Dataset of Spatial Room Impulse Responses in a Variable Acoustics Room for Six Degrees-of-Freedom Rendering and Analysis

Thomas McKenzie*, Leo McCormack, Christoph Hold

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

*thomas.mckenzie@aalto.fi

Abstract—Room acoustics measurements are used in many areas of audio research, from physical acoustics modelling and speech enhancement to virtual reality applications. This paper documents the technical specifications and choices made in the measurement of a dataset of spatial room impulse responses (SRIRs) in a variable acoustics room. Two spherical microphone arrays are used: the mh Acoustics Eigenmike em32 and the Zylia ZM-1, capable of up to fourth- and third-order Ambisonic capture, respectively. The dataset consists of three source and seven receiver positions, repeated with five configurations of the room's acoustics with varying levels of reverberation. Possible applications of the dataset include six degrees-of-freedom (6DoF) analysis and rendering, SRIR interpolation methods, and spatial dereverberation techniques.

Index Terms—Spatial room impulse response, six degrees-of-freedom, variable acoustics, higher-order Ambisonics

I. INTRODUCTION

This paper details the acquisition of a dataset of SRIRs measured in the Arni variable acoustics room at the Acoustics Lab, Aalto University, Espoo, Finland – consisting of three source positions, seven receiver positions and five different configurations of the room's acoustics. The source and receiver positions were selected to simulate a possible performance space or stage, whereby receivers were distributed closer to the walls and sources were placed in the centre of the room; with the sources all pointed towards the south wall (which could be interpreted as the 'front' of the stage). Source and receiver positions are depicted in Fig. 1, with exact coordinates presented in Table I. The seven receiver positions were chosen as a line array of five receivers (R1 - R5), with an inter-receiver array distance of 1.25 m at a distance of 1.3 m from the south wall, and two near the north wall (R6 and R7). The three source positions were chosen as follows: S1 was placed slightly outside the geometry enclosed by the receiver positions (i. e. their planar convex hull), in order to produce a potentially challenging scenario for a reproduction workflow. S2 was placed within the convex hull of the receiver positions and was relatively central. S3 was placed nearer to the north wall, inside the convex hull of the receiver positions. All positions and orientations of the three sources and seven receivers were aligned to the best of the ability of the authors using laser level meters. However, small positional errors up to 2-3 cm, and orientation errorsup to 1-2° are possible.
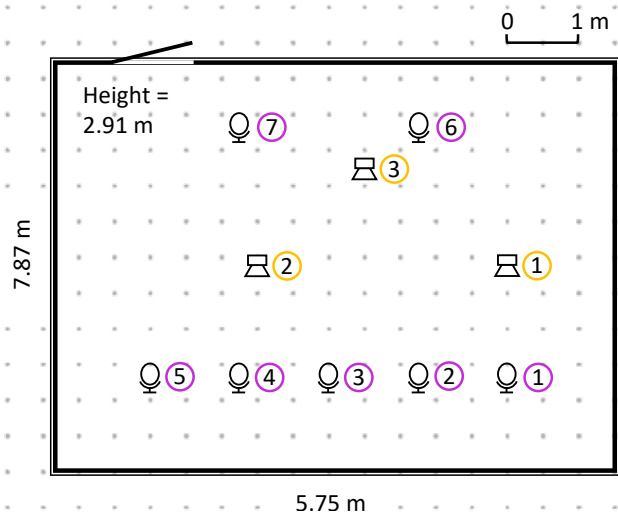

Fig. 1. Illustration of the source and receiver positions. Microphones and microphones were oriented faced north and south, respectively, according to the illustration orientation.

Loudspeakers were used as the sources, with their height (z-axis) measured from the floor to the centre of the coaxial driver. As according to Fig. 1 and Table I, the x (counting west), y (counting north) coordinates of the loudspeakers are given with respect to their stand mounting position. Note that the distance between the stand mounting and the centre of the driver along the x-axis and y-axis is 0 cm and 11.5 cm, respectively. Therefore, to align the listed loudspeaker coordinates to be with respect to the centre of the loudspeaker driver, rather than to the position of the stands, subtract 11.5 cm from the given y-axis coordinates. Microphone height was measured from the floor to the centre of the array. Source and receiver heights were 150 cm and 155 cm, corresponding to the approximate average height of human mouth and ear height, respectively [1]. For potential applications in which signals with varying amount of reverberation are important, the dataset was measured with five configurations of the variable acoustics room. The reverberation times for the five configurations [2] are presented in Table II.

A. Specification

The dataset was captured with two different spherical microphone array (SMA) models: the mh Acoustics Eigenmike em32 and the Zylia ZM-1, which are 32 and 19 capsule SMAs

TABLE I. Coordinates of the source (S) and receiver (R) positions, corresponding to the numbering in Fig. 1. To align the loudspeaker with respect to the centre of the loudspeaker driver, rather than with respect to the position of the stands, subtract 11.5 cm from the given y-axis coordinates.

Source/Receiver Number	Position (m)		
	x	y	z
S1	1.5	2.8	1.5
S2	5.0	2.8	1.5
S3	3.5	4.245	1.5
R1	1.5	1.3	1.55
R2	2.75	1.3	1.55
R3	4.0	1.3	1.55
R4	5.25	1.3	1.55
R5	6	1.3	1.55
R6	2.75	4.745	1.55
R7	5.25	4.745	1.55

TABLE II. Reverberation times of the variable acoustics room for the five absorption configurations.

% Absorbers Enabled	Reverberation Time T_{20}(s)		
	500 Hz	1000 Hz	2000 Hz
100	0.30	0.32	0.48
75	0.38	0.40	0.56
50	0.50	0.54	0.68
25	0.72	0.76	0.84
0	1.28	1.22	1.12

(a) Eigenmike SMA in position R2, 0% absorbers enabled.

(b) Zylia SMA in recording state, 75% absorbers enabled.

Fig. 2. Photographs of the loudspeakers and spherical microphone arrays with different absorption configurations.

capable of up to fourth-order and third-order Ambisonic capture, respectively. SRIRs were measured using 5 second exponential sine sweeps [3]. Genelec 8331A coaxial loudspeakers were used for sweep playback at a peak A-weighted amplitude of approximately 80 dB, with an RME Fireface UCX audio interface. The Eigenmike SRIRs were recorded using the proprietary Firewire interface, and the seven Zylia ZM-1 microphones used the on-board USB interfaces. The Eigenmike SRIRs were recorded with the same SMA, which required moving to each receiver position separately, whereas the 7 Zylia SRIRs were recorded simultaneously using the Zylia 6DoF recording application. All audio was recorded at 24-bit resolution and 48 kHz sample rate. Photographs of the recording setup with both the Eigenmike and the Zylia SMAs are presented in Fig. 2, with different variable acoustics absorption configurations.
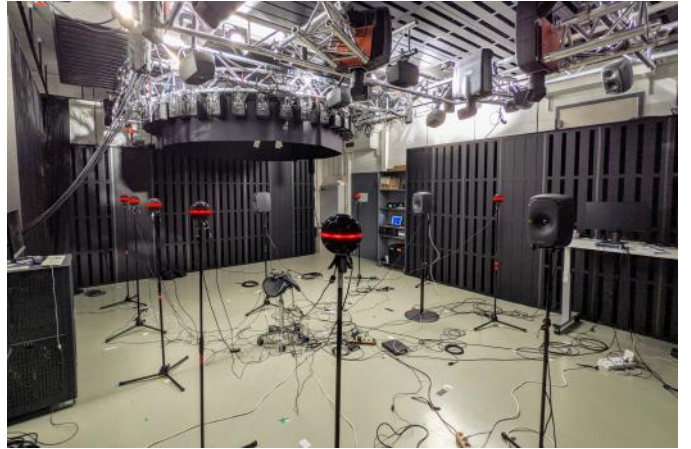
Sweep recordings were repeated, and only the cleanest sweep was used. Impulse responses were generated through deconvolution of the measured sweeps with the inverse of the original sweep. The impulse responses were aligned temporally using onset detection based on the exceeding of a threshold magnitude, with a pre-onset gap of 5 samples (approximately 0.1 ms), and then delayed based on the geometric distances from the sources to receivers. Impulse responses were truncated to a length of 2.5 s.

II. ENCODING

The raw SMA SRIRs were also converted to the spherical harmonic (SH) domain in third- and fourth-order Ambisonics using the open-source `array2sh` software tool described in [4] for both SMAs. Tikhonov regularisation during encoding limited the gain to a set maximum +15 dB. The encoded

SRIRs follow the Ambisonic Channel Order (ACN) and semi-normalised (SN3D) conventions; often collectively referred to as the AmbiX format. Note that this encoder also applies diffuse-field equalisation of the encoded signals above the spatial aliasing frequency, as recommended by Gerzon in [5]. This equalisation only affects the colouration of the signals, and not their spatial characteristics. Objective metrics, as described in [6], [7], are depicted in Fig. 3 for the Eigenmike and Zylia, which demonstrate the effect of this diffuse-field equalisation. Note that the theoretical spatial aliasing frequencies are approximately 3.3 kHz and 5.1 kHz for the Zylia and Eigenmike, respectively. The latter frequencies are also demonstrated in the spatial correlation metric as the point where the spatial correlation of the highest encoding order deviates from 1. However, it is also clear that the spatial aliasing frequency is order-dependent, and that the low-frequency cutoff for usable components (where the level difference metric deviates from 0 dB, which is a product of the regularisation) is also order-dependent. Therefore, the approximate usable frequency ranges for each SH order, which narrow with increasing order, are provided in Table III for convenience.
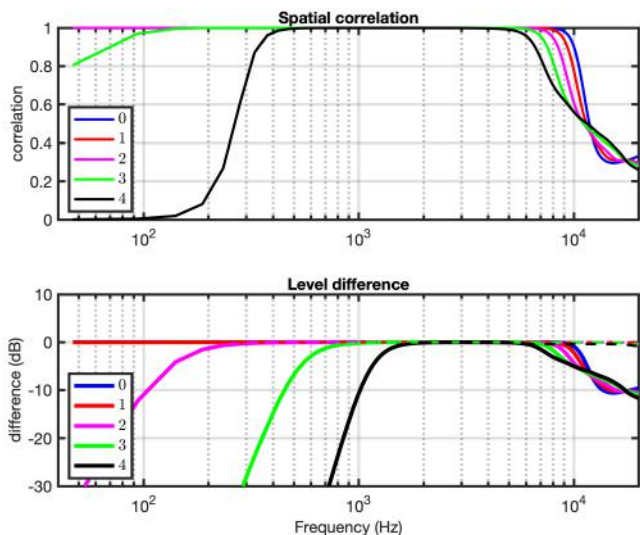
(a) Eigenmike em32 objective metrics results.
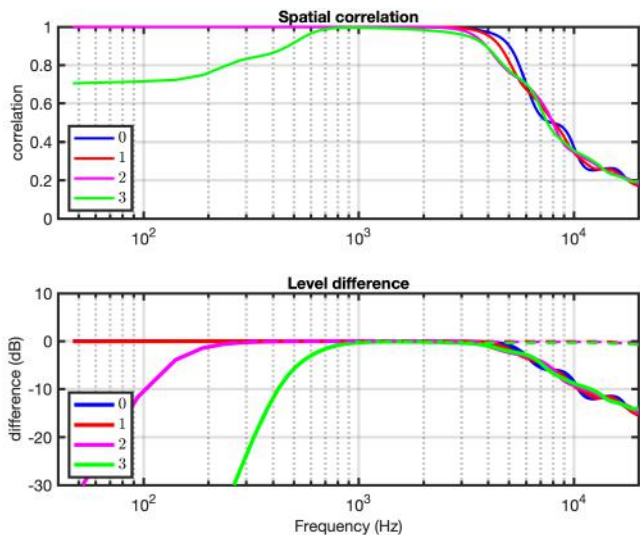
(b) Zylia ZM-1 objective metrics results.

Fig. 3. Objective metrics averaged over all components per SH order. The dotted lines in the level difference plots are after the patterns have been diffuse-field equalised above the array spatial aliasing frequency.

III. ROTATION CORRECTION FOR THE EIGENMIKE

Due to a poor fit of the Eigenmike to its shock mount, alignment of the array orientation was required. Here, the known direction of S1 from the perspective of receiver position R1 was used to derive a compensation rotation matrix to align a direction of arrival (DoA) estimate with this known true direction. The direct sound part of the (100% absorbers enabled) SRIR was isolated, and the DoA found using the active-intensity vector [8], which is derived from a first-order encoding of the array SRIR. The required compensation rotation matrix $\mathbf{R} \in \mathbb{R}^{3\times3}$ was then calculated and used to create a SH domain rotation matrix $\mathbf{R}_{\mathrm{SH}} \in \mathbb{R}^{(N+1)^2 \times (N+1)^2}$ [9]. The included SH domain conversions of the SRIR in the database have this compensation pre-applied. Since the relative Eigenmike

TABLE III. Approximate frequency bandwidths (Hz) where the SH encoded SRIRs exhibit the correct spatial patterns (spatial correlation of 1 and a level difference of 0 dB with respect to ideal SH patterns), given the employed encoding approach.

SH Order	Eigenmike em32		Zylia ZM-1	
	lower	upper	lower	upper
0th	50	8500	50	3500
1st	50	8000	50	3500
2nd	250	7000	250	3200
3rd	900	6000	900	3000
4th	1800	5100	-	-

orientations at different receiver positions were kept constant, this rotation offset is assumed to be appropriate for all 7 receiver positions. However, if using the raw/space-domain Eigenmike SRIRs for e.g. DoA estimation, then the Euler compensation rotation matrix should be used to correctly align them. The Zylia SRIR measurements do not require any rotation correction, since its forward facing direction is clearly indicated.

IV. DATASET DOWNLOAD

The dataset is available under a Creative Commons Attribution 4.0 International license at https://doi.org/10.5281/zenodo.5720724, downloadable in Spatially Oriented Format for Acoustics (SOFA) [10] format. The dataset is available as both the raw capsule recordings and encoded SH format conversions, for both the Eigenmike em32 and Zylia ZM-1. Supplementary data includes the MATLAB script and the appropriately windowed SRIR that were used for calculating the compensation rotation matrices for the Eigenmike, along with photographs and schematics of the source and receiver positions.

REFERENCES

[1] M. Roser, C. Appel, and H. Ritchie, "Human height," https://ourworldindata.org/human-height, 2013, accessed: 27/10/2021.

[2] K. Prawda, S. J. Schlecht, and V. Välimäki, "Evaluation of reverberation time models with variable acoustics," in *Sound and Music Computing Conference*, 2020, pp. 145–152.

[3] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *AES 108th Convention*, 2000.

[4] L. McCormack, S. Delikaris-Manias, A. Farina, D. Pinardi, and V. Pulkki, "Real-time conversion of sensor array signals into spherical harmonic signals with applications to spatially localized sub-band sound-field analysis," in *AES 144th Convention*, 2018.

[5] M. A. Gerzon, "The design of precisely coincident microphone arrays for stereo and surround sound," in *AES 50th Convention*, 1975.

[6] S. Moreau, J. Daniel, and S. Bertet, "3D sound field recording with higher order ambisonics–objective measurements and validation of a 4th order spherical microphone," in *AES 120th Convention*, 2006.

[7] A. Politis and H. Gamper, "Comparing modeled and measurement-based spherical harmonic encoding filters for spherical microphone arrays," in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2017, pp. 224–228.

[8] F. J. Fahy and V. Salmon, "Sound intensity," *Journal of the Acoustical Society of America*, vol. 88, no. 4, pp. 2044–2045, 1990.

[9] J. Ivanic and K. Ruedenberg, "Rotation matrices for real spherical harmonics. direct determination by recursion," *Journal of Physical Chemistry A*, vol. 102, no. 45, pp. 9099–9100, 1998.

[10] P. Majdak, Y. Iwaya, T. Carpentier, R. Nicol, M. Parmentier, A. Roginska, Y. Suzuki, K. Watanabe, H. Wierstorf, H. Ziegelwanger, and M. Noisternig, "Spatially Oriented Format for Acoustics: a data exchange format representing head-related transfer functions," in *AES 134th Convention*, 2013, pp. 1–11.