ICSV 21
13-17 July, 2014, Beijing/China

The 21st International Congress on Sound and Vibration

13-17 July, 2014, Beijing/China

MEASUREMENT AND ANALYSIS OF A SPATIALLY SAMPLED BINAURAL ROOM IMPULSE RESPONSE DATASET

Darius Satongar and Yiu. W. Lam

University of Salford, UK
e-mail: dsatongar@gmail.com

Chris Pike

BBC Research and Development, Salford, UK

This paper presents a freely-available spatially-sampled binaural room impulse response dataset, measured at the University of Salford's ITU-R BS.1116-compliant listening room. The Salford BBC Spatially-sampled Binaural Room Impulse Response dataset (SBSBRIR) was measured at a number of listening positions within the room. The dataset includes measurements of 12 loudspeakers, positioned at ear-height, using a head-and-torso simulator with a 2° head- azimuth resolution and 15 different listener positions. The dataset can be used in the subjective and objective evaluation of domestic spatial audio reproduction. Measurement details, procedure and initial validation tests will be presented alongside the application of the dataset for future work.

1. Introduction

Assessment of spatial audio reproduction systems often focuses on the central listening position (CLP) or *sweet spot*. However, when used in domestic listening environments many listeners will not be seated centrally within the loudspeaker array. It is therefore important to assess reproduction outside of the CLP. Blind subjective assessment of the effect of changing listening position is not possible in-situ, since it involves moving the subject. Binaural simulation using measured binaural room impulse responses (BRIRs) could be used instead, such measurements may also be used for objective analysis of sound reproduction at different listening positions. This paper presents the measurement and application of a freely available spatially-sampled binaural room impulse response dataset (SBSBRIR) measured at the University of Salford in collaboration with BBC Research and Development.

2. Measurement Details

The dataset includes measurements of 12 loudspeakers, positioned at ear-height, using a Brüel & Kjær (B&K) head-and-torso simulator (HATS) at 15 different listener positions, rotated with a 2° head-azimuth resolution. The measurement procedure and setup details are documented in this section.

2.1 Geometry

Figure 1 shows the measurement setup with the loudspeakers and measurement positions in the room. BRIR measurements were made at the filled points (•) giving 15 measurement positions. Head-azimuth angle is taken from the HATS pointing towards loudspeaker 1 (0° azimuth or position (0,0)m) and positive head-azimuth values represent anti-clockwise rotation of the head between 0° and 359°. All listening positions are denoted in metres relative to the central listening position. The loudspeaker and dummy head ear height was set at 1.06m measured from the floor. The diagram shows the doors and the soft furnishing at the rear of the room.

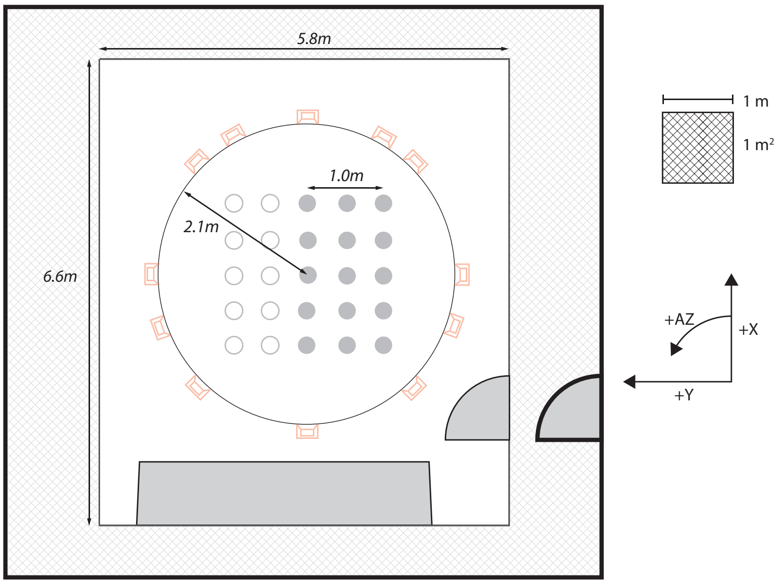

Figure 1. Plan view of the measurement setup and coordinate system used in the dataset.

2.2 Measurement Technique

Due to the large number of measurements required for the dataset, efficiency was considered when designing the measurement procedure, in order to reduce the measurement time and also avoid redundancy in the dataset. All measurements were made at a sampling rate of 48kHz using an RME UFX audio interface with an RME ADI-8 converter for six of the loudspeaker outputs.

The exponential swept-sine method was used to measure the impulse responses [1]. Signals emitted from each loudspeaker were overlapped to reduce the measurement time for each head-azimuth direction, c.f. [2]. The sweep measurements were controlled remotely using MATLAB. The measured impulse responses were post-processed to remove measured hardware and software delays. The HATS was automatically rotated in the horizontal plane in 2° steps using a B&K Turntable System Type 9640 controlled from MATLAB via a GPIB interface. This meant that the listening room could be sealed throughout all measurements at each listening position.

2.3 Equipment and Environment

The listening room was chosen due to its BS-1116-1 [3] conformity. The room has 6.6m x 5.8m x 2.8m room has a mid-frequency reverberation time of 0.27s and background noise of 5.7dBA. The listening room can also be used to carry out listening tests on loudspeakers according to BS 6840-13/IEC 268-13.
Twelve Genelec 8030A loudspeakers were measured in the dataset, positioned on a circle with a radius of 2.1m. Positive azimuth moves anti-clockwise around the circle, where 0° represents (2.1,0)m,

as noted in Figure 1. The loudspeaker azimuth positions used were: 0°, 30°, 45°, 90°, 110°, 135°, 180°, 225°, 250°, 270°, 315°, 330°. These loudspeaker positions were chosen to include five reproduction system layouts: stereo, ITU 5.0 [3], a square (with centre front loudspeaker), a square (with no centre front loudspeaker) and an octagon. Levels were set by generating pink noise and aligning each loudspeaker gain to achieve 74.5dB SPL at centre listening position. No equalisation was applied to compensate for the loudspeaker response.

Calibrated B&K type 4190 microphones were used in a B&K HATS Type 4100. This HATS has no ear canal simulator, microphones were mounted at the entrance of the closed ear canal. Alongside the BRIR measurements, an omnidirectional measurement microphone was used to measured room impulse responses from each loudspeaker at all 25 listening positions. This can be used in objective analysis of the dataset, to characterise the room, without the influence of the HATS.

3. Subjective Validation: Localisation

An auditory virtual environment (AVE) has been developed using the SBSBRIR dataset, to simulate spatial audio reproduction in an existing auditory environment. The purpose of this simulation is to allow testing of various domestic spatial audio reproduction methods at multiple listening positions in a direct blind comparison. Spatial audio reproduction systems are often tested at the CLP but the difficulty of repositioning subjects means that assessing variation in sound quality across a listening area is uncommon.

To validate use of an AVE for this purpose, it should be shown that any artifacts present during off-centre listening (as caused by time-of-arrival changes from loudspeakers, loudspeaker directivity effects, and room effects) are maintained. An initial step towards this validation is to assess whether listeners' ability to localise sound sources is consistent between the real in-situ scenario and the AVE simulation. In this paper localisation refers to the direction-of-arrival of an auditory event in the horizontal plane, it does not include the distance or elevation of the event.

A localisation test was undertaken in which participants were asked to indicate the direction-of-arrival of sound sources under different 'auralisation' methods: (1) In-situ, where real loudspeakers reproduced the phantom sources and (2) AVE (binaural), where a head-tracked dynamic binaural system, simulating loudspeakers in the listening environment, reproduced phantom sources. Spatial audio reproduction methods were used to create sound sources over a chosen loudspeaker layout using monophonic audio items.

3.1 Independent Variables

A selection of reproduction systems, source directions and source stimuli were chosen as shown in Table 1 and were tested at both a central (0,0)m and non-central (-0.5,-0.5)m listening position, in order to assess the performance of the AVE under a broad range of realistic domestic reproduction scenarios. Alongside stimuli emitted from a single loudspeaker, amplitude-panning techniques Vector Base Amplitude Panning (VBAP) [6] and Ambisonics were implemented on a selection of five loudspeaker layouts using three different audio items. Ambisonic panning coefficients were calculated using a velocity decode method. This was implemented by taking Moore-Penrose pseudo-inverse of the re-encoding matrix C [7, p. 159]. The re-encoding matrix contains the spherical harmonic coefficients corresponding to the direction of each loudspeaker in the selected array.
The three source stimuli used in the test were: Noise - repeated pink noise bursts with rectangular window, 1s long [500ms noise, 500ms silence]; Music - repeated piano scale extract, 8s long; Voice - repeated female spoken voice, 28s long.

Table 1. Definition of samples according to independent variables. Numerical subscripts indicate Ambisonic order. CF (centre front) and NCF (not centre front) denote angular orientation of the square/octagonal layouts.

Sample	Reproduction Method	Loudspeaker Layout	Phantom Source Position	Audio Item
01	VBAP	Stereo $\pm$ 30	15 $^\circ$	Noise
02	$Ambisonic_1$	Square (CF)	0 $^\circ$	Voice
03	Mono	110 $^\circ$	-	Music
04	$Ambisonic_3$	Octagon (CF) 30	45 $^\circ$	Music
05	VBAP	Octagon (CF)	100 $^\circ$	Voice
06	$Ambisonic_2$	Square (CF)	115 $^\circ$	Noise
07	Mono	315 $^\circ$	-	Voice
08	VBAP	ITU 5.0	290 $^\circ$	Voice
09	VBAP	ITU 5.0	190 $^\circ$	Music
10	$Ambisonic_1$	Square (NCF)	0 $^\circ$	Noise

It is important to note that the capabilities of different reproduction methods (VBAP, Mono, Ambisonics), loudspeaker layouts, audio items and listening positions are not the focus for this test; rather the comparison of the localisation results under a selection of these variables between in-situ or AVE auralisation is of interest.

3.2 Pointing Method

The egocentric (head- or nose-pointing) technique [9] was chosen, where, upon hearing an auditory event, participants turn to face in the perceived direction of the sound source, pointing with a laser pointer attached to their head - a trigger button is then pressed by the participant to record their judgement and begin the next stimulus presentation. The laser pointer was attached to the head-set and could be centrally aligned to their gaze to reduce over- or under-shooting the perceived direction [10]. This was done by mounting the laser on an adjustable ball-joint to reduce the caused by varied headphone positioning [11]. A potential disadvantage of this pointing method, is that it primarily measures changes in frontal localisation acuity. However the method allows for more accurate reporting of direction and recording of biomechanical data can be used to analyse effects on the localisation process. Carlile et al. [9] highlight the benefits of this method, it is a natural action and head-tracking can be performed feasibly. An optical motion tracking system (4 VICON Bonita cameras and Tracker software) was used to track the participants' head position both for analysis and as input to the AVE rendering software. The tracking system can be used to capture biomechanical data with high precision and accuracy. Additional dependent variables can therefore be considered in the analysis, including analysis of head rotation patterns and translational head movement.

3.3 Procedure

Participants were volunteers from the University of Salford Acoustics Research Centre. All had experience with acoustics or audio and considered themselves to be *"audio experts"* when asked in a pre-test questionnaire. There were 15 participants in the test. Participants were given an instruction guide on the test procedure. They were then guided into the listening room in which loudspeakers were hidden behind an acoustically transparent curtain. Participants were given a controller with a button for submitting localisation decisions and a knob for audio volume control. They were allowed to adjust the volume at any point in the test. In total, 120 stimuli were presented (2 systems, 2 listening positions, 10 samples with 3 repeats per sample). The order of stimuli was randomised for each subject. After a training session, participants performed the test. The training consisted of a short trial test until participants felt comfortable with the method. No feedback on localisation performance

was given.
Importantly, a calibration stage was performed to allow alignment of coordinate systems of the Vicon tracking system, the AVE system, and the physical room geometry. This was done by asking the participant to point the head-mounted laser at a marker placed at 0° on the curtain.

3.4 Auditory Virtual Environment: Binaural Reproduction System

The AVE uses BRIRs from the SBSBRIR dataset in a dynamic binaural renderer with head azimuth tracking. A modified version of the SoundScape Renderer (SSR) [12] was used for real-time BRIR convolution with loudspeaker input signals, received from a Max/MSP patch, which controlled the test. Vicon tracking data was sent over the OSC protocol [13] to the Max/MSP and SSR software. Stax SR-207 electrostatic headphones were used in the test. Headphone compensation filters were applied to reduce the effect of the headphone-to-ear transfer function (HpTF). HpTF measurements were made on the B&K HATS and the compensation filters were applied to the BRIRs offline, before use in real-time rendering.

3.5 Results

3.5.1 Localisation Error

Letowski and Letowski [14] highlight that the mean unsigned error (MUE) between perceived source direction and real source direction gives a general approximation of localisation error (LE), encompassing both precision and accuracy into a single measure. Mean signed error (ME) and standard deviation (SD) of the signed error distribution can be used to approximate precision and accuracy respectively. Figure 2 shows the MUE with 95% confidence intervals for all samples, listening positions and auralisation methods. Table 2 shows the MUE, ME and SD across all samples and subjects, with each listening position shown independently.

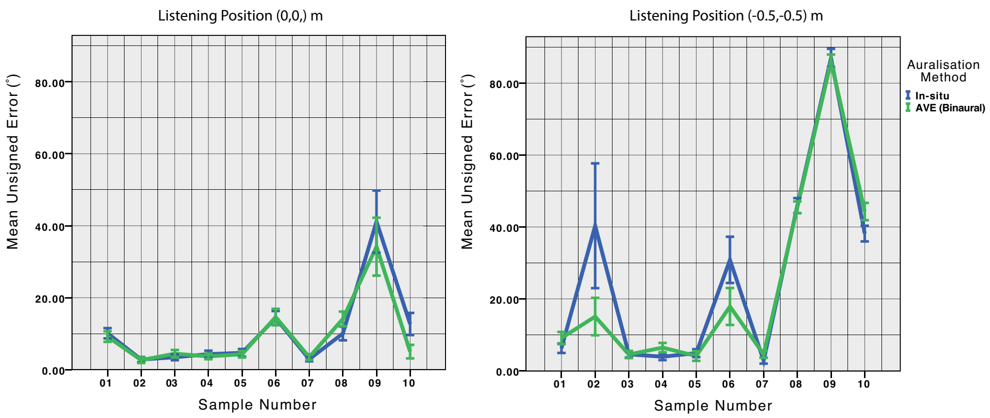

Figure 2. The mean unsigned error for in-situ and AVE auralisation methods at listening positions (0,0)m and (-0.5,-0.5)m across all subjects.

3.5.2 Time-of-Judgement

Each localisation judgement can be described by the head movements made by the participant during the stimulus presentation. The time taken to make judgements can be analysed to get an approximation of the difficulty of the localisation task and make comparisons between the two auralisation methods. If a judgement takes longer, we can assume that the localisation task was more challenging or complex. Figure 3 shows the mean time-of-judgement (ToJ) values for each sample

Table 2. Localisation error across all samples and subjects for in-situ and AVE auralisation at (0,0)m and (-0.5,-0.5)m listening positions.

Listening Position	Auralisation Method	MUE (°)	ME (°)	SD (°)
(0,0)	In-situ	10.7	1.1	18.4
(0,0)	AVE	9.6	-3.0	16.1
(0,0)	*In-situ - AVE*	*1.1*	*4.1*	*2.3*
(-0.5,-0.5)	In-situ	26.5	-4.3	42.0
(-0.5,-0.5)	AVE	23.8	-0.8	36.0
(-0.5,-0.5)	*In-situ - AVE*	*2.7*	*-3.5*	*8.0*

and auralisation method at each listening position.

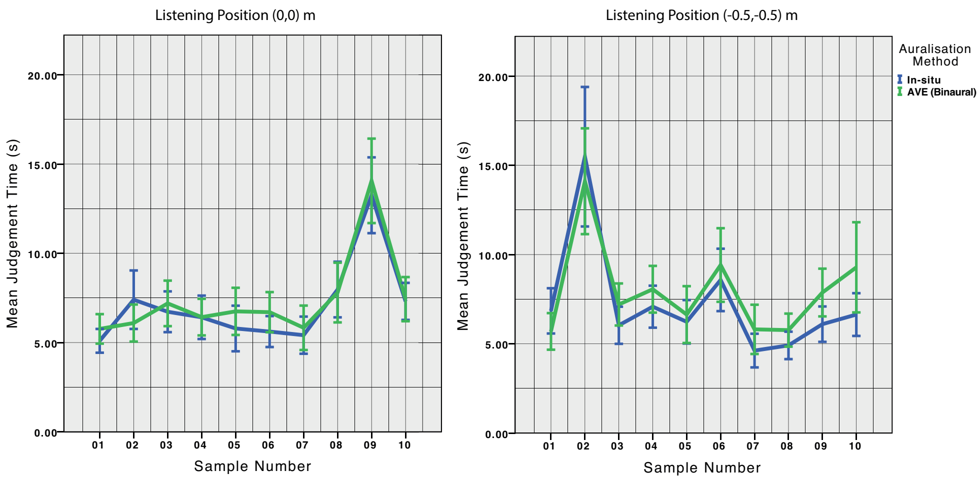

Figure 3. The mean time-of-judgement for in-situ and AVE auralisation methods at listening positions (0,0)m and (-0.5,-0.5)m across all subjects. All plots show 95% confidence intervals.

3.6 Discussion

MUE differences between the in-situ and AVE systems appear to be small with respect to the variance in MUE across samples. The observed MUE differences are also similar to the 0.9° value presented in [15] when testing real vs HRTF loudspeaker sound sources; the MUE magnitudes for each method are much larger here however, due to the use of amplitude-panning algorithms rather than just single loudspeakers. SD and ME differences also seem to be relatively small, the SD at (-0.5,-0.5)m shows the largest difference. The SD differences between auralisation method are larger than shown in [15]. This is again thought to have been caused by the use of stimuli that are difficult to localise, due to the amplitude panning applied.

Further analysis of the samples with the largest LE revealed that, in certain circumstances, judgement error distributions were multi-modal. In particular in-situ results showed a more even distribution between modes, this could have been due to the collapsing of a phantom image into a loudspeaker with head translation. In these scenarios, Gaussian models of localisation error are less appropriate. There is clearly a significant difference in MUE between auralisation methods for samples 2 and 6 at the off-centre listening position (-0.5,-0.5)m, which were both created using Ambisonic panning. The reasons for this are not clear, further investigation is required.

For ToJ there is a clear similarity in the values shown in Figure 3. This is a surprising but reassuring result which suggests similar complexity of the localisation task between the two auralisation methods. Both systems exhibit the same sample-dependent variations in ToJ at both listening positions. Results with increased ToJ values also correspond closely with values of reduced localisation accuracy from Figure 2, supporting the assertion that ToJ relates to task complexity.

4. Access and Use of the SBSBRIR Dataset

The dataset presented in this paper is freely available for download from the University of Salford Institutional Repository[1]. SBSBRIR by the University of Salford and the British Broadcasting Company is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License[2].

The dataset is saved in the Measured Impulse Response Object (*miro*) format[3] for MATLAB™, as described in [17]. The *miro* class enables a large amount of technical and positional information to be stored in a conveniently accessible package with functions included to convert the data to a number of output formats. There are a total of 180 *miro* files; one file per loudspeaker at each of the 15 listening positions. Each *miro* file contains 360 BRIRs (in 1° head azimuth steps) and one omni-directional room IR. BRIRs have been linearly interpolated to 1° head-azimuth resolution.

Any contributions or amendments to the dataset will be welcomed to the correspondence email address.

5. Summary

A dataset of spatially-distributed binaural room impulse response measurements has been presented, recorded at the University of Salford. The purpose of which is to provide a dataset for assessing domestic loudspeaker-based spatial audio systems at non-central listening positions using binaural analysis and simulation.

This dataset was also used in the validation of an AVE through a localisation test. The test highlighted that in general differences in localisation error between in-situ and AVE presentation were small and comparable with a previous study. However, for certain samples created using Ambisonics, significant differences were measured at the off-centre listening position. Further study is needed to identify the cause of this difference. Time-of-judgement was considered and found to match well between to the two auralisation methods for all samples at both listening positions.

Following these validation tests, the AVE will be used to investigate the perception of spatial audio systems at non-central listening positions in blind direct-comparison tests. Alongside localisation, other auditory cues influenced by changing listening position will be considered.

REFERENCES

1 A. Farina, *Advancements in impulse response measurements by sine sweeps* in 122nd Audio Engineering Society Convention, 2007.

[1] `http://usir.salford.ac.uk/30868/`
[2] `http://creativecommons.org/licenses/by-nc-sa/4.0/`
[3] Available at: `http://www.audiogroup.web.fh-koeln.de/`

[2] P. Majdak, P. Balazs, and B. Laback, *Multiple Exponential Sweep Method for Fast Measurement of Head-Related Transfer Functions* in J. Audio Eng. Soc., vol. 55, no. 7/8, pp. 623–637, 2007.

[3] ITU, *Methods For The Subjective Assessment Of Small Impairments In Audio Systems Including Multichannel Sound Systems* (Rec. ITU-R BS.1116-1) 1997.

[4] P. Novo, *Auditory Virtual Environments* in Communication Acoustics, J. Blauert, Ed. Berlin, Germany: Springer-Verlag, 2005.

[5] ITU, *Multichannel stereophonic sound system with and without accompanying picture* (Rec. ITU-R BS.775-2), 2006.

[6] V. Pulkki, *Virtual Sound Source Positioning Using Vector Base Amplitude Panning* J. Audio Eng. Soc., vol. 45, no. 6, pp. 456–466, 1997.

[7] J. Daniel, *Représentation de champs acoustiques, application á la transmission et á la reproduction de scénes sonores complexes dans un contexte multimédia* University of Paris, 2001.

[8] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. London, UK: The MIT Press, 2001.

[9] S. Carlile, P. Leong, and S. Hyams, *The nature and distribution of errors in sound localization by human listeners* Hearing Research, vol. 114, pp. 179–196, 1997.

[10] J. Lewald, G. J. Dörrscheidt, and W. H. Ehrenstein, *Sound localization with eccentric head position* Behav. Brain Res., vol. 108, no. 2, pp. 105–25, Mar. 2000.

[11] H. Wierstorf, A. Raake, and S. Spors, *Localization of a virtual point source within the listening area for Wave Field Synthesis* in 133rd Audio Engineering Society Convention, 2012, vol. c, pp. 1–9.

[12] J. Ahrens, M. Geier, S. Spors. *The SoundScape Renderer: A Unified Spatial Audio Reproduction Framework for Arbitrary Rendering Methods* in 124th Audio Engineering Society Convention, Amsterdam, The Netherlands, 2008.

[13] M. Wright and A. Freed. *Open Sound Control: A New Prtocol for Communicating with Sound Synthesizers* Proceedings of the 1997 International Computer Music Conference,

[14] T. R. Letowski and S. T. Letowski, *Auditory Spatial Perception : Auditory Localization* 2012.

[15] H. Wierstorf, S. Spors, and A. Raake, *Perception and evaluation of sound fields* in 59th Open Seminar on Acoustics, 2012.

[16] M. Rychtarikova, T. Van Den Bogaert, G. Vermeir, and J. Wouters, *Binaural Sound Source Localization in Real and Virtual Rooms* J. Audio Eng. Soc., vol. 57, no. 4, pp. 205–220, 2009.

[17] P. Stade, B. Bernschütz and M. Rühl, *A Spatial Audio Impulse Response Compilation Captured at the WDR Broadcast Studios* in 27th Tonmeistertagung - VDT International Convention, 2012.