
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

McKenzie, Thomas; Schlecht, Sebastian J.; Pulkki, Ville

Acoustic analysis and dataset of transitions between coupled rooms

Published in:

ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings

DOI:

[10.1109/ICASSP39728.2021.9415122](https://doi.org/10.1109/ICASSP39728.2021.9415122)

Published: 01/01/2021

Document Version

Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:

McKenzie, T., Schlecht, S. J., & Pulkki, V. (2021). Acoustic analysis and dataset of transitions between coupled rooms. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings* (Vol. 2021-June, pp. 481-485). (IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings). IEEE. <https://doi.org/10.1109/ICASSP39728.2021.9415122>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.

ACOUSTIC ANALYSIS AND DATASET OF TRANSITIONS BETWEEN COUPLED ROOMS

Thomas McKenzie*, Sebastian J. Schlecht*[†] and Ville Pulkki*

*Acoustics Lab, Department of Signal Processing and Acoustics, Aalto University, Espoo, Finland

[†]Department of Media, Aalto University, Espoo, Finland

ABSTRACT

The measurement of room acoustics plays a wide role in audio research, from physical acoustics modelling and virtual reality applications to speech enhancement. While vast literature exists on position-dependent room acoustics and coupling of rooms, little has explored the transition from one room to its neighbour. This paper presents the measurement and analysis of a dataset of spatial room impulse responses for the transition between four coupled room pairs. Each transition consists of 101 impulse responses recorded using a fourth-order spherical microphone array in 5 cm intervals, both with and without a continuous line-of-sight between the source and microphone. A numerical analysis of the room transitions is then presented, including direct-to-reverberant ratio and direction of arrival estimations, along with potential applications and uses of the dataset.

Index Terms— Room Impulse Response, Acoustic Measurement, Coupled Rooms, Room Transition

1. INTRODUCTION

Several characteristics of reverberation are dependent on the source and microphone positions in a room, such as direct-to-reverberant ratio, early reflections and modal coupling, while reverberation time remains largely constant. However, when coupled rooms are considered, features such as double-slope decays, boundary diffraction and portaling effects emerge [1], all of which vary with inter-room position and coupling aperture size. The transition between coupled rooms is therefore a complex interaction, and poses significant challenges for acoustical modelling.

Room acoustics measurements are widely used in virtual reality systems offering six degrees-of-freedom immersive experiences, dereverberation algorithms for more efficient speech recognition [2] and even aiding reconstruction of historic monuments [3]. Measurement of the acoustic response of a room, the acquisition of the room impulse response (RIR), is typically achieved using a loudspeaker as the sound source playing an exponential sine sweep, and a microphone as the receiver [4]. RIRs measured with spherical microphone

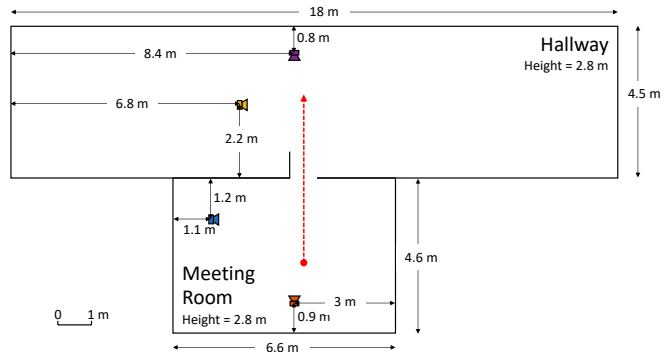


Fig. 1: Room geometry and source locations (denoted by loudspeakers) for the *Meeting Room to Hallway* transition. Microphone orientation and trajectory denoted by red arrow. Red and purple source locations have continuous line-of-sight (CLOS) between the source and microphone, blue and yellow do not (colours correspond to those in Fig. 4).

arrays, which use the principles of Ambisonics to encode microphone signals into spherical harmonic (SH) format [5], are known as spatial room impulse responses (SRIRs). These allow for greater flexibility post measurement, as they can be reproduced over loudspeaker arrays or headphones and allow directional analysis.

Recent literature has investigated how RIRs change with different receiver positions inside a single room, both for virtual reality [6] and dereverberation applications [7]. While it is possible to interpolate between measured RIRs [8], the perceptual requirements for inter-measurement distance vary with auditory stimuli, whereby sounds with limited frequency bandwidth can forgive larger distances between measurements [9], and the greater diffuseness of late reverberation allows for different measurement distances for different parts of the impulse response [6]. However, little has been published on the transition from one room to another. Due to diffraction and portaling effects from the coupling aperture between rooms, and with situations that can arise such as no continuous line-of-sight between the source and receiver, the transition is significantly more complex.

This paper presents the analysis and measurement of a dataset of SRIRs for the transition between four coupled room

The authors would like to thank the Human Optimised XR (HumOR) Project for funding this research.

pairs (see Fig 1 for an example), intended for use in acoustics and signal processing research. The paper is laid out as follows: the measurement process is detailed in Section 2, some numerical analysis of the data is presented in Section 3 including direct-to-reverberant ratio and direction of arrival estimations, and potential applications of the dataset are discussed in Section 4. A link to download the measured dataset, along with formatting information and accompanying materials, is presented in Section 5.

2. ROOM TRANSITION IMPULSE RESPONSE MEASUREMENT

Four room transitions were measured, each repeated for four different source locations: two inside each room, one of which kept a continuous line-of-sight (CLOS) between the source and microphone for the entire transition, and the second with no CLOS for a microphone in the opposite room.

The four transitions measured were:

- Meeting Room to Hallway
- Office to Anechoic Chamber
- Office to Kitchen
- Office to Stairwell

The rooms are situated in the Acoustics Lab at Aalto University, Espoo, Finland. An illustration of room geometries and source locations for the *Meeting Room to Hallway* transition is presented in Fig. 1 (corresponding figures for the other transitions are included in the full dataset).

2.1. Measurement Methodology

The impulse responses were measured using 5 second exponential sine sweeps [4]. A directional source was preferred over an omnidirectional source in order to mimic radiation characteristics for practical applications such as human speakers, therefore a Genelec 8331A coaxial loudspeaker was used for sweep playback at a peak A-weighted amplitude of 83.1 dB (measured at a distance of 1.5 m using a Sinus Tango sound pressure level meter), and an RME Fireface UCX audio interface. The SRIRs were recorded using an mh acoustics em32 Eigenmike, a 32 capsule spherical microphone array capable of fourth-order Ambisonic capture, using the proprietary Firewire interface. All audio was recorded at 24-bit resolution with a 48 kHz sample rate. Loudspeaker and microphone heights were 150 cm and 155 cm, corresponding to the approximate average height of human mouth and ear height, respectively [10]. A 5 cm distance between measurements was used for the 5 m trajectory (from 2.5 m inside the first room to 2.5 m inside the second), resulting in a total of 101 measurements per source location, which for 4 source locations totals 404 SRIRs per room coupling, and 1616 SRIRs in the whole dataset. The measurement setup for the *Meeting Room to Hallway* transition is shown in Fig. 2.



Fig. 2: Impulse response measurement setup for the *Meeting Room to Hallway* transition with a spherical microphone array and a loudspeaker. The tape on the floor indicates the microphone positions.

2.2. Deconvolution and Post-Processing

The sweeps were then encoded into fourth-order SH format with ACN component ordering [11] and SN3D normalisation [12], using the EigenUnits Encoder¹ plugin. Sweeps were exported with an additional two second tail, and impulse responses were generated through deconvolution of the encoded SH sweeps with the inverse of the original sweep. Impulse responses were then aligned temporally using onset detection based on the exceeding of a threshold magnitude, with a pre-onset gap of 5 samples (approximately 0.1 ms), before being truncated to a length of 1.5 s with in and out half-hanning amplitude windowing of 5 samples and 0.1 s, respectively. If the noise floor becomes an issue, such as when the source and receiver are in opposing rooms with no CLOS, an SRIR denoising process could be employed as in [13].

Time-domain plots of the omnidirectional SH channel (first channel of the SH files) of the *Meeting Room to Hallway* transition, for the source in the meeting room is presented in Fig. 3 (corresponding figures for the other transitions are included in the full dataset). There are clear differences between the transitions with and without CLOS. When the source is occluded (no CLOS), the direct sound is greatly attenuated in the coupled room, whereas the effect is more

¹<https://mhacoustics.com/eigenunits>

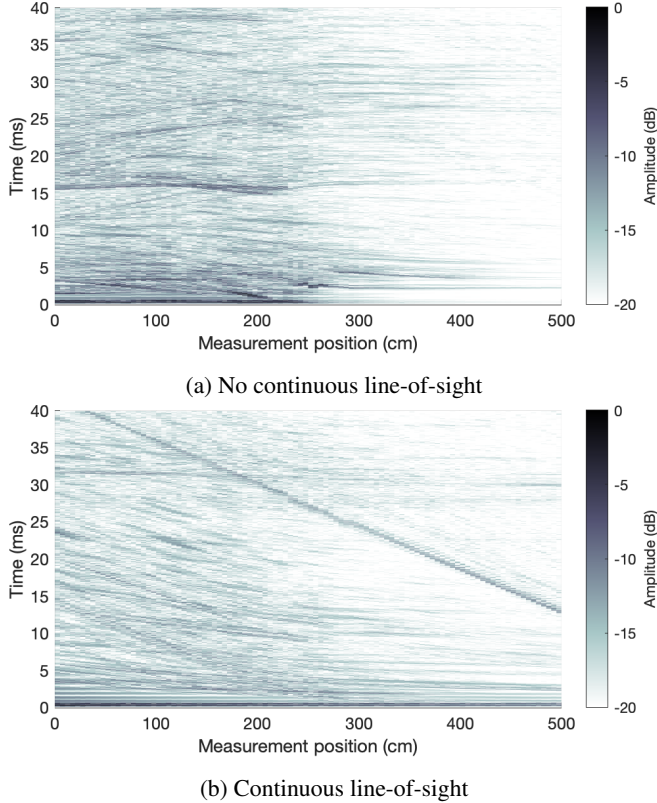


Fig. 3: Time-domain plots of the *Meeting Room to Hallway* transition (first 40 ms) for the sources in the meeting room, with and without a continuous line-of-sight between the source and microphone (see Fig. 1). Measurement position 0 cm is in the meeting room, and 500 cm in the hallway. The impulse responses are truncated such that direct sound arrives at the beginning.

gradual with CLOS. In the CLOS plot, a strong reflection arriving at a time decreasing from above 40 ms to approximately 15 ms is from the opposing room wall. There are also similar reflections from the meeting room wall between measurement positions 0 cm and 250 cm, due to the source facing the wall.

3. NUMERICAL ANALYSIS OF ROOM TRANSITIONS

3.1. Direct-to-Reverberant Ratio

The direct-to-reverberant ratio (DRR) is used to elicit acoustical properties such as source distance. The DRR of the transitions was estimated for the omnidirectional SH channel as

$$\text{DRR} = 10 \log_{10} \frac{\int_{t_0-\sigma}^{t_0+\sigma} h^2(t) dt}{\int_{t_0+\sigma}^{\infty} h^2(t) dt} \quad (1)$$

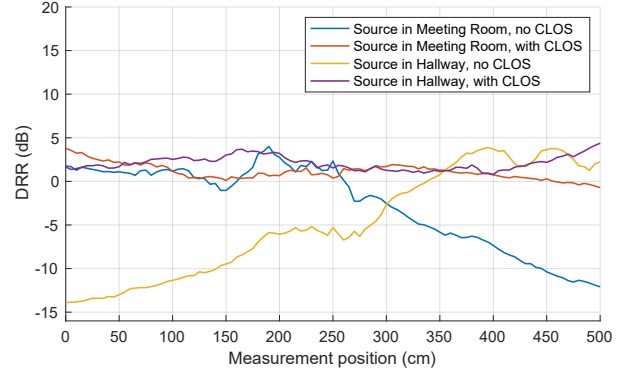


Fig. 4: Estimated direct-to-reverberant ratio of the *Meeting Room to Hallway* transition, for four measured source locations, with and without a continuous line-of-sight (CLOS) between the source and microphone (colours correspond to source locations shown in Fig. 1). Measurement position 0 cm is in the meeting room, and 500 cm in the hallway.

where h denotes the time-domain impulse response, t_0 is the time of the direct impulse and σ is the correction variable, which in this case was 10 ms [14]. Fig. 4 presents the DRR for all source locations and measurement positions of the *Meeting Room to Hallway* transition (corresponding plots for the other transitions are included in the full dataset). For the two source locations with no CLOS, the DRR decreases significantly when the microphone leaves the room with the source. This happens earlier when the source is in the hallway due to the door occluding the direct path (see again Fig. 1). For the source locations with CLOS, the DRR is relatively even apart from small increases when approaching the source, as well as when the microphone has just entered the opposing room from the source, due to the occlusion of reflections from the room containing the source.

3.2. Direction of Arrival

Direction of arrival (DOA) is a metric that can be used to measure the directivity of room reflections and direct sound of SRIRs. The DOA of the transitions was estimated above 3 kHz (due to the order dependent filtering necessary for higher order spherical microphone arrays [15]) using a fourth-order SH steered plane-wave decomposition beamformer, that calculates the power at each chosen location on the sphere [16]. DOA was estimated in one degree resolution for seven arrivals, referring to the direct sound and loudest early reflections. The horizontal DOA for all source locations and measurement positions of the *Meeting Room to Hallway* transition is presented in Fig. 5 (corresponding figures for the other transitions are included in the full dataset). Azimuth is denoted in degrees where a positive increase moves anticlockwise. Colour intensity is normalised separately for each measurement to the maximum power detected in that

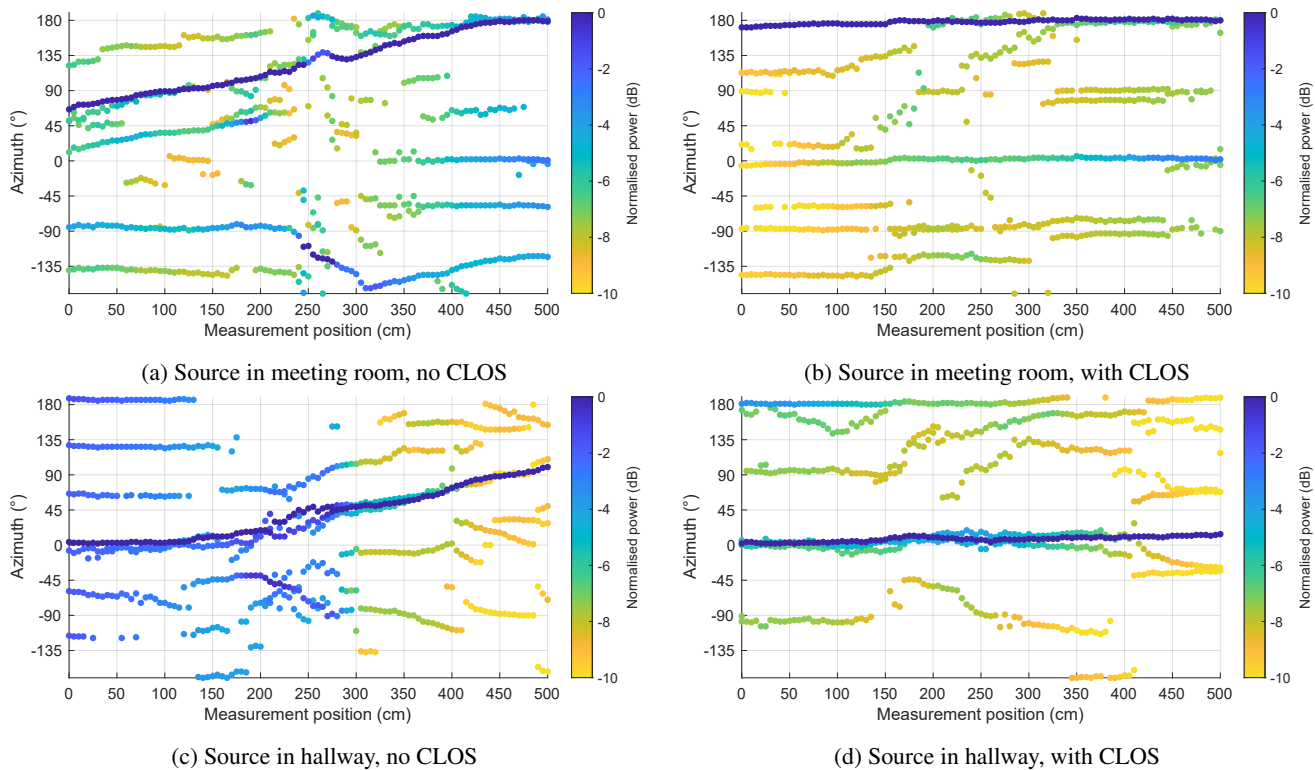


Fig. 5: Estimated direction of arrival of direct sound and early reflections for the *Meeting Room to Hallway* transition, with and without a continuous line-of-sight (CLOS) between the source and microphone (see Fig. 1). Measurement position 0 cm is in the meeting room, and 500 cm in the hallway. Azimuth values are presented from -170° to 190° for aided visibility around $\pm 180^\circ$, and colour intensity is normalised separately to each measurement’s maximum power value.

measurement, in order to illustrate the relative intensity of the dominant source direction to the other reflections. The plots demonstrate the portaling effect for the source locations with no CLOS, as the dominant arrival tends toward the median plane once through the coupling aperture. Additionally, strong reflections and changes to the relative power of arrivals occur around the coupling aperture. The plots for the source with CLOS are simpler, however, showing relatively consistent DOA, though the relative power of the reflection from the source to the opposite room’s far wall increases as the microphone approaches.

4. APPLICATIONS AND CONCLUSION

This paper has detailed the measurement of spatial room impulse responses for the transition between four coupled room pairs. Preliminary numerical analysis demonstrates the complex characteristics of the transition: direct-to-reverberant ratio decreases through the coupling aperture and with occluded line-of-sight between the source and microphone, and direction of arrival estimations show the effect of the aperture on room reflections, as well as the portaling effect when the direct sound is occluded and the source and microphone are in

opposing rooms. Potential applications of this dataset include parametric room acoustics modelling [17], dereverberation algorithms [18, 19], virtual reality [20], machine learning [21], and impulse response interpolation [22].

5. DATASET DOWNLOAD

The dataset is available under a Creative Commons license at <http://doi.org/10.5281/zenodo.4095493>, downloadable in either Spatially Oriented Format for Acoustics (SOFA) [23] or Wav format. Supplementary data includes illustrations of room geometries and source locations, DRR graphs and estimated DOA plots.

6. REFERENCES

- [1] Alexis Billon, Vincent Valeau, Anas Sakout, and Judicaël Picaut, “On the use of a diffusion model for acoustically coupled rooms,” *Journal of the Acoustical Society of America*, vol. 120, no. 4, pp. 2043–2054, 2006.
- [2] Emanuël A. P. Habets, Sharon Gannot, and Israel Cohen, “Late reverberant spectral variance estimation

- based on a statistical model,” *IEEE Signal Processing Letters*, vol. 16, no. 9, pp. 770–773, 2009.
- [3] Barteld N. J. Postma and Brian F. G. Katz, “Acoustics of Notre-Dame cathedral de Paris,” in *International Congress on Acoustics*, Buenos Aires, 2016, pp. 1–10.
- [4] Angelo Farina, “Simultaneous measurement of impulse response and distortion with a swept-sine technique,” in *AES 108th Convention*, Paris, 2000, pp. 1–23.
- [5] Michael A. Gerzon, “Periphony: with-height sound reproduction,” *Journal of the Audio Engineering Society*, vol. 21, no. 1, pp. 2–10, 1973.
- [6] Annika Neidhardt, Alby Ignatious Tommy, and Anson Davis Pereppadan, “Plausibility of an interactive approaching motion towards a virtual sound source based on simplified BRIR sets,” in *AES 144th Convention*, Milan, 2018, pp. 1–11.
- [7] Marco Jeub, Magnus Schäfer, and Peter Vary, “A binaural room impulse response database for the evaluation of dereverberation algorithms,” in *IEEE International Conference on Digital Signal Processing*, Santorini, 2009, pp. 1–5.
- [8] Niccolò Antonello, Enzo De Sena, Marc Moonen, Patrick A. Naylor, and Toon Van Waterschoot, “Room impulse response interpolation using a sparse spatio-temporal representation of the sound field,” *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 25, no. 10, pp. 1929–1941, 2017.
- [9] Annika Neidhardt and Boris Reif, “Minimum BRIR grid resolution for interactive position changes in dynamic binaural synthesis,” in *AES 148th Convention*, Online, 2020, pp. 1–10.
- [10] Max Roser, Cameron Appel, and Hannah Ritchie, “Human height,” <https://ourworldindata.org/human-height>, 2013, Accessed: 15/9/2020.
- [11] Michael Chapman, Thomas Musil, Johannes Zmöllnig, Hannes Pomberger, Franz Zotter, Alois Sontacchi, and Winfried Ritsch, “A standard for interchange of Ambisonic signal sets; including a file standard with metadata,” in *International Symposium on Ambisonics and Spherical Acoustics*, 2009, vol. 1996, pp. 1–6.
- [12] Jérôme Daniel, *Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia*, Phd thesis, University of Paris, 2001.
- [13] Pierre Massé, Thibaut Carpentier, Olivier Warusfel, and Markus Noisternig, “A robust denoising process for spatial room impulse responses with diffuse reverberation tails,” *Journal of the Acoustical Society of America*, vol. 147, no. 4, pp. 2250–2260, 2020.
- [14] Pavel Zahorik, “Direct-to-reverberant energy ratio sensitivity,” *Journal of the Acoustical Society of America*, vol. 112, no. 5, pp. 2110–2117, 2002.
- [15] Jérôme Daniel and Sébastien Moreau, “Further study of sound field coding with higher order Ambisonics,” in *AES 116th Convention*, Berlin, 2004, pp. 1–14.
- [16] Archontis Politis, *Microphone array processing for parametric spatial audio techniques*, Phd thesis, Aalto University, 2016.
- [17] Fabian Brinkmann, Hannes Gamper, Nikunj Raghuvanshi, and Ivan Tashev, “Towards encoding perceptually salient early reflections for parametric spatial audio rendering,” in *AES 148th Convention*, Online, 2020, pp. 1–11.
- [18] Yotam Peled and Boaz Rafaely, “Method for dereverberation and noise reduction using spherical microphone arrays,” in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2010, pp. 113–116.
- [19] Adrian Herzog and Emanuel A. P. Habets, “Direction and reverberation preserving noise reduction of ambisonics signals,” *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 28, pp. 2461–2475, 2020.
- [20] Franz Zotter, Matthias Frank, Christian Schorkhuber, and Robert Holdrich, “Signal-independent approach to variable-perspective (6DoF) audio rendering from simultaneous surround recordings taken at multiple perspectives,” in *Proceedings of the 46th DAGA*, 2020, pp. 1–5.
- [21] Michael Lovedee-Turner and Damian Murphy, “Application of machine learning for the spatial analysis of binaural room impulse responses,” *Applied Sciences*, vol. 8, no. 1, 2018.
- [22] Prasanga Samarasinghe, Thushara Abhayapala, Mark Poletti, and Terence Betlehem, “On room impulse response between arbitrary points: An efficient parameterization,” *International Symposium on Communications, Control and Signal Processing*, pp. 153–156, 2014.
- [23] Piotr Majdak, Yukio Iwaya, Thibaut Carpentier, Rozenn Nicol, Matthieu Parmentier, Agnieszka Roginska, Yôiti Suzuki, Kanji Watanabe, Hagen Wierstorf, Harald Ziegelwanger, and Markus Noisternig, “Spatially Oriented Format for Acoustics: a data exchange format representing head-related transfer functions,” in *AES 134th Convention*, Rome, 2013, pp. 1–11.