PAN-AR: A Multimodal Dataset of Higher-Order Ambisonics Room Impulse Responses, Ambient Noise and Spherical Pictures

Filippo Denti
filippo.denti@studenti.unimi.it
Laboratory of Music Informatics (LIM), University of Milan
Department of Computer Science
Milan, Lombardy, Italy

Davide Fantini
davide.fantini@unimi.it
Laboratory of Music Informatics (LIM), University of Milan
Department of Computer Science
Milan, Lombardy, Italy

Federico Avanzini
federico.avanzini@unimi.it
Laboratory of Music Informatics (LIM), University of Milan
Department of Computer Science
Milan, Lombardy, Italy

Giorgio Presti
giorgio.presti@unimi.it
Laboratory of Music Informatics (LIM), University of Milan
Department of Computer Science
Milan, Lombardy, Italy

ABSTRACT

This work presents PAN-AR, a dataset of spatial room impulse responses (SRIRs) encoded in the higher-order Ambisonics (HOA) format. The dataset encompasses measurements obtained in four distinct environments: a printer room, a meeting room, a classroom, and an underground parking area. The SRIRs have been recorded for different configurations of the emitter and receiver positions, including up to six emitter positions and two receiver positions. The dataset also includes samples of the ambient noise and spherical pictures captured at the receiver positions, along with other metadata. These characteristics render PAN-AR well-suited for extended reality applications, as it enables the dynamic simulation of virtual sound sources in the space surrounding the listener. Moreover, the dataset favors the creation of ecological virtual auditory environments, as it encompasses scenarios that are commonly encountered in the everyday lives of workers and students. In addition, the ambient noise recordings and the spherical pictures provide a more immersive and realistic experience. Finally, an analysis of the acoustical properties of the measured SRIRs is provided in the paper.

CCS CONCEPTS

• **Applied computing** → **Sound and music computing**; • **Human-centered computing** → *Mixed / augmented reality*; *Virtual reality*.

KEYWORDS

Dataset, reverberation, room impulse responses, Ambisonics, extended reality, spatial audio, virtual acoustics, spherical images

AM '24, September 18–20, 2024, Milan, Italy

ACM ISBN 979-8-4007-0968-5/24/09
https://doi.org/10.1145/3678299.3678332

ACM Reference Format:
Filippo Denti, Davide Fantini, Federico Avanzini, and Giorgio Presti. 2024. PAN-AR: A Multimodal Dataset of Higher-Order Ambisonics Room Impulse Responses, Ambient Noise and Spherical Pictures. In *Audio Mostly 2024 - Explorations in Sonic Cultures (AM '24), September 18–20, 2024, Milan, Italy.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3678299.3678332

1 INTRODUCTION

The sound waves emitted in an environment reflect on the encountered surfaces resulting in the persistence in time of sound. This phenomenon, known as reverberation, is influenced by the position in space and the materials of such surfaces. This results in specific characteristics of the reverberated sound, including intensity, duration in time, timbre, and spatial distribution. Artificial reverberation is the process of simulating an acoustic reverberation effect. Since the 1960s, a number of digital methods have been developed to this end [37]. Convolution-based techniques represent a specific category of these methods, which assume that a linear and time-invariant (LTI) system models the reverberant environment. Accordingly, a room impulse response (RIR) is recorded in the environment of interest for specific emitter and receiver positions, and its reverberation effect is simulated via convolution with an arbitrary audio signal. Despite the considerable computational load of the convolution operation, the obtained reverberation accurately simulates the effect perceived in the real environment. For this reason, convolution-based techniques are well-suited for auralization purposes, i.e., the rendering of a source soundfield in an acoustic environment, by simulating the listening experience at a given position in the modeled space [23]. For a realistic auralization, the virtual simulation of a sound source in the environment of interest should consider different aspects. Multiple RIRs should be measured in the same environment by considering different configurations of the emitter and receiver positions. Further, the measurement of an RIR with a monophonic microphone is inadequate for a realistic auralization process. In the physical world, a listener perceives the various directions of arrival of the sound waves. In the field of spatial audio, this phenomenon is virtually replicated through the use of a number of techniques developed to simulate immersive virtual auditory environments (VAEs) [2, 16]. In the context of convolution-based techniques, the use of spherical microphone arrays allows the acquisition of spatial RIRs (SRIRs), which capture the directional properties of the incoming sound waves. This enables the dynamic simulation of the auralization process, whereby the spatial audio rendering can be adapted in accordance with the listener's rotation. One of the principal formats for encoding spatial audio signals is Ambisonics [42].

RIRs are indispensable tools in the study of various fields including virtual acoustics, extended reality (XR), music production, teleconferencing, speech enhancement, reverb matching, dereverberation, indoor localization, auditory scene analysis, hearing-impairment research, and so on. For instance, in teleconferencing systems, RIRs can be employed to simulate the presence of remote speakers in the same room as the listener [1, 19, 41]. Further, some approaches to reverberation matching utilize RIRs as a target to simulate the reverberation of a given environment [5, 6, 20, 34]. Moreover, RIRs have been employed to study, design, and test hearing aids [7, 21, 22, 29, 40]. In the context of XR, SRIRs are of particular interest as they permit the creation of VAEs, which adapt to the user's movements. The design of VAEs that are experienced as real as possible is a hot research topic. SRIRs can be employed to investigate the extent to which listeners are capable of discerning between reference and simulated sound sources. Several criteria can be used to this end, including authenticity [4], plausibility [25], transfer-plausibility [39] and co-immersion [35]. Further, these criteria can be evaluated under different settings of the reality-virtuality continuum [28]. For instance, audio augmented reality (AAR) involves the seamless superimposition of virtual sound sources onto existing ones [3, 15, 31], while audio augmented virtuality (AAV) entails the creation of VAEs using real-world auditory content [10].

The variety of applications encompassed by RIRs has led to the publication of numerous datasets, each designated to address specific use cases. In this paper, we present PAN-AR (Panoramas, Ambient Noise & Ambisonics RIRs), a dataset of higher-order Ambisonics (HOA) SRIRs, in particular second-order Ambisonics, which is well-suited, but not limited, to applications in the XR context. The SRIRs have been recorded for different configurations of receiver and emitter positions in four rooms: a printer room, a meeting room, a classroom, and an underground parking area. The selected environments and positions were chosen to capture ecologically-valid scenarios, which are representative of typical experiences of workers and students in everyday life. Further, we considered at least one configuration in each room where the receiver and emitter positions have been exchanged. The dataset also includes additional data such as ambient noise samples, spherical pictures, and other metadata pertaining to the environments. These additional data are provided with the intention of obtaining a more comprehensive representation of the captured environments. The ambient noise allows for a more immersive simulation within the XR scene, while the spherical pictures provide a visual stimulus that is consistent with the listener's position in space. The PAN-AR dataset is publicly available on Zenodo [8]. Further, we also provide the SRIRs in SOFA format [26] following the SingleRoomSRIR convention[1].

[1]https://sofacoustics.org/data/database/pan-ar/

This paper is structured as follows. Section 2 provides an overview of similar SRIRs datasets presented in the literature. Section 3 describe the dataset PAN-AR including the measured environments, the measurement setup, the post-processing operations, and the dataset organization. Section 4 provides an acoustical analysis of the measured SRIRs. Section 5 concludes the paper.

2 PREVIOUS WORK

The existing literature contains numerous examples of RIRs datasets, reflecting the considerable interest in the field of artificial reverberation. Nevertheless, only a minority of these datasets encompass SRIRs, which allows for a dynamic auralization that accounts for the listener's rotation. These datasets typically provide SRIRs encoded in Ambisonics, which is the prevailing format for spatial audio content. Table 1 shows an overview of some of the Ambisonics SRIRs datasets that have been proposed in the literature. These datasets exhibit some distinguishing characteristics. A dataset is composed of multiple SRIRs recorded in one or more environments of interest. The majority of the datasets consider only one environment. Some of these datasets employ varying furniture arrangements to alter the reverberation properties of a room, for example, using absorption panels [17, 27]. Conversely, other datasets measure SRIRs in multiple rooms [9, 38], such as the one described in this paper. Typical measured environments include concert halls, meeting rooms, classrooms, living rooms, and pubs, among others. In each environment, one or more configurations of the emitter and receiver positions can be considered. In case of multiple environments, SRIRs are typically measured for only one configuration [38], or slightly more [9]. On the contrary, when a single environment is considered, several configurations of emitter and receiver positions are usually encompassed. In general, the number of emitter positions is greater than that of the receiver [17, 18, 24, 32, 33], but this is not always the case [9, 27, 30]. The most prevalent microphone used for the acquisition of Ambisonics SRIRs is the Eigenmike em32, which permits the encoding of up to fourth-order [9, 17, 24, 27, 32, 33]. However, alternative microphones [18, 27, 30] or custom-designed solutions [38] have been employed. In addition to SRIRs, such datasets may also include other Ambisonics recordings in the considered environments. These may include babble [9], ambient noise (sometimes referred to as background noise or room tone) [9, 38], and musical instruments [24]. A minority of the dataset also includes visual data about the environments, such as spherical photographs [17] and 3D models [17, 18].

The analysis of the existing datasets, as presented in Table 1, indicates that our dataset, PAN-AR, has some advantages over the others. PAN-AR is the sole dataset that includes both ambient noise recordings and spherical pictures of the environments. Furthermore, it is only one contemplating multiple environments along with multiple emitter and receiver positions within each environment.

3 DATASET

3.1 Measured environments

The PAN-AR dataset includes Ambisonics SRIRs that were measured in four environments of the Department of Computer Science of the University of Milan (Italy) in 2023. The environments are depicted in Fig. 1 and include a printer room, a meeting room, a classroom, and

Table 1: Comparison of some of the published SRIRs datasets in Ambisonics format.

Dataset	Ambisonics order	Number of environments	Environment type	Number of positions Emitter	Number of positions Receiver	Device Emitter	Device Receiver	Sample rate [kHz]	Other recordings	Visual data
ACE Challenge [9]	4th	7	Office, meeting and lecture rooms, building lobby	1	2	Fostex 6301B	Eigenmike em32	48	Babble, fan and ambient noise	No
METU SPARG [32]	4th	1	Classroom	244	1	Genelec 6010A	Eigenmike em32	48	No	No
ARTE [38]	4th	13	Office, cafè, library, living room, etc.	1	1	Tannoy V8	Custom	44.1	Ambient noise	No
3D-MARCo [24]	4th	1	Concert hall	13	1	Genelec 8331A	Eigenmike em32	48	Musical instruments	No
Schütze et al. [33]	4th	1	Living room	15	2	Genelec 8030	Eigenmike em32	48	No	Visual model
Motus [17]	4th	1	Room with varying furniture	4	1	Genelec 1030A	Eigenmike em32	48	No	3D models, spherical photo
McKenzie et al. [27]	4th, 3rd	1	Room with varying furniture	3	7	Genelec 8331A	Eigenmike em32, Zylia ZM-1	48	No	No
Grimm et al. [18]	1st	1	Pub	28	4	Genelec 8030, 8020, B&K 4295	Core Sound TetraMic	48	No	3D models
HOMULA-RIR [30]	2nd	1	Seminar room	2	25	Genelec 8020D	Spatial Mic Dante	48	No	No
PAN-AR (our)	2nd	4	Printer and meeting rooms, classroom, underground parking	≤ 6	≤ 2	Focal Alpha 65	Core Sound OctoMic	192, 96	Ambient noise	Spherical photo

an underground parking area. These spaces were specifically chosen to capture the acoustic properties of ecological scenarios, which are experienced in the everyday lives of workers and students. For instance, the meeting room represents a well-suited environment for immersive teleconferencing applications, whereas the classroom is indicated for distance learning or remote workshops. Further, we selected spaces with acoustic properties that differed as much as possible, while avoiding an excessive number of environments. For instance, the environments encompass a broad range of scales, from a small-sized room, such as the printer room, to a vast area such as the underground parking. In particular, the printer room has an irregular shape, with the longest dimension measuring less than 5 meters. The meeting room has a rectangular base, with a length of approximately 6 meters. The classroom has an almost rectangular base, with a length of around 13 meters. The underground parking has an irregular shape, with the longest dimension measuring more than 40 meters. The exact dimensions of each environment are provided in the planimetries reported in the dataset.

For each environment, we selected six positions, with the exception of the printer room, where five positions were selected due to the limited space. Fig. 2 depicts the planimetries of the four environments, along with the selected positions. The positions are labeled with the letters A through F (A through D for the printer room). The positions were selected to obtain uniform coverage of the environment. Given these positions, we considered six (five for the printer room) configurations of the emitter and receiver positions used to measure the SRIRs. Five (four for the printer room) configurations were obtained by placing the receiver in position A and measuring the SRIR with the emitter placed in each of the remaining positions. Position A was positioned to be roughly at the room center in the majority of the environments. The remaining configuration was obtained by exchanging the positions of the receiver and emitter: the emitter was positioned in A, whereas the receiver was positioned in one of the remaining positions. The position selected for the receiver in this configuration is marked with the letter X as a subscript in the planimetries shown in Fig. 2. The heights of the positions were selected to represent ecological scenarios where the emitter and receiver positions correspond to the heads of speakers and listeners, which may be either standing or seated. The receiver was oriented towards the north cardinal direction for all the measurements. In some cases, the emitter was oriented toward the receiver, while in others, it was oriented in a different direction. The orientation of the emitter for each SRIRs is provided in the dataset along with other metadata. Some of the emitter positions were selected in order to elicit the reverberation of adjacent environments and, in some cases, to occlude the direct path to the receiver completely or partially. For position A, we performed recordings of the ambient noise for 3 minutes capturing the sound of the environment without any explicit sound source. Further, we took spherical pictures in the two positions where the receiver was placed in each environment.

3.2 Measurement setup

The SRIRs were measured with the logarithmic sine sweep technique [12]. The impulse responses were obtained through deconvolution of the signal captured by the receiver. The employed sine sweeps were 15 seconds long, ranging from 10 Hz to 95.9 kHz with a sample rate of 192 kHz. The sweeps were propagated in

(a) Printer room (b) Meeting room (c) Classroom (d) Underground parking

Figure 1: Pictures of the four environments measured in PAN-AR dataset, each showing a configuration of the emitter and receiver positions.

the environment through the emitter represented by a Focal Alpha 65 loudspeaker (see Fig. 3a). The loudspeaker was connected to a laptop via the Motu UltraLite-MK3 Hybrid audio interface. Then, the reverberated sweep was captured through the receiver represented by a Core Sound OctoMic microphone (see Fig. 3b), placed perpendicular to the floor. The OctoMic is composed by a spherical microphone array of eight capsules allowing up to second-order Ambisonics recordings. The OctoMic is a mixed-order 2H1V microphone, which means that the horizontal and vertical planes are covered with a second-order and first-order spatial resolution, respectively [36]. The multichannel recordings were performed with the Zoom F8n multitrack recorder connected to the OctoMic. We recorded the audio signals with 192 kHz sample rate and 24 bit depth. The same measurement procedure was followed for the ambient noise. The devices' gains were calibrated to achieve a sufficiently intense signal, without clipping, and they were coherent within each environment. The exact gain settings for both playback and recording devices can be found in the dataset's metadata. Finally, the spherical pictures were taken through a smartphone in equirectangular projection format with a resolution of 10240×5120 pixels and with the center of the image facing the north cardinal direction.

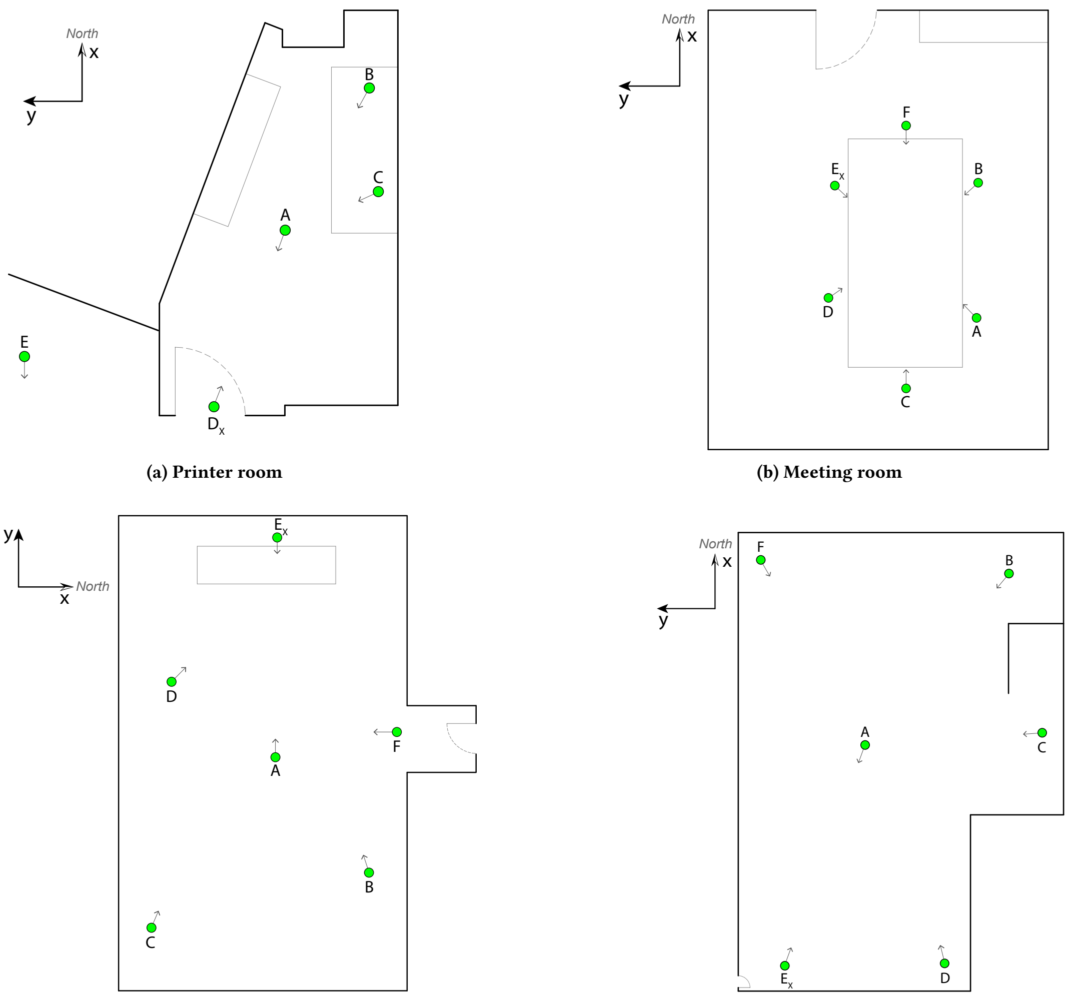
3.3 Post-processing

The raw recordings performed by the Zoom F8n are A-format audio files at 192 kHz with eight channels. To obtain the SRIRs provided in the dataset, some post-processing operations were conducted. First, we applied the deconvolution operation to obtain the impulse responses from the recorded sine sweeps. Then, the recordings were downsampled to 96 kHz, and the onsets preceding the SRIRs were removed. It is possible to reconstruct the sound propagation delay according to the distances between the emitter and the receiver reported in the dataset's metadata. The A-format files were then converted in B-format obtaining the second-order Ambisonics SRIRs having nine channels. The conversion was conducted with the VVOctoEncode plugin by VVAudio[2]. The first calibration option provided for the OctoMic microphone was selected during the encoding procedure to correct its response. This option is the one providing the least processing noise and excellent spatial location cues as indicated by the microphone's manufacturer[3]. The encoding was performed using the AmbiX format encompassing the conventions ACN for channel ordering and SN3D for normalization. The same conversion from A-format to B-format has been conducted for the ambient noise recordings. Also, ambient noise recordings have been cleaned to remove sporadic impulsive noises and edited to allow for seamless playback when looped. The spherical pictures have been converted to cubic projections and then post-processed to adjust wrong junctions and to apply color correction. Finally, the pictures were encoded again in equirectangular coordinates. Fig. 4 shows a spherical photo obtained for the printer room.

3.4 Dataset organization

The dataset is organized following the directory structure shown in Fig. 5. Each directory in the root is dedicated to one of the four environments. Inside each environment directory, there is a directory for each of the versions provided for the recorded audio files. The directory `A_Format_192kHz` includes the raw version in A-format at the original sample rate of 192 kHz. The directory `B_Format_96kHz_AmbiX` include the versions at 96 kHz encoded in HOA Ambisonics using the AmibiX format. Inside each of these directories, the SRIRs and the ambient noise recordings are stored in the `IRs` and `Ambient_noise` directories, respectively. The naming convention used for the audio files is *env_pos_type_format*`.wav`, where *env* is the environment number from `1` to `4`. *pos* is the emitter position from the letter `B` to `F`, plus the letter `X` for the configuration where the receiver position has been exchanged with the emitter position. *pos* is omitted in case of ambient noise. *type* is the type of recording that can be `IR` or `Ambience` for the SRIRs and the ambient noise, respectively. *format* is the audio format that can be `A` for the

[2] https://www.vvaudio.com/products/VVOctoEncode
[3] https://www.core-sound.com/products/octomic



(a) Printer room (b) Meeting room (c) Classroom (d) Underground parking area

Figure 2: Planimetries of the four environments measured in the PAN-AR dataset, showing the considered positions and the orientations of the emitter and receiver. The receiver is always pointing to the north and in position A, except for the configuration where the emitter and the receiver positions are exchanged; in this configuration, the emitter is in position A, while the receiver is in the position marked with the letter X in subscript.

A-format and `B_AmbiX_opt1` for the B-format AmbiX encoding using the first option for microphone calibration.

Further, for each environment directory, we include the directory `Images` with images related to the environments. Inside each `Images` directory, there are the directories: `Panoramas` including the spherical pictures taken in the two receiver positions in the environment, `Planimetry` including the environment planimetry, `Pictures` including some pictures taken in the environment showing the emitter and receiver positions. Additionally, we included in the root two CSV files, namely `IRs.csv` and `Ambient_noise.csv`, regarding all the environments. These files include additional information on SRIRs and ambient noise recordings. For each measured SRIR, `IRs.csv` reports the distance and the angle between the emitter and receiver, the gain of the audio interface that reproduced the sine sweep, the gain of the multitrack recorder, and additional notes regarding the measurement conditions. The last two metadata are also provided in `Ambient_noise.csv` for the ambient noise recordings.

4 SRIRS ANALYSIS

This section presents an analysis of the SRIRs reported in the dataset. This analysis encompasses a set of acoustical parameters obtained using the AURORA software [11, 13] in accordance with the ISO standard 3382 [14]. These parameters were computed for each environment and each recording configuration, based on the omnidirectional channel of the SRIRs. This analysis is intended to provide a characterization of the acoustic properties of each environment. The selected acoustical parameters include reverberation time T_{20},

(a) Focal Alpha 65 (b) Core Sound OctoMic

Figure 3: Pictures of (a) the emitter, a broad-band loudspeaker, and (b) the receiver, a 8-capsules microphone, which were used for the SRIRs measurement.

Figure 4: Spherical picture of the printer room after the post-processing operations.

early decay time EDT, center time T_S, and clarity index C_{80}. The parameter T_{20} is a measure of the reverberation time, which is quantified by tripling the time employed by the RIR to decay from −5 to −25 dB. The early decay time EDT estimates the decay time of the early reflections as the time required by the RIR to decay by 10 dB. The center time T_S represents the center of gravity of the squared RIR. The clarity index C_{80} is the energy ratio between the early reflections and the late reverberation parts of the RIR. This is computed assuming that early reflections and late reverberation are divided at 80 ms.

Table 2 reporters the values of the acoustical parameters for each position in each environment. In addition, Fig. 6 shows the mean values for each environment computed per one-third octave band. It can be observed that the reverberation time T_{20} is consistent with the size of the environments. The shortest reverberation time was observed in the printer room, while the longest was observed in the underground parking. Nevertheless, the meeting room exhibits a slightly longer reverberation time than the classroom, despite the smaller size. This can be attributed to the configuration of the meeting room, which has a shoebox shape and almost no furniture but a table and some chairs. Conversely, the classroom has an irregular shape, is partially acoustically treated, and has a considerable number of seating arrangements, which serve to diffuse and absorb the sound reflections. The large size of the underground parking may also affect the clarity index C_{80}, which is notably low in this environment. The underground parking is also the environment with the largest values for the EDT and T_S parameters, while the other environments exhibit similar values. The acoustical parameters of the printer room exhibit a generally larger standard deviation compare to the other environments. This can be attributed to the SRIR recorded with the emitter situated outside the room, that is in position E. For example, this position exhibits a long reverberation time, which is likely the result of the reverberation of the spaces adjacent to the printer room.

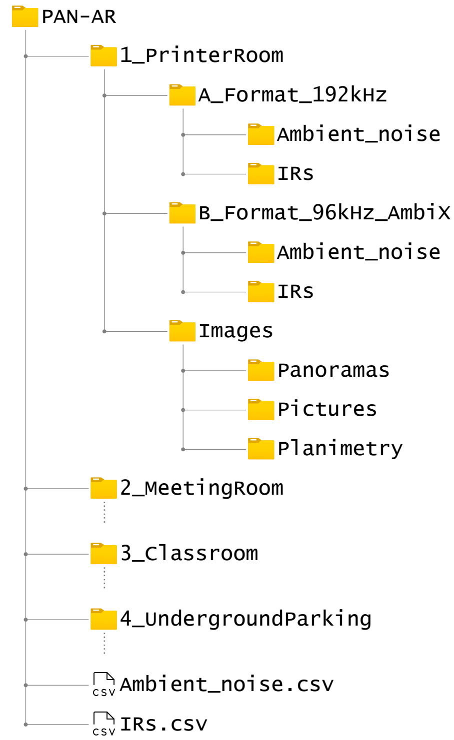
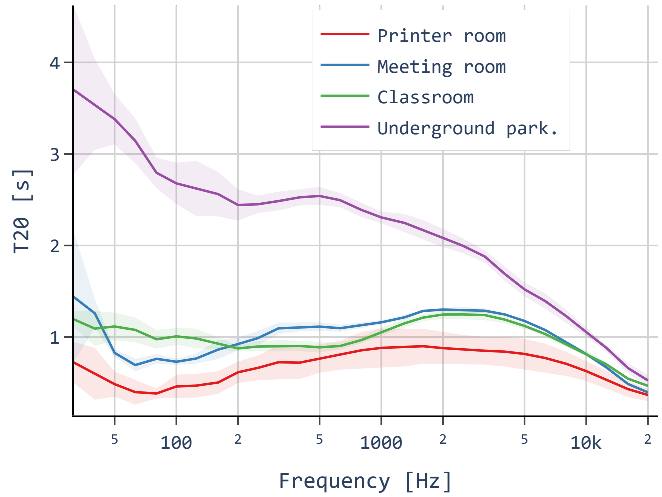


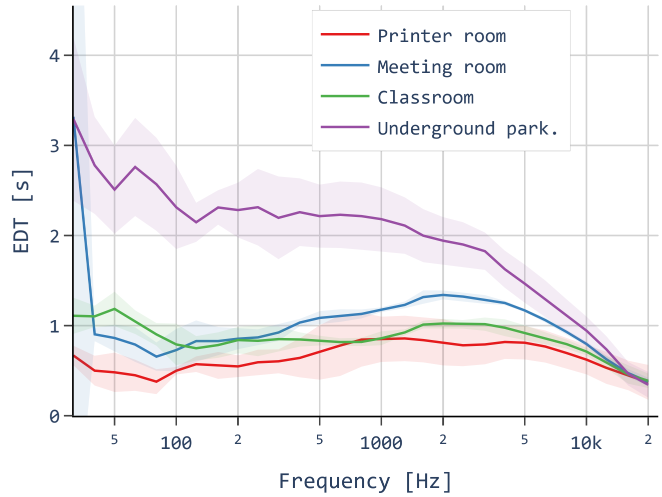
Figure 5: Directory tree representing the dataset organization.

5 CONCLUSIONS

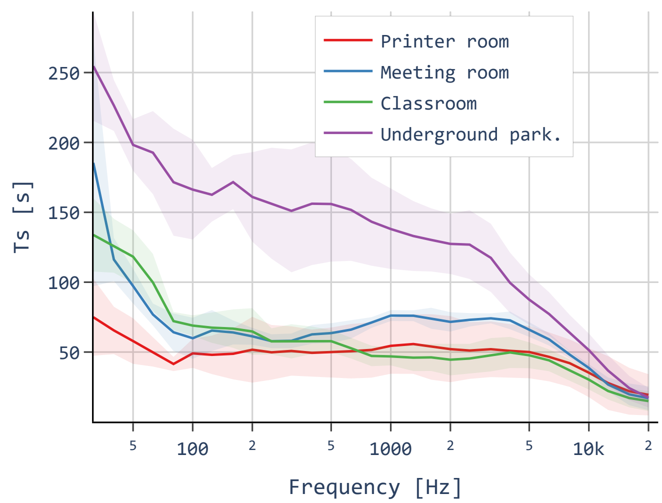
In this paper, we presented PAN-AR, a dataset of SRIRs encoded in second-order Ambisonics. The dataset encompasses measurements of a printer room, a meeting room, a classroom, and an underground parking area. For each environment, the SRIRs have been recorded for six configurations of the emitter and receiver positions. One of these configurations is obtained by exchanging the receiver position with one of the emitter positions. Additionally, PAN-AR includes samples of the ambient noise and spherical pictures captured for the receiver positions. PAN-AR also includes metadata to interpret the SRIRs and ambient noise measurements. Further, we provide the SRIRs in the SOFA format, which is a widespread standard to exchange spatial acoustic data. In this paper, we also conducted an analysis of the SRIRs by computing some acoustical parameters, which include the reverberation time T_{20}, the early decay time EDT, the center time T_S, and the clarity index C_{80}.



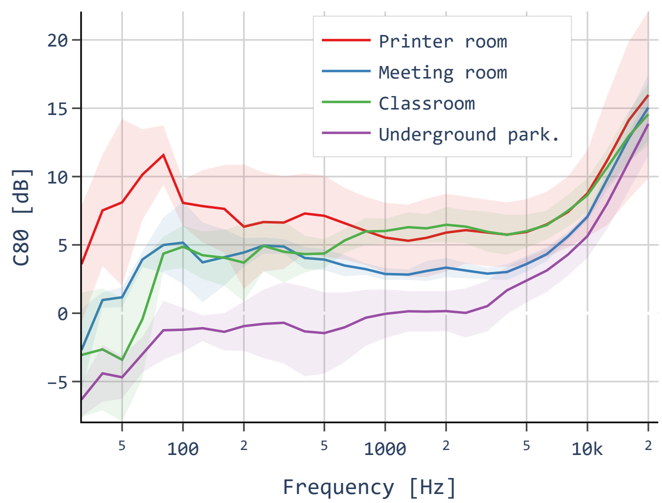
(a) Reverberation time T_{20}



(b) Early decay time EDT



(c) Center time T_S



(d) Clarity index C_{80}

Figure 6: Values of the acoustical parameters for each environment computed per one-third octave band (logarithmic scale). The solid lines represent the averaged values for each position inside the environments, while the area is the standard deviation.

The characteristics of PAN-AR make it well-suited for XR applications. In this context, the SRIRs included in PAN-AR permit the simulation of virtual sound sources in the space surrounding the users, with the spatial audio rendering adapted to their movements. Furthermore, the dataset can be employed to generate ecologically valid VAEs. This is made possible by the captured environments, which represent common scenarios in the everyday lives of workers and students. In addition, the ambient noise recordings can be employed to provide more immersive VAEs, and the spherical pictures may enhance the overall experience, adding visual stimuli consistent with the listener's position in space.

ACKNOWLEDGMENTS

This work is part of SONICOM, a project that has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101017743 (https://doi.org/10.3030/101017743) .

REFERENCES

[1] Jens Ahrens, Matthias Geier, Alexander Raake, and Claudia Schlegel. 2010. Listening and conversational quality of spatial audio conferencing. In *Audio Engineering Society Conference: 40th International Conference: Spatial Audio: Sense the Sound of Space*. Audio Engineering Society.

[2] Durand R Begault. 1994. *3-D sound for virtual reality and multimedia*. Academic Press.

[3] Jake Bhattacharyya, Lorenzo Picinali, Alessandro Vinciarelli, and Stephen Brewster. 2024. Investigating the Influence of Environmental Acoustics and Playback Device for Audio Augmented Reality Applications. In *Audio Engineering Society*

Table 2: Acoustical parameters computed with AURORA [11, 13] according to the standard ISO 3382 [14]. The values were computed for the omnidirectional channel of the SRIRs in each position of the four environments. The mean and standard deviation for each environment is also reported.

		T_{20} [ms]	EDT [ms]	T_S [ms]	C_{80} [dB]
Printer room	B	661	583	29.10	9.45
	C	608	429	11.81	14.13
	D	815	598	38.37	8.37
	E	1101	1103	79.96	1.98
	X	703	652	31.29	8.67
	Mean	777.6	673.0	38.1	8.5
	Std	175.5	227.4	22.7	3.9
Meeting room	B	1107	967	56.56	4.82
	C	1106	996	48.03	5.80
	D	1115	1024	49.29	5.53
	E	1113	971	45.47	5.96
	F	1126	1010	44.59	6.20
	X	1111	968	46.59	5.64
	Mean	1113.0	989.3	48.4	5.7
	Std	6.6	22.2	4.0	0.4
Classroom	B	1026	761	32.96	8.06
	C	1000	789	33.63	7.97
	D	1008	826	47.68	6.20
	E	1045	856	40.94	6.69
	F	1014	845	26.90	8.91
	X	954	703	30.29	8.67
	Mean	1007.8	796.7	35.4	7.8
	Std	28.0	52.9	6.9	1.0
Underground parking	B	2115	1769	87.35	2.92
	C	1838	1200	68.32	4.52
	D	2229	1834	123.08	0.25
	E	2210	1820	96.47	1.85
	F	2165	1824	117.32	0.82
	X	1985	1308	69.46	4.49
	Mean	2090.3	1625.8	93.7	2.5
	Std	138.1	265.6	1.2	1.7

Conference: AES 2024 International Audio for Games Conference. Audio Engineering Society.

[4] Jens Blauert. 1996. *Spatial hearing: the psychophysics of human sound source localization.* The MIT Press. https://doi.org/10.7551/mitpress/6391.001.0001

[5] Riccardo Bona, Davide Fantini, Giorgio Presti, Marco Tiraboschi, Juan Isaac Engel Alonso-Martinez, and Federico Avanzini. 2022. Automatic parameters tuning of late reverberation algorithms for audio augmented reality. In *Proceedings of the 17th International Audio Mostly Conference* (St. Pölten, Austria) *(AM '22)*. Association for Computing Machinery, New York, NY, USA, 36–43. https://doi.org/10.1145/3561212.3561236

[6] Jay Coggin and Will Pirkle. 2016. Automatic design of feedback delay network reverb parameters for impulse response matching. In *Audio Engineering Society Convention 141*. Audio Engineering Society.

[7] Jens Cubick and Torsten Dau. 2016. Validation of a virtual sound environment system for testing hearing aids. *Acta Acustica united with Acustica* 102, 3 (2016), 547–557. https://doi.org/10.3813/AAA.918972

[8] Filippo Denti, Davide Fantini, Federico Avanzini, and Giorgio Presti. 2024. *PANAR: A Multimodal Dataset of Higher-Order Ambisonics Room Impulse Responses, Ambient Noise and Spherical Pictures)*. https://doi.org/10.5281/zenodo.13134269

[9] James Eaton, Nikolay D Gaubitch, Alastair H Moore, and Patrick A Naylor. 2016. Estimation of room acoustic parameters: The ACE challenge. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 24, 10 (2016), 1681–1693. https://doi.org/10.1109/TASLP.2016.2577502

[10] Davide Fantini, Giorgio Presti, Michele Geronazzo, Riccardo Bona, Alessandro Giuseppe Privitera, and Federico Avanzini. 2023. Co-immersion in Audio Augmented Virtuality: the Case Study of a Static and Approximated Late Reverberation Algorithm. *IEEE Transactions on Visualization and Computer Graphics* 29, 11 (2023), 4472–4482. https://doi.org/10.1109/TVCG.2023.3320213

[11] Angelo Farina. 1995. Auralization software for the evaluation of a pyramid tracing code: results of subjective listening tests. In *ICA95 (International Conference on Acoustics), Trondheim (Norway)*. 26–30.

[12] Angelo Farina. 2000. Simultaneous measurement of impulse response and distortion with a swept-sine technique. In *Audio engineering society convention 108*. Audio Engineering Society.

[13] Angelo Farina. 2010. AURORA software. http://pcfarina.eng.unipr.it/aurora_xp/index.htm. version 4.3.

[14] International Organization for Standardization. 2009. *ISO 3382-1: International Standard ISO/DIS 3382-1: Acoustics – Measurement of room acoustic parameters – Part 1: Performance spaces.* International Organization for Standardization.

[15] Sebastià V Amengual Garí, W Owen Brimijoin, Henrik G Hassager, and Philip W Robinson. 2019. Flexible binaural resynthesis of room impulse responses for augmented reality research. In *EAA Spatial Audio Signal Processing Symposium*. Paris, France, 161–166. https://doi.org/10.25836/sasp.2019.31

[16] Michele Geronazzo and Stefania Serafin. 2023. *Sonic interactions in virtual environments* (1 ed.). Springer Nature, Cham. https://doi.org/10.1007/978-3-031-04021-4

[17] Georg Götz, Sebastian J Schlecht, and Ville Pulkki. 2021. A dataset of higher-order Ambisonic room impulse responses and 3D models measured in a room with varying furniture. In *2021 Immersive and 3D Audio: from Architecture to Automotive (I3DA)*. IEEE, 1–8. https://doi.org/10.1109/I3DA48870.2021.9610933

[18] Giso Grimm, Maartje Hendrikse, and Volker Hohmann. 2022. *Pub environment.* https://doi.org/10.5281/zenodo.5886987

[19] Rikiya Hara and Takafumi Shimizu. 2021. The effect of room sound absorption on a teleconference system and the differences in subjective assessments between elderly and young people. *Applied Acoustics* 179 (2021), 108050. https://doi.org/10.1016/j.apacoust.2021.108050

[20] Sebastian Heise, Michael Hlatky, and Jörn Loviscach. 2009. Automatic adjustment of off-the-shelf reverberation effects. In *Audio Engineering Society Convention 126*. Audio Engineering Society.

[21] Jan Ole Jungmann, Radoslaw Mazur, and Alfred Mertins. 2015. Joint time-and frequency-domain reshaping of room impulse responses. In *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 733–737. https://doi.org/10.1109/ICASSP.2015.7178066

[22] James M Kates, Kathryn H Arehart, and Lewis O Harvey. 2019. Integrating a remote microphone with hearing-aid processing. *The Journal of the Acoustical Society of America* 145, 6 (2019), 3551–3566. https://doi.org/10.1121/1.5111339

[23] Mendel Kleiner, Bengt-Inge Dalenbäck, and Peter Svensson. 1993. Auralization – An overview. *Journal of the Audio Engineering Society* 41, 11 (1993), 861–875.

[24] Hyunkook Lee and Dale Johnson. 2019. An open-access database of 3D microphone array recordings. In *Audio Engineering Society Convention 147*. Audio Engineering Society.

[25] Alexander Lindau and Stefan Weinzierl. 2012. Assessing the plausibility of virtual acoustic environments. *Acta Acustica united with Acustica* 98, 5 (2012), 804–810. https://doi.org/10.3813/AAA.918562

[26] Piotr Majdak, Franz Zotter, Fabian Brinkmann, Julien De Muynke, Michael Mihocic, and Markus Noisternig. 2022. Spatially oriented format for acoustics 2.1: Introduction and recent advances. *Journal of the Audio Engineering Society* 70, 7/8 (2022), 565–584. https://doi.org/10.17743/jaes.2022.0026

[27] Thomas McKenzie, Leo McCormack, and Christoph Hold. 2021. Dataset of Spatial Room Impulse Responses in a Variable Acoustics Room for Six Degrees-of-Freedom Rendering and Analysis. arXiv:2111.11882 [eess.AS]

[28] Paul Milgram and Fumio Kishino. 1994. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems* 77, 12 (1994), 1321–1329.

[29] Pauli Minnaar, Sylvain Favrot, and Jörg M Buchholz. 2010. Improving hearing aids through listening tests in a virtual sound environment. *The Hearing Journal* 63, 10 (2010), 40–42. https://journals.lww.com/thehearingjournal/fulltext/2010/10000/improving_hearing_aids_through_listening_tests_in.7.aspx

[30] Federico Miotello, Paolo Ostan, Mirco Pezzoli, Luca Comanducci, Alberto Bernardini, Fabio Antonacci, and Augusto Sarti. 2024. HOMULA-RIR: A Room Impulse Response Dataset for Teleconferencing and Spatial Audio Applications Acquired Through Higher-Order Microphones and Uniform Linear Microphone Arrays. arXiv:2402.13896 [eess.AS]

[31] Annika Neidhardt, Christian Schneiderwind, and Florian Klein. 2022. Perceptual matching of room acoustics for auditory augmented reality in small rooms-literature review and theoretical framework. *Trends in Hearing* 26 (2022), 23312165221092919. https://doi.org/10.1177/23312165221092919

[32] Orhun Olgun and Hüseyin Hacıhabiboğlu. 2019. *METU SPARG Eigenmike em32 Acoustic Impulse Response Dataset v0.1.0.* https://doi.org/10.5281/zenodo.2635758

[33] Julia Schütze, Christoph Kirsch, Kirsten C. Wagener, Birger Kollmeier, and Stephan D. Ewert. 2021. *Living room environment.* https://doi.org/10.5281/

zenodo.5747753 Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummer 352015383 – SFB 1330, Project C4 and C5..

[34] Justin Shen and Ramani Duraiswami. 2020. Data-driven feedback delay network construction for real-time virtual room acoustics. In *Proceedings of the 15th International Audio Mostly Conference* (Graz, Austria) *(AM '20)*. Association for Computing Machinery, New York, NY, USA, 46–52. https://doi.org/10.1145/3411109.3411145

[35] G Christopher Stecker, Travis M Moore, Monica Folkerts, Dmitry Zotkin, and Ramani Duraiswami. 2018. Toward objective measures of auditory co-immersion in virtual and augmented reality. In *Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality*. Audio Engineering Society.

[36] Chris Travis. 2009. A new mixed-order scheme for ambisonic signals. In *Proc. of the Ambisonics Symposium, Graz, Austria*.

[37] Vesa Valimaki, Julian D. Parker, Lauri Savioja, Julius O. Smith, and Jonathan S. Abel. 2012. Fifty years of artificial reverberation. *IEEE Transactions on Audio, Speech, and Language Processing* 20, 5 (2012), 1421–1448. https://doi.org/10.1109/TASL.2012.2189567

[38] Adam Weisser, Jörg M Buchholz, Chris Oreinos, Javier Badajoz-Davila, James Galloway, Timothy Beechey, and Gitte Keidser. 2019. The ambisonic recordings of typical environments (ARTE) database. *Acta Acustica United With Acustica* 105, 4 (2019), 695–713. https://doi.org/10.3813/AAA.919349

[39] Stefan A Wirler, Nils Meyer-Kahlen, and Sebastian J Schlecht. 2020. Towards transfer-plausibility for evaluating mixed reality audio in complex scenes. In *Audio Engineering Society Conference: 2020 AES International Conference on Audio for Virtual and Augmented Reality*. Audio Engineering Society.

[40] Tao Zhang, Fred Mustiere, and Christophe Micheyl. 2016. Intelligent hearing aids: The next revolution. In *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 72–76. https://doi.org/10.1109/EMBC.2016.7590643

[41] Hassan EL-Banna Zidan, U Peter Svensson, and Johan L Nielsen. 2015. Room acoustical parameters of two electronically connected rooms. *The Journal of the Acoustical Society of America* 138, 4 (2015), 2235–2245. https://doi.org/10.1121/1.4930946

[42] Franz Zotter and Matthias Frank. 2019. *Ambisonics – A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Springer Cham. https://doi.org/10.1007/978-3-030-17207-7