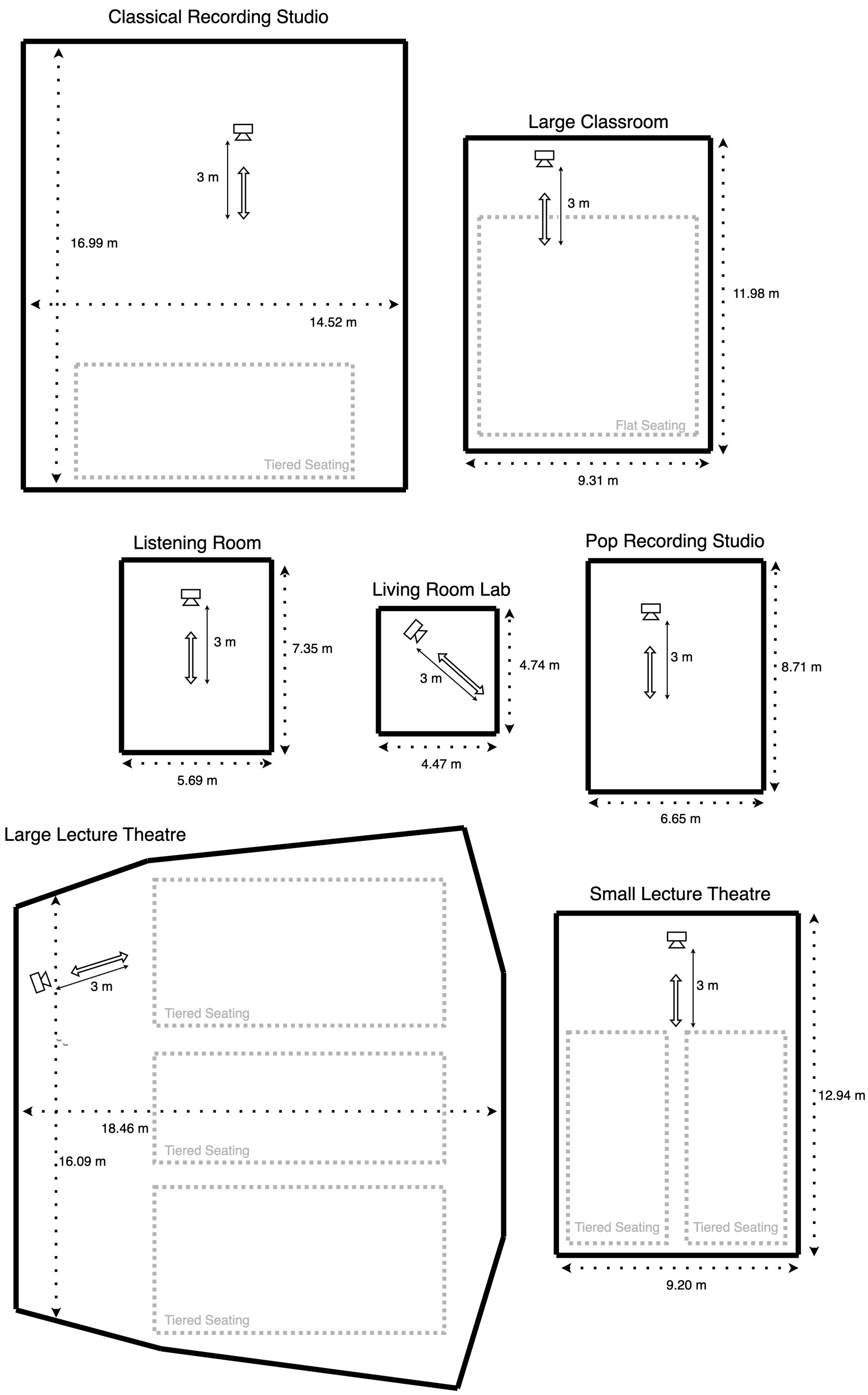
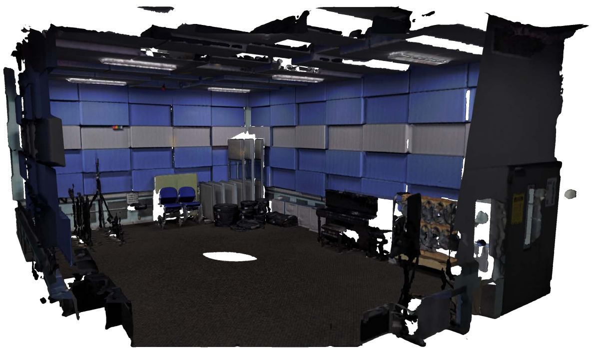


Figure 1: Diagrams of capture positions within each room, from 1 m to 3 m at 0.5 m intervals from the face of the loudspeaker, drawn with the same relative scale. The dimensions of each room are annotated. Loudspeaker illustration not to scale.

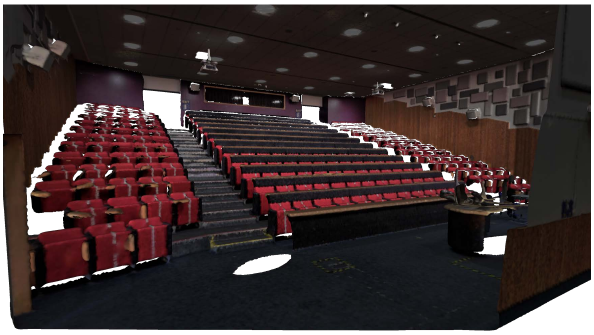
Figure 2: Illustrative images of each room extracted from the LiDAR scans.

(a) Listening Room.

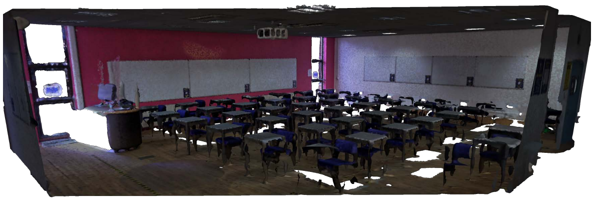
(b) Pop Recording Studio.



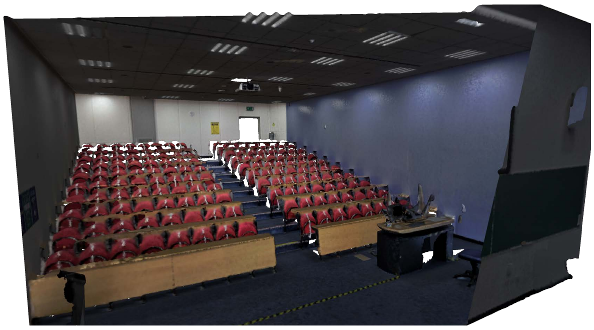
(c) Living Room Lab (LiDAR scan unavailable due to lab decommisioning).



(d) Large Lecture Theatre.



(e) Large Classroom.

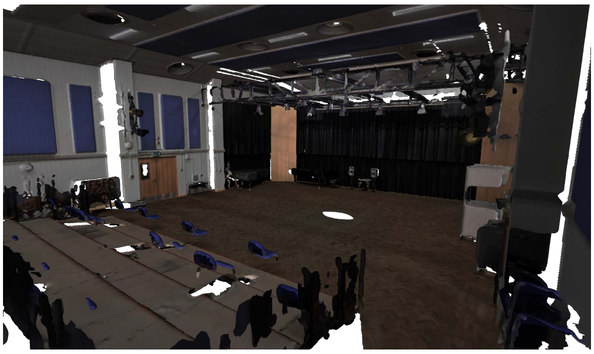


(f) Small Lecture Theatre.

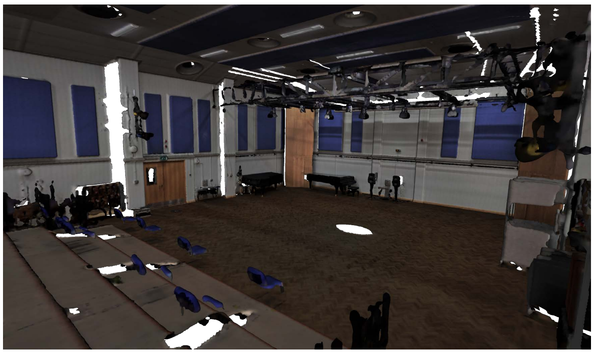
scans described in Section 3.2, are presented in Figure 2.

A compromise was sought between room count and receiver configurations. With a representative source position within the room, receivers were placed consecutively at distances from 1 m to 3 m, at 0.5 m intervals, in a single plane outwards from the source position. Measurements were recorded using a first-order Ambisonic microphone, Soundfield ST450mkII, and a KEMAR 45BB head-and-torso simulator. At each source-receiver distance, the BRIRs were measured at azimuthal head orientations of 15° intervals between -45° and +45°. The swept-sine method was used [16] with a Genelec 8040B producing the excitation signal.

Audio was captured at a sample rate of 48 kHz and a bit depth of 32 bits. A MOTU 1248 audio interface was used for microphone pre-amplification, analog-to-digital and digital-to-analog conversion. A laser measurement device, Fluke 414D, was used to measure source to receiver distances, heights, relative positions and room dimensions. All aspects of the measurement method were consistent between the rooms including equipment, measurement intervals and elevation of the source and receiver at 1.5



(g) Classical Recording Studio, Curtains Closed.



(h) Classical Recording Studio, Curtains Open.

m relative to the main floor surface.

The RIRs were post-processed using Matlab. For each room, for each of the 1O-RIRs and BRIRs, normalisation to -1 dBFS was performed on the RIR captured at 1 m from the source. The same gain coefficient was then applied to each of the remaining RIRs to maintain relative levels.

3.2 Visual

To accurately model the geometry of the chosen rooms and spaces in the proposed dataset, a Faro Focus S Laser Scanner [1], a highly accurate LiDAR-based scanner, was used. Laser-based scanning, using LiDAR, is widely used within film, television and game production to capture geometrically accurate measurements of an environment with respect to the scanner. The output of the scanner is not generally useful in its raw form and often requires further processing, manual clean-up and editing by a digital artist. With the right level of human intervention, a compact model that can be efficiently rendered using a wide range of digital platforms can be achieved at the cost of significant manual input from highly skilled digital artists.

All but one room, *Living Room Lab*, was captured using the LiDAR scanner, At the time of the optical measurements *Living Room Lab* had already been decommissioned.

During the capture phase, multiple locations within each room were sampled in order to minimise holes in the final reconstructions, caused by object and environmental occlusions. Multiple scans of the environment were processed using the Faro SCENE processing software [2], which converts the raw angle/distance/colour data into a 3D point cloud representation. Further processing steps performed alignment, to convert multiple spatial observations into a single consistent 3D point cloud for each room. The resulting consistent point cloud was manually cleaned to remove erroneous points before conversion from a 3D point cloud to 3D mesh. The final processing stage was to bake the point cloud-based colours onto the mesh surface and store in a texture atlas. Modelling the rooms using a 3D mesh and texture atlas allows for efficient storage and rendering in a game engine, e.g. Unity or Unreal Engine, to support inclusion of virtual environments for immersive experience.

Stereo 360° images were also captured at two elevations simultaneously using Ricoh Theta V cameras, at each receiver position, to facilitate 3D scene reconstruction of the sort described by Kim and Hilton [17].

4 Analysis of Rooms

A separate set of reverb time measurements were performed via the interrupted noise method [1] with two source positions and six receiver positions distributed around the room – three per source position. A measurement device, NTi XL2, and signal generator, NTi Minirator, were used. Figure 3 displays a plot of RT60 T20 measurements in one-third-octave bands for each room in the dataset.

Acoustic parameters, described by ISO 3382-1:2009, were calculated on the omni-directional, W, component of the 1O-ARIR measured at 2 m for each room. Table 1 provides these acoustic parameters alongside values for RT60 T20 – calculated from the separate set of reverb time measurements – averaged over 200 Hz–4 kHz one-third-octave bands,

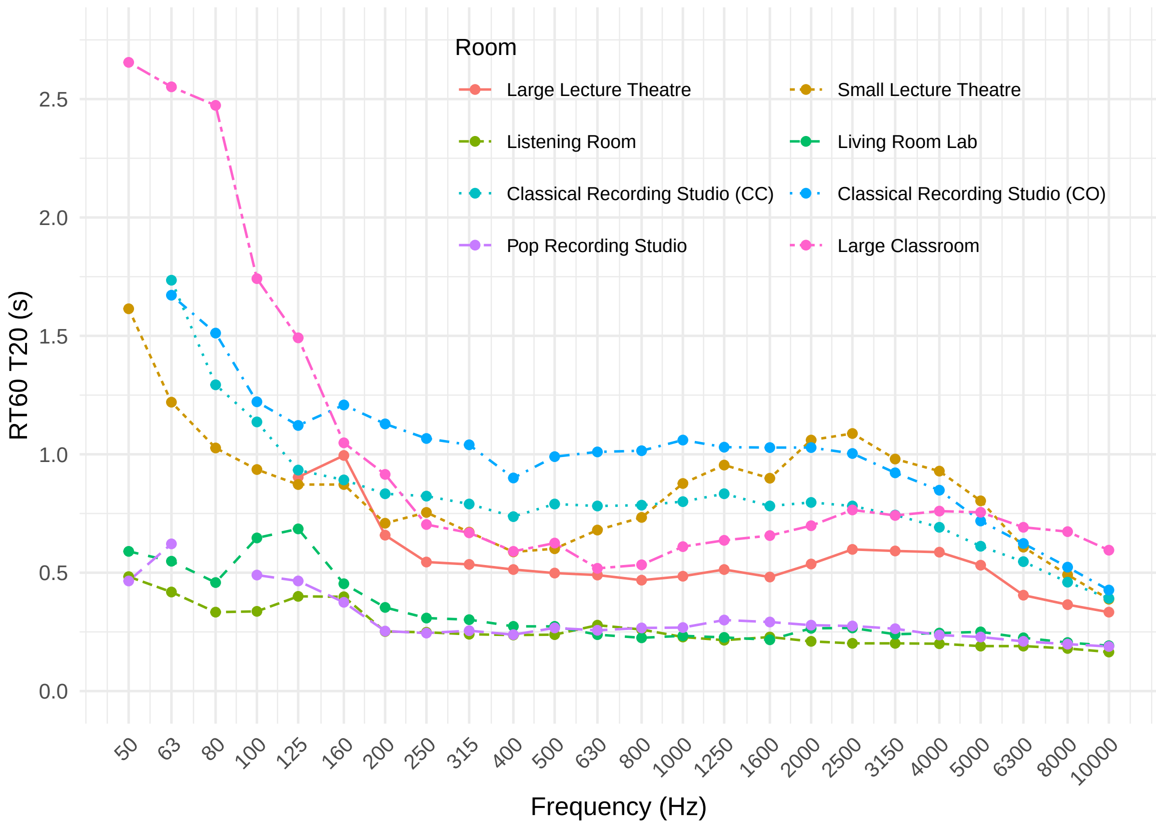


Figure 3: One-third-octave band measurements of $RT_{60}(T_{20})$ in measured rooms.

volume (V) and bass ratio (BR).

5 Suggested Uses

We propose that the consistent measurement procedure, range of rooms and inclusion of BRIRs and 1O-ARIRs make this dataset highly relevant to the development, and objective and perceptual testing, of systems which parametrically represent the acoustical features of rooms, such as the RSAO [2, 3, 4, 5, 6, 7] . Furthermore, the inclusion of optical measurements, including LiDAR scans from which precise geometry can be extracted, make this dataset suitable for testing approaches which estimate acoustical parameters from optical inputs, e.g., Kim et al. [13], Remaggi et al. [14], or as a ground truth for the development of acoustical simulations.

6 Access

The dataset can be accessed from `DOI : 10.15126/surreydata.900689`. The impulse responses are made available both as multichannel .WAV files and in the Spatially Oriented Format for Acoustics (SOFA) [18, 19]. 3D pictures are included as image files and processed LiDAR scans are provided as .PLY files.

7 Future Work

A second version of the SurrRoom dataset has been proposed. To achieve the highest standard of precision in the consistency of measurements between rooms, a motorised platform, capable of accurately positioning receivers at precise distances from the source and at precise azimuthal rotations, is under development. For each of the rooms in the dataset,

Room	$RT_{60}T_{20}$ (s)	V (m^3)	EDT (s)	C50 (dB)	C80 (dB)	D50 (%)	BR
Listening Room	0.24	104	0.19	18.47	25.97	98.60	1.29
Pop Recording Studio	0.27	231	0.22	15.67	23.59	97.35	1.26
Living Room Lab	0.27	55	0.26	14.65	22.67	96.65	1.86
Large Lecture Theatre	0.56	102	0.17	15.09	18.40	96.90	1.60
Large Classroom	0.70	327	0.21	11.38	14.74	93.20	1.82
Classical Recording Studio (CC)	0.78	1600	0.15	11.60	13.11	93.45	1.10
Small Lecture Theatre	0.82	469	0.43	9.01	12.63	88.75	1.11
Classical Recording Studio (CO)	1.00	1600	0.71	9.27	10.82	89.25	1.11

Table 1: Analysis of rooms with reverberation time (RT60 T20) averaged over 200 Hz–4 kHz one-third-octave bands, volume (V), and analysis of the omnidirectional W component of the 1O-ARIR at 2.0 m distance, averaged between 500 Hz and 1 kHz frequency bands: Early Decay Time (EDT); ratio of early to late time energy, Clarity, with early time limits of 50 ms (C_{50}) and 80 ms (C_{80}); Definition, with early time of 50 ms (D_{50}); and Bass Ratio (BR) calculated as $(RT60_{125} + RT60_{250})/(RT60_{500} + RT60_{1000})$.

capture of the measurements took approximately eight hours. Much of this time was spent ensuring accurate placement and orientation of receivers after manual movement between each measurement position. If the majority of the capture process could be automated then our capacity for capturing measurements at smaller intervals and in greater quantities could be significantly increased.

Acknowledgments

This research was supported by UKRI EPSRC and BBC Prosperity Partnership AI4ME: Future Personalised Object-Based Media Experiences Delivered at Scale Anywhere EP/V038087 and UKRI Innovate UK Polymersive PR: 105168. Thanks also to Andrés Estrella Terneux for his assistance with the impulse response measurements.

References

[1] ISO, "ISO 3382-1:2009: Acoustics — Measurement of Room Acoustic Parameters — Part 1: Performance Spaces," Technical report, 2009.

[2] Remaggi, L., Jackson, P., and Coleman, P., "Estimation of Room Reflection Parameters for a Reverberant Spatial Audio Object," in *Audio Engineering Society Convention 138*, Audio Engineering Society, 2015.

[3] Coleman, P., Franck, A., Menzies, D., and Jackson, P. J. B., "Object-Based Reverberation Encoding from First-Order Ambisonic RIRs," in *The 142nd Audio Engineering Society Convention*, Audio Engineering Society, 2017.

[4] Blanco Galindo, M., Jackson, P., Coleman, P., and Remaggi, L., "Microphone Array Design for Spatial Audio Object Early Reflection Parametrisation from Room Impulse Responses," in *ICSV 24 Proceedings*, International Institute of Acoustics and Vibration (IIAV), London, UK, 2017.

[5] Coleman, P., Franck, A., Jackson, P. J. B., Hughes, R. J., Remaggi, L., and Melchior, F., "Object-Based Reverberation for Spatial Audio," *The Journal of the Audio Engineering Society*, 65(1/2), pp. 66–77, 2017.

[6] Remaggi, L., Jackson, P., Coleman, P., and Parnell, T., "Estimation of Object-Based Reverberation Using an Ad-Hoc Microphone Arrangement for Live Performance," in *The 144th Audio Engineering Society Convention*, Audio Engineering Society, 2018.

[7] Chitreddy, S. and Jackson, P., "Source Distance Perception with Reverberant Spatial Audio Object Reproduction of Real Rooms," in *Forum Acusticum*, p. 2079, 2020, doi:10.48465/fa.2020.0883.

[8] 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 *The 138th Audio Engineering Society Convention*, Audio Engineering Society, 2015.

[9] Stewart, R. and Sandler, M., "Database of Omnidirectional and B-format Room Impulse Responses," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 165–168, 2010, ISSN 2379-190X, doi:10.1109/ICASSP.2010.5496083.

[10] Lokki, T., Pätynen, J., Kuusinen, A., Vertanen, H., and Tervo, S., "Concert Hall Acoustics Assessment with Individually Elicited Attributes," *The Journal of the Acoustical Society of America*, 130(2), pp. 835–849, 2011, ISSN 0001-4966, doi:10.1121/1.3607422.

[11] Bacila, B. I. and Lee, H., "360° Binaural Room Impulse Response (BRIR) Database for 6DOF Spatial Perception Research," in *Audio Engineering Society Convention 146*, Audio Engineering Society, 2019.

[12] Murphy, D. T. and Shelley, S., "OpenAIR: An Interactive Auralization Web Resource and Database," in *The 129th Audio Engineering Society Convention*, Audio Engineering Society, 2010.

[13] Kim, H., Hughes, R. J., Remaggi, L., Jackson, P. J. B., Hilton, A., Cox, T. J., and Shirley, B., "Acoustic Room Modelling Using a Spherical Camera for Reverberant Spatial Audio Objects," in *The 142nd Audio Engineering Society Convention*, Audio Engineering Society, 2017.

[14] Remaggi, L., Kim, H., Neidhardt, A., Hilton, A., and Jackson, PJ., "Perceived Quality and Spatial Impression of Room Reverberation in VR Reproduction from Measured Images and Acoustics," in *Proceedings of ICA*, 2019.

[15] Cieciura, C., Mason, R., Coleman, P., and Paradis, M., "Creating Object-Based Stimuli to Explore Media Device Orchestration Reproduction Techniques," in *The 145th Convention of The Audio Engineering Society*, AES, 2018.

[16] Farina, A., "Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique," in *Audio Engineering Society Convention 108*, Audio Engineering Society, 2000.

[17] Kim, H. and Hilton, A., "3D Scene Reconstruction from Multiple Spherical Stereo Pairs," *International Journal of Computer Vision*, 104(1), pp. 94–116, 2013, ISSN 1573-1405, doi: 10.1007/s11263-013-0616-1.

[18] Majdak, P., Iwaya, Y., Carpentier, T., Nicol, R., Parmentier, M., Roginska, A., Suzuki, Y., Watanabe, K., Wierstorf, H., Ziegelwanger, H., and Noisternig, M., "Spatially Oriented Format for Acoustics: A Data Exchange Format Representing Head-Related Transfer Functions," in *Audio Engineering Society Convention 134*, Audio Engineering Society, 2013.

[19] Majdak, P., Zotter, F., Brinkmann, F., De Muynke, J., Mihocic, M., and Noisternig, M., "Spatially Oriented Format for Acoustics 2.1: Introduction and Recent Advances," *Journal of the Audio Engineering Society*, 70(7/8), pp. 565–584, 2022.