Acoustic Room Modelling using a Spherical Camera for Reverberant Spatial Audio Objects

H. Kim[1], R. Hughes[2], L. Remaggi[1], P. Jackson[1], A. Hilton[1], T. Cox[2] and B. Shirley[2]

[1]Centre for Vision Speech and Signal Processing, University of Surrey, Guildford, UK
[2]Acoustics Research Centre, University of Salford, UK

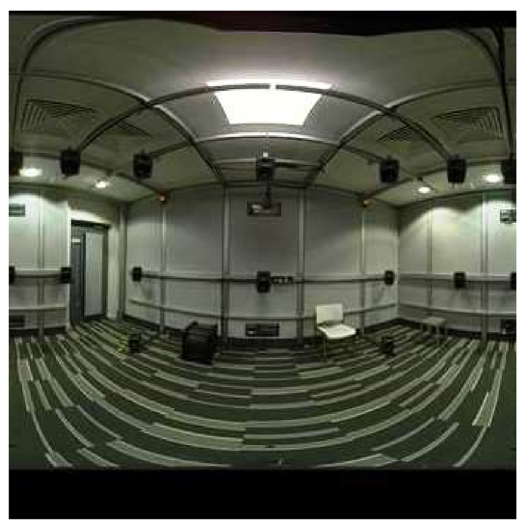
Abstract

A cuboid-based room geometry estimation method using a spherical camera is proposed, to produce frequency-dependent acoustic predictions. Results are compared to measurements through calculated reverberant spatial audio object parameters.

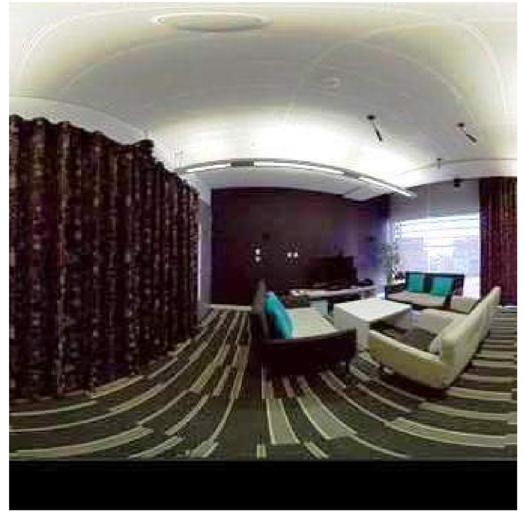
Introduction

In spatial audio reproduction, knowledge of the acoustics of a space could allow for more accurate reproduction of a captured environment. In scenarios where acoustical measurements are not available, the ability to predict the room acoustics is a useful tool. In this paper, visual capture processing are exploited.

Dataset



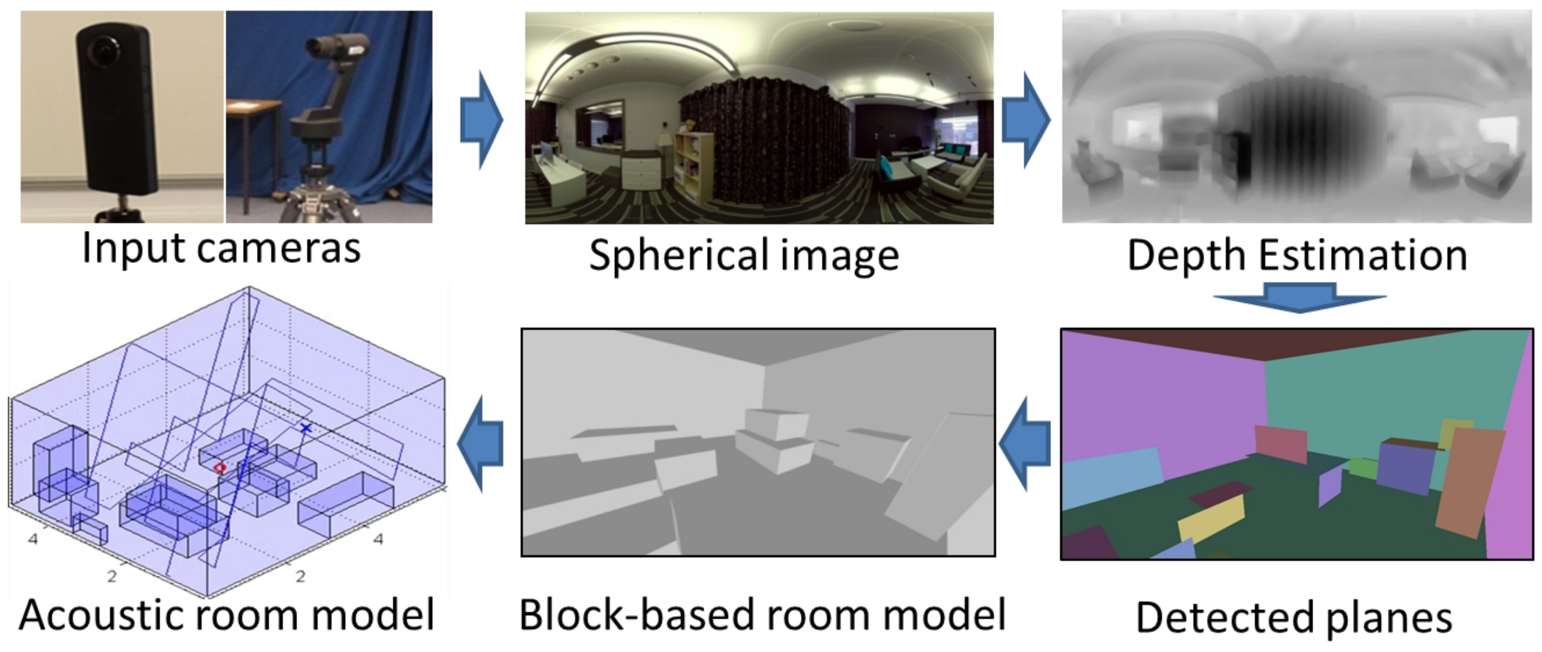
Listening Room (LR): 5.60×5.00×2.90 m³.



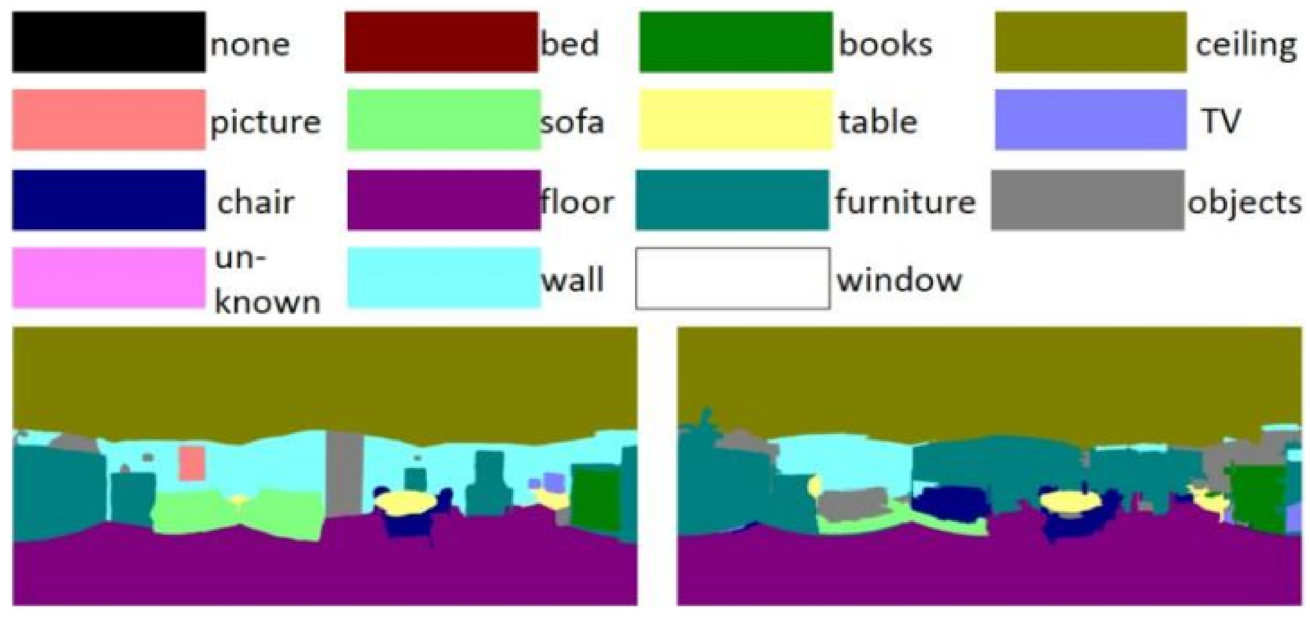
Usability Lab (UL): 5.60×5.20×2.90 m³.

Meeting Room (MR): 5.60×4.30×2.30 m³.

Room Geometry Estimation and Object Identification from Two 360° Cameras

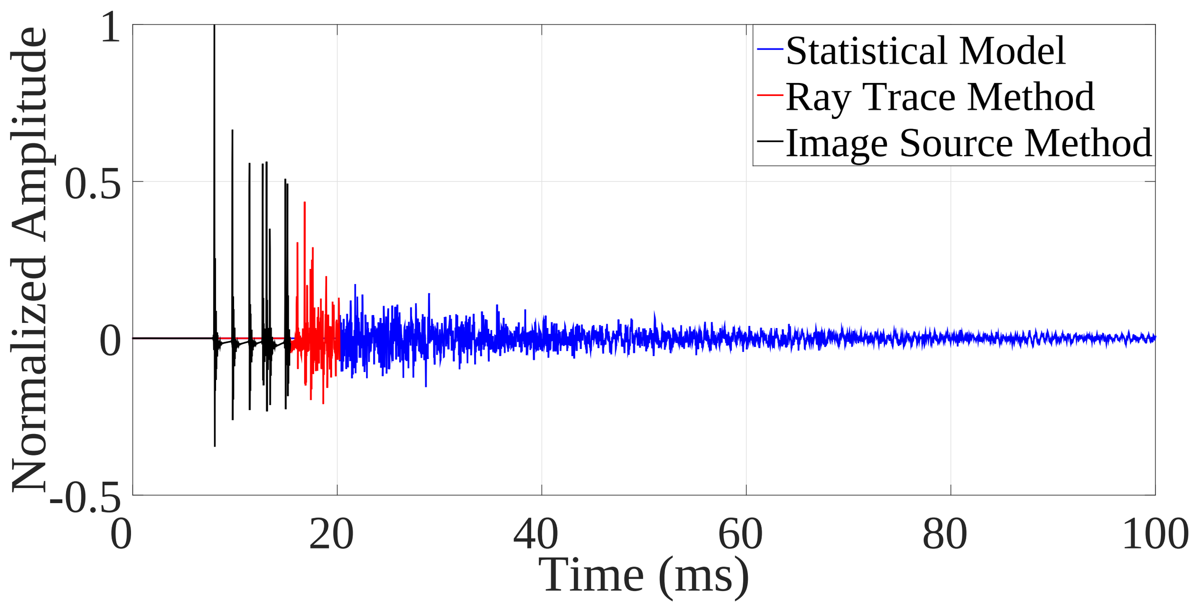


Block diagram of the proposed system. The room geometry is estimated from the depth map by calculating the disparity angle [1].



CNN architecture for semantic object classification was used to identify the objects and room boundaries [2].

Acoustic Modeling



RIR simulated by joint techniques.

Reverberant Spatial Audio Object Estimation

To evaluate the estimated RIRs, they are compared to recorded RIRs by observing their reverberant spatial audio objects [3].

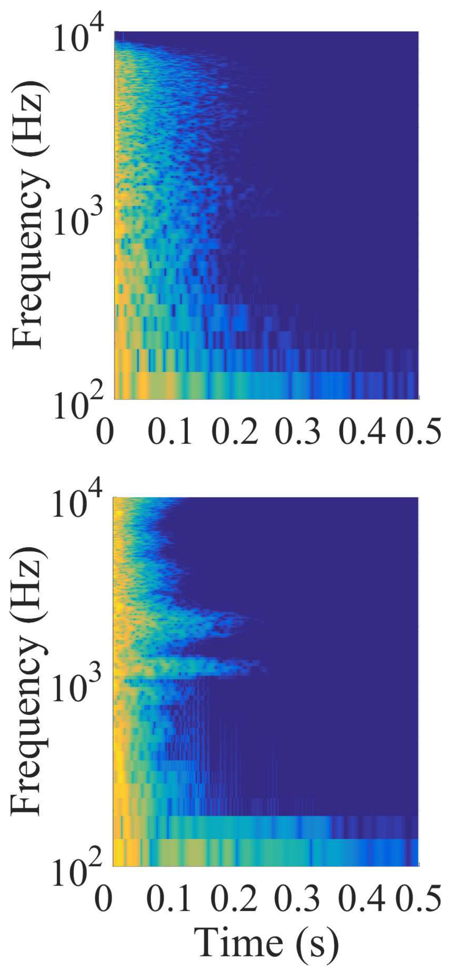
Direct Sound and Early Reflection Objects Parameters

- Times of arrival (TOAs) estimated through the clustered-dynamic projected phase-slope algorithm (C-DYPSA) [4].
- Directions of arrival (DOAs) estimated using a delay-and-sum-beamformer.

Late Diffuse Object Parameters

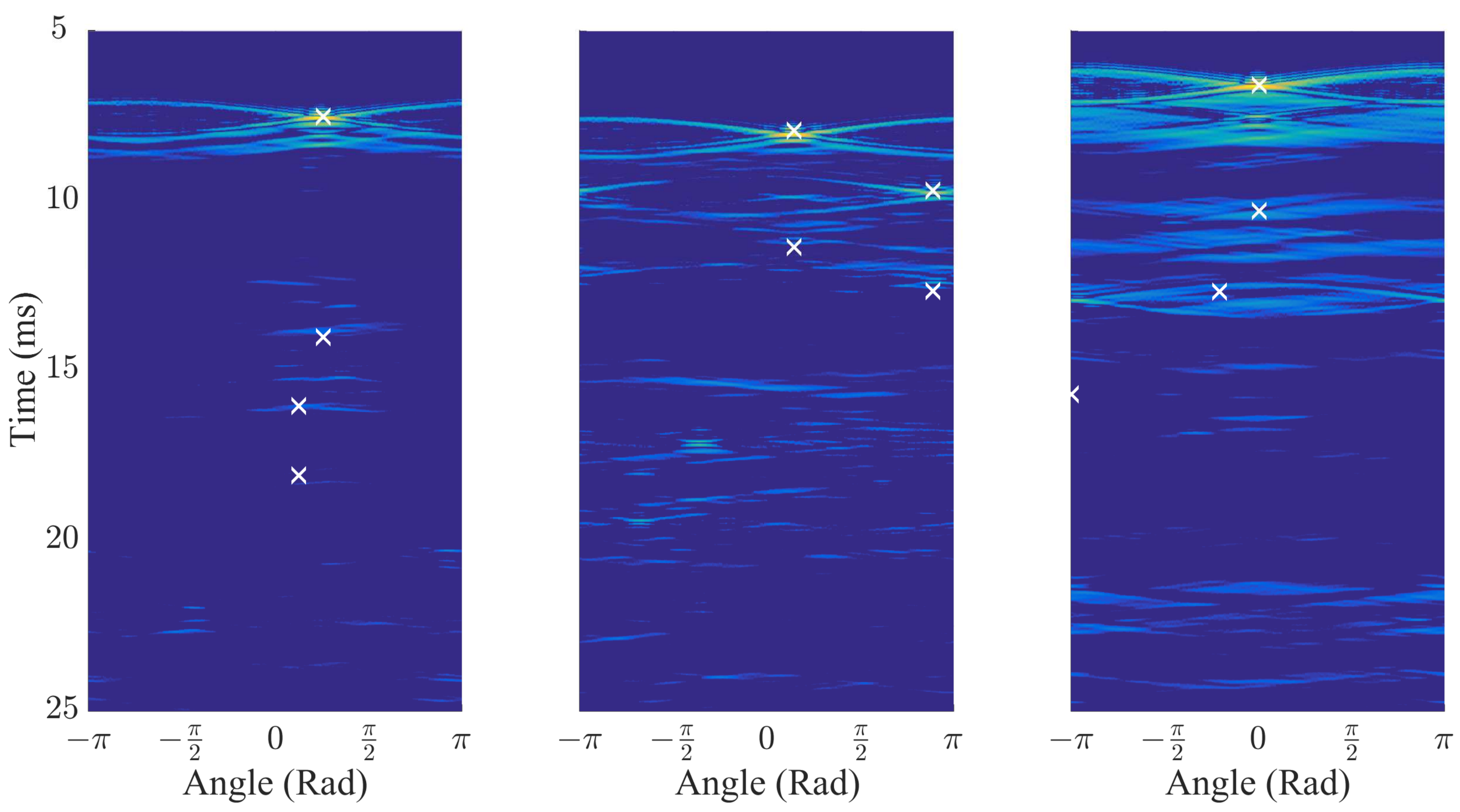
- Mixing time calculated using the model-based perceptual mixing time [5].
- Reverberation time (RT60) calculated for each octave band between 125 Hz and 2 kHz.

Synthetic vs real RIRs

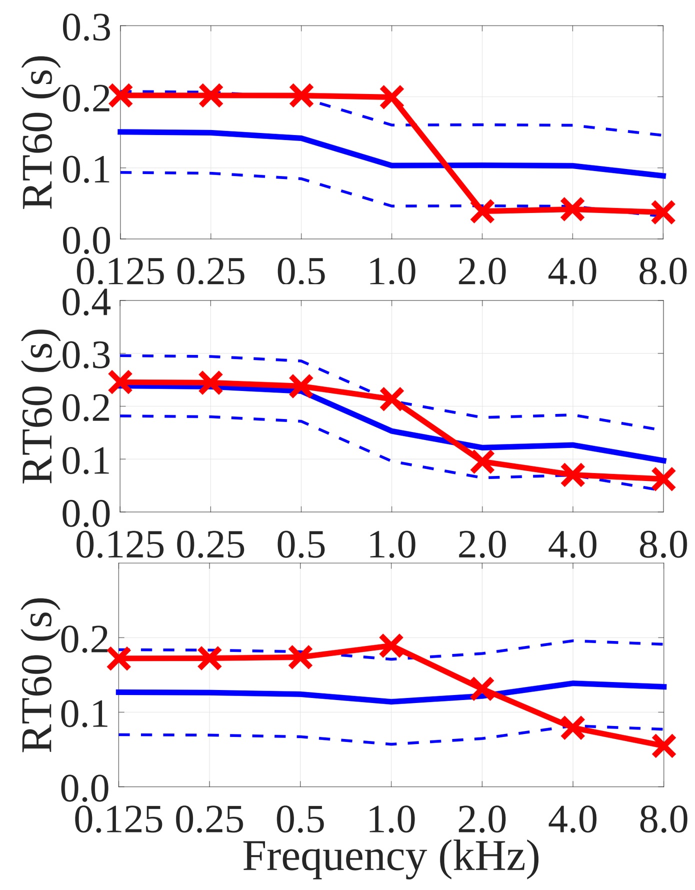


Spectrograms of the recorded (top) and synthetic (bottom) RIRs.

Experimental Results



Recorded RIRs and estimated TOA-DOA parameters (white crosses) visualization, for LR (left), UL (center) and MR (right). Blue means -30 dB, yellow 0 dB.



RT60 for simulated and recorded RIRs, for LR (top), UL (center) and MR (right).

Conclusion

A method has been outlined for prediction of RIRs based on visual information. Room geometry was estimated through vertical spherical stereo systems using commercial off-the-shelf cameras, and aligned cuboid representations of the room were reconstructed. Experiments were conducted by comparing the RSAO parameters of the estimated RIRs with the recorded ones. Results show plausible agreement between predictions obtained using estimated geometry and measurements.

References

[1] Kim et al., "Block world reconstruction from spherical stereo image pairs", Computer Vision and Image Understanding, 2015.
[2] Kim et al., "Room Layout Estimation with Object and Material Attributes Information Using a Spherical Camera", in Fourth 3DV Conference, 2016.
[3] Remaggi et al., "Estimation of room reflection parameters for a reverberant spatial audio object", 138th AES Convention, 2015.
[4] Remaggi et al., "Acoustic reflector localization: novel image source reversion and direct localization methods", IEEE TASLP, 2017.
[5] Coleman et al., "Object based reverberation for spatial audio", J. AES, 2015.