Tampere University Rotated Circular Array Dataset

1st Arjun Venkat Venkatakrishnan
Tampere University, Finland
arjunvenkat.venkatakrishnan@tuni.fi

2nd Pasi Pertilä
Tampere University, Finland
pasi.pertila@tuni.fi

3rd Mikko Parviainen
Tampere University, Finland
mikko.parviainen@tuni.fi

***Abstract*— Advancements in deep learning have resulted in new techniques to address sophisticated audio processing tasks, such as sound localization and recognition. However, supervised training of deep neural networks (DNNs) requires a significant amount of training data. Existing datasets are either recorded or allow synthetic recordings through impulse responses (IRs) via convolution. Recorded datasets often lack sufficient and versatile material for supervised DNN training. On the other hand, impulse response databases allow large scale dataset creation, provided that suitable IRs are available. However, existing IR datasets do not cater to the data requirements of moving and crossing sources problem in sound localization, due to insufficient angular resolution. This work introduces a versatile room IR dataset to address this problem. Various diverse environments such as office rooms, meeting rooms, corridor, and an anechoic chamber are chosen for the data collection. The chosen rooms have varying characteristics, such as reverberation times (T60) and volumes. The data is collected by placing the speaker at three different distances from a rotated microphone array, thus mimicking the moving source condition. Direction of arrival (DoA) estimation is performed by spatializing the sound signal with the collected IRs to verify their quality. The dataset will be publicly available.**

***Index Terms*—sound localization, deep learning, direction of arrival, impulse responses**

I. Introduction

The ability to solve complex audio tasks such as sound localization and recognition using DNNs [1]–[4] paved the way for the creation of several datasets. Each dataset varies in size and the objective which it aims to achieve. New algorithms and techniques, such as transfer learning and semi-supervised learning, reduce data requirements for training DNNs. However, existing datasets are not fully adequate for localizing moving and crossing sound sources. This work aims to tackle these needs by enabling large scale and versatile creation of labeled spatial audio data for supervised training of DNNs.

Analysis of various IR datasets [5]–[12] indicates a number of data needs left unaddressed by them. Firstly, diverse acoustic conditions enable the performance of various experiments, yet most of the collected data is limited to an average of three or four different spaces – often not including anechoic environments, for instance. Secondly, crossing and moving source problems require non-static receivers or sources, but existing IR datasets exhibit little to no variation in microphone and speaker positions. Thirdly, supervised training of DNNs requires large and diverse datasets, which may be obtained by collecting data with fine angular resolution (and thus more spatial configurations).

Funding by Academy of Finland project no. 321881

To address these issues, the Tampere University Rotated Circular Array impulse response (TUNI-RCAIR) dataset offers IRs with fine resolution angular data collected from different acoustic conditions. IRs are collected from spaces which vary in dimensions and reverberation times (T60). The recording environments are three office rooms, two meeting rooms, a corridor and an anechoic chamber. Varying reverberation conditions provide a parallel corpus of IRs, i.e. the same spatial configurations are available in varying rooms and distances. A static loudspeaker is used as the source and a rotating microphone array on a tripod acts as the receiver, which mimics the impression of a moving source by capturing the direct path wavefront from multiple orientations. Three different IR capture distances between the microphone array and the loudspeaker are involved during the data collection. The recordings are collected over a range of 180° for every 5°. Fine resolution in sound source angles is thus achieved by rotating the microphone array, which provides IRs aimed at simulating moving sources. The dataset contains 777 IRs for the array (37 different angles, 3 distances and 7 rooms; 5439 IRs in total, if we consider the seven microphones individually).

With the work presented in this paper, one can generate the data required for moving sources and crossing sources scenarios to research these difficult problems with the aid of supervised DNNs. This IR database will be useful in solving challenging sound localization problems, such as moving sources and multiple moving sources, which are still identified as current challenges [10].

The paper is divided into the following sections: Section II presents related works and makes comparisons with the TUNI-RCAIR dataset. Section III describes the data collection setup and environments in detail. Section IV details the evaluation of the IR dataset in terms of DoA estimation error through spatialization with audio signals. The final section presents brief conclusions and future work.

II. Related Works

The impulse response dataset finds its application in tasks such as speech recognition, speech enhancement, and sound localization. There are multiple impulse response datasets, in which the impulses are collected from different recording environments. Some of the existing IR datasets and the tasks they aim to address are described below. A brief overview is

provided about the utilized recording equipment, microphone and speaker placement, as well as the intended purpose of the dataset.

- **RWCP Sound Scene Database** [5] – Data collected using 14-channel linear and 54-channel spherical microphone arrays and Diatone DS-7 loudspeaker, with B&K type 4128 head torso used as the source. This dataset was designed to simulate sounds in different acoustic environments. For the IRs, consistent angular data was collected from five different rooms with angles between [50°,. . . ,130°] in 20° steps via a linear microphone array. The initial distance between the source and the speaker was two meters.
- **Reverb challenge dataset** [6] – Consists of both real and simulated data collected from three different rooms using omnidirectional microphones and loudspeaker. Targeted for speech enhancement and automatic speech recognition tasks. Eight-channel circular array was used for data collection and the loudspeaker position was varied twice across the rooms. There are six unique microphone/speaker positions. No angular resolution is involved in this dataset.
- **DCASE dataset** [7] – As a part of sound event detection and localization challenge in the DCASE series. Consists of real impulse responses collected from five large indoor spaces. A spherical microphone array was used for recording the IRs. The distance between the source and speaker was set to one and two meters respectively. The highest resolution of speaker angles is 10°, between [-40°,. . . ,40°] degrees.
- **Center for Digital Music (C4DM)-RIR dataset** [8] – Includes omnidirectional impulse responses collected from three different rooms. Data collected using omnidirectional DPA 4006 microphones and a B-format microphone. The microphones were moved to different positions across the rooms used for data collection, with 468 positions in total. Various angles were also employed during data collection.
- **Acoustic Characterisation of Environment corpus (ACE)** [9] – Data collected from seven different rooms using two-channel laptop, notebook, three-channel mobile phone, 32-channel Eigen microphone and eight-channel linear array microphones. Targeted at speech recognition and enhancement tasks. The microphone positions were set at two different distances in each room. This dataset was designed to identify room reverberation characteristics.
- **BUT Speech@FIT Reverb Database** [11] – IRs were collected from nine different rooms, mainly for speech recognition and speech enhancement tasks. Different microphone setups were involved in data collection: mounted on walls, placed over a table, placed on the ceiling, etc., since the focus was towards speech recognition. Two types of microphones were used for data collection: omnidirectional microphone and electret condenser microphone module. The loudspeaker position was varied in the range of three to ten positions across the rooms used for data collection.
- **Multichannel Impulse Response Database** [12] – Impulse responses were collected from a single room at the BIU acoustics lab. However, the room has configurable reverberation levels. Three different levels were chosen for data collection. The receiver and source were, respectively, eight AKG CK 32 with three different array combinations and Fostex 6301BX loudspeaker arrays. The position of the sources varied with 15° steps in the range [-90°,. . . ,90°]. This database mainly targets source separation tasks.

In addition to these datasets there are binaural IR datasets, which involve the collection of Head Related Impulse Responses (HRIR) and Head Related Transfer Function (HRTF), their frequency domain counterpart.

In the Aachen Impulse Response Database (AIR) [13] dataset, binaural IRs were collected using HMS2 artificial head by head acoustics in four different rooms. In the CIPIC [14] dataset, HRTF was collected from 45 different subjects from a single room. ER-7C (Etymotic Research) probe microphone and an array of Bose loudspeakers, placed at different positions, were used for data collection. Azimuth was varied between [-45°,...,45°], in 5° intervals, by moving the loudspeakers. The RIEC [15] dataset consists of HRTFs collected from 105 different subjects in an anechoic chamber. Angular resolution in terms of azimuth was set to 5° for the range of [-180°,...,180°]. FG3329, Knowles microphones and Fostex loudspeaker array were used for collecting HRTFs. In the LISTEN [16] dataset, HRIRs were collected in an anechoic chamber. TANNOY system 600 was used as the source and B&K omnidirectional microphone type 4149 was used to collect the IRs. The distance between the source and the receiver was 0.5 m.

Table I compares the datasets discussed above to the TUNI-RCAIR dataset. The datasets are compared on the basis of number of spaces used, distance range between the source and speaker, and the angular range in which the microphones or speakers were varied. The unique variations in the microphone and speaker positions are also shown, counting the microphone arrays as a single entity rather than as individual microphones. Unique variations refer to the change in microphone position or the speaker position, while collecting the impulse responses. The table also highlights whether angular resolution is used for data collection. The comparison clearly highlights the need for an IR dataset with fine angular resolution, collected from different acoustic environments, which the TUNI-RCAIR dataset addresses.

III. Data collection and Space characteristics

This section describes the setup of the data collection and the spaces used for it.

TABLE I: Comparison of impulse response datasets. Distance range represents the range of distances between the speaker/source and microphone array/receiver. Microphone/Speaker positions indicates unique changes in the positions of microphones and/or the loudspeaker across all the recording environments. The microphone array is regarded as a single entity. Angle range refers to azimuth angles. NA denotes Not Applicable

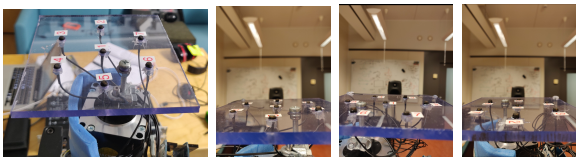
Dataset	**Rooms**	**Distance range** (m)	**Microphone/Speaker positions**	**Angular resolution** (deg°)	**Angle range** (deg°)
AIR	4	0.5 – 10.2	18	NA	NA
RWCP	5	2	25	20°	50° – 130°
Reverb	3	0.5 – 2	6	NA	NA
DCASE	5	1 – 2	504	10°	-40° – 40°
C4DM-RIR	3	2 – 12	468	NA	NA
ACE	7	0.5 – 2	14	NA	NA
BUT Speech@FIT Reverb	9	0.5 – 15	41	NA	NA
Multichannel IR database	3	1 – 2	234	15°	-90° – 90°
TUNI-RCAIR	7	0.7 – 2.1	777	5°	-90° – 90°

A. Data collection setup

The microphone array consists of seven microphones, with the first one in the middle and the other six microphones placed on a circle of diameter 8 cm. The array setup is similar to UMA-8 [17] and is closer to the microphone arrays used in latest smart speakers such as Amazon Echo [18]. This setup is connected to a step motor, controlled by the Arduino stepper library. The step motor is mounted on Raspberry Pi and connected to a laptop. The Raspberry Pi is also connected to an external power source running the step motor. The step motor and the microphone array are placed over a tripod. The microphones are connected to OctaMic – RME audio interface for pre-amplification. Fireface UFX (RME) is used as an ADC connected to the OctaMic. DPA 4060 microphones act as the receiver and a Genelec 8010A loudspeaker is the source in this setup. Figure 1(a) represents the microphone array mounted on a step motor. The individual microphones are numbered for identification.

The source speaker and the receiver microphone array are kept at three different distances: 0.7 m, 1.4 m, and 2.1 m. The Arduino stepper library is used to control and change the step motor resolution. Maximum Length Sequence (MLS) [19] signal with 48 kHz sampling rate is used as the excitation signal to generate the IRs. The distance between the receiver and the source speaker is calculated from the central microphone. The microphone array is rotated from -90° to 90° at a fine resolution of 5° at a time and the recordings are captured for each angle. When microphones 1 and 2, with coordinates [0,0,0] and [40,0,0] respectively, are facing the loudspeaker, the angle between the source and the receiver is 0°. Refer to Table II for individual microphone coordinates.

Figures 1(b), 1(c), and 1(d) represent the microphone array and loudspeaker positions for the angles -90°, 0°, and 90°.



(a) Microphone array (b) -90° (c) 0° (d) 90°

Fig. 1: Microphone array with each microphone numbered. Angle specified between microphone 1 and loudspeaker.

B. Spaces used for data collection

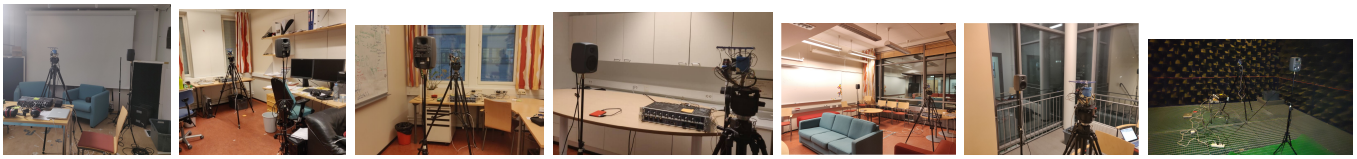
Seven different spaces used for data collection are described below and displayed in Figures 2(a) – 2(g).

1) **Audio Lab (ALB)** – A studio-like setup with some furniture and lights. It falls under the category of low reverberation but not as low as an anechoic chamber.
2) **Office Room-1 (ORA)** – A typical work environment in the university campus. This work space contains tables, sofa, and chairs.
3) **Office Room-2 (ORB)** – Another type of office setup, with different dimensions and acoustic characteristics than Office Room-1.
4) **Small Meeting Room (SMR)** – One of the smallest places used for data collection. It is highly reverberant in nature. It contains a table and cupboards.
5) **Large Meeting Room (LMR)** – The Large Meeting Room is bigger than other rooms from which data is collected. Contains a table, a few chairs and sofas.
6) **Glass Corridor (GCO)** – IRs are collected from one of the staircases of the university premises. It is highly reverberant, the most among the spaces involved in data collection.
7) **Anechoic Chamber (AEC)** – IRs are also collected from an anechoic chamber. It is the least reverberant of all the spaces involved in data collection.

Table III lists physical dimensions, volumes and reverberation times of the rooms. The volume of each room is the product of its measured length, width, and height. This is the empty volume, although all the rooms except for the anechoic chamber and the glass corridor contain various furniture. The reverberation time is calculated using Room EQ wizard [20] software. Logarithmic sine sweep is used as the excitation signal [21]. Genelec 8010A loudspeaker is used as the source and Earthworks M30 microphone as the receiver.

TABLE II: Microphone coordinates along x, y, and z axes.

Microphone	**x-axis** (mm)	**y-axis** (mm)	**z-axis** (mm)
1	0	0	0
2	40	0	0
3	20	30	0
4	-20	30	0
5	-40	0	0
6	-20	-30	0
7	20	-30	0



(a) Audio Lab (b) Office Room-1 (c) Office Room-2 (d) Small Meeting Room (e) Large Meeting Room (f) Glass Corridor (g) Anechoic Chamber

Fig. 2: Data collection spaces.

Space	**ID**	**Dimensions** (m)	**Volume** (m^3)	**Reverberation Time T60** (ms)
Audio Lab	ALB	$(2 \times 2.3 \times 2.6)$	11.96	270
Office Room-1	ORA	$(3.9 \times 4.9 \times 3.2)$	61.15	486
Office Room-2	ORB	$(2.7 \times 5.1 \times 3.2)$	44.06	314
Small Meeting Room	SMR	$(2.5 \times 3.7 \times 2.5)$	23.13	416
Large Meeting Room	LMR	$(6.6 \times 5.2 \times 3.2)$	109.82	319
Glass Corridor	GCO	$(7.3 \times 2.7 \times 3.4)$	67.01	1700
Anechoic Chamber	AEC	$(7.8 \times 7.8 \times 7.8)$	474.56	60

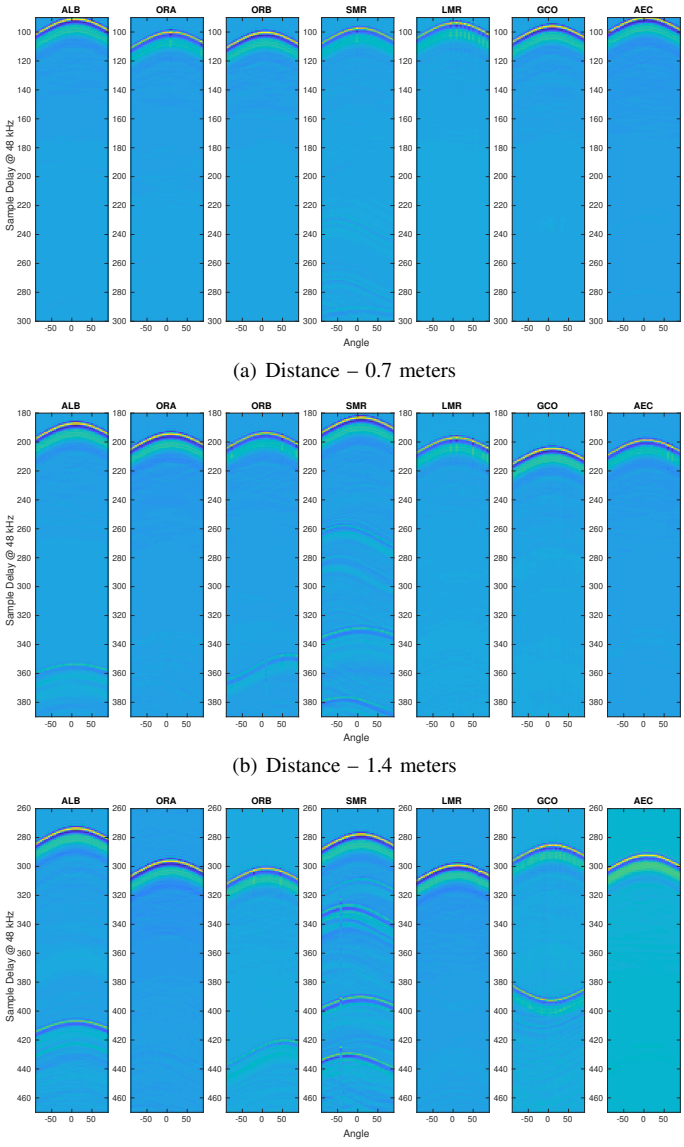
TABLE III: Spaces used for data collection and their characteristics. Dimensions are expressed as (width×length×height).

IV. Evaluation of the Dataset

The quality of the recordings was checked by generating plotting both impulse and frequency responses. In additi the plots, sanity checks were performed on the generated Using the collected IRs DoA estimation was performed t derstand the collected data and identify any faulty recor (e.g. due to external disturbances).

The estimation was performed through spatializing sp or music signals by convolving them with the collected This process was done for each room for all the diff distances involved in data collection. Only the direct pat from the impulse response was used for spatialization. W Gaussian noise was spatialized with the collected IRs acro rooms and distances. After the spatialization, the appro segment of the convolved data was chosen. Time Differ Of Arrival (TDOA) was estimated for each microphone combination using Generalized Cross-Correlation with P Transform (GCC-PHAT) [22]. The TDOA value is obt as the delay that maximizes the GCC-PHAT function. B on the array geometry of Table II and the estimated delays, the DoA was estimated using Steered-Response P algorithm with Phase Transform (SRP-PHAT) [23]. Befor recordings were collected, the alignment between the m phone array and the speaker had been performed manu Once the DoA was obtained for a particular recording ronment, the alignment bias was identified by comparin actual and estimated DoA and finding a value that minim their offset.

Table IV represents the bias and error between actual estimated DoA across all distances for the rooms. Fig 3(a), 3(b), and 3(c) depict the direct path IR values a all distances and rooms for microphone-2. Refer to Fig. for the microphone array geometry and numbering. Fron plots one could notice a curve denoting the rotating array. lower parts of the images show sound reflections obser as additional curvatures. When the loudspeaker is close to the wall, the additional reflections can be seen in Figures 3(b) and 3(c), especially in the Small Meeting Room, owing to its size and reverberations.



(a) Distance – 0.7 meters

(b) Distance – 1.4 meters

(c) Distance – 2.1 meters

Fig. 3: IR values for direct path and first reflections for microphone 2 at different distances for array angles $[-90°, \ldots, 90°]$. The order of the rooms is Audio Lab, Office Room-1, Office Room-2, Small Meeting Room, Large Meeting Room, Glass Corridor, and Anechoic Chamber.

V. Conclusion

The main purpose of this work was to create a robust impulse response dataset. Although there are multiple existing IR datasets, our work addresses the lack of datasets for sound localization tasks, including moving sources and crossing sources. The spaces used for data collection vary in dimensions and acoustic properties such as reverberation time. Given the large training material requirements of supervised deep learning, there is a need to generate large amounts of data. The TUNI-RCAIR dataset allows simulating moving source data in versatile rooms and different distances. Future work involves collecting data from new spaces and increasing the size of the dataset.

References

[1] Soumitro Chakrabarty and Emanuël AP Habets, "Multi-speaker doa estimation using deep convolutional networks trained with noise signals," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 1, pp. 8–21, 2019.

[2] Nelson Yalta, Kazuhiro Nakadai, and Tetsuya Ogata, "Sound source localization using deep learning models," *Journal of Robotics and Mechatronics*, vol. 29, no. 1, pp. 37–48, 2017.

[3] Weipeng He, Petr Motlicek, and Jean-Marc Odobez, "Deep neural networks for multiple speaker detection and localization," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 74–79.

[4] Sharath Adavanne, Archontis Politis, and Tuomas Virtanen, "Localization, detection and tracking of multiple moving sound sources with a convolutional recurrent neural network," *arXiv preprint arXiv:1904.12769*, 2019.

[5] Satoshi Nakamura, Kazuo Hiyane, Futoshi Asano, Takanobu Nishiura, and Takeshi Yamada, "Acoustical sound database in real environments for sound scene understanding and hands-free speech recognition," in *LREC*, 2000.

[6] Keisuke Kinoshita, Marc Delcroix, Sharon Gannot, Emanuël AP Habets, Reinhold Haeb-Umbach, Walter Kellermann, Volker Leutnant, Roland Maas, Tomohiro Nakatani, Bhiksha Raj, et al., "A summary of the reverb challenge: state-of-the-art and remaining challenges in reverberant speech processing research," *EURASIP Journal on Advances in Signal Processing*, vol. 2016, no. 1, pp. 1–19, 2016.

TABLE IV: Error measure between actual source angles and estimated DoA

Room ID	**Distance** (m)	**Bias** (deg°)	**Error** (deg°)
ALB	0.7	-0.28	0.55
	1.4	-0.02	0.59
	2.1	0.33	0.66
ORA	0.7	1.07	0.55
	1.4	-0.05	0.59
	2.1	-0.62	0.71
ORB	0.7	-2.13	0.82
	1.4	-1.77	0.69
	2.1	-0.80	0.67
SMR	0.7	1.07	0.55
	1.4	-0.05	0.59
	2.1	-0.62	0.71
LMR	0.7	-2.40	0.84
	1.4	-1.67	0.79
	2.1	0.70	0.68
GCO	0.7	0.03	0.61
	1.4	-0.28	0.63
	2.1	0.86	0.84
AEC	0.7	-4.61	0.68
	1.4	-1.73	1.0
	2.1	-1.73	1.04

[7] Sharath Adavanne, Archontis Politis, and Tuomas Virtanen, "A multi-room reverberant dataset for sound event localization and detection," in *Detection and Classification of Acoustic Scenes and Events 2019 Workshop (DCASE2019)*, 2019.

[8] Rebecca Stewart and Mark Sandler, "Database of omnidirectional and b-format room impulse responses," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2010, pp. 165–168.

[9] James Eaton, Nikolay D Gaubitch, Alastair H Moore, and Patrick A Naylor, "The ace challenge—corpus description and performance evaluation," in *2015 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*. IEEE, 2015, pp. 1–5.

[10] Christine Evers, Heinrich Löllmann, Heinrich Mellmann, Alexander Schmidt, Hendrik Barfuss, Patrick Naylor, and Walter Kellermann, "The locata challenge: Acoustic source localization and tracking," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1620–1643, 2020.

[11] Igor Szoke, Miroslav Skacel, Ladislav Mosner, Jakub Paliesek, and Jan Cernocky, "Building and evaluation of a real room impulse response dataset," *IEEE Journal of Selected Topics in Signal Processing*, vol. 13, no. 4, pp. 863–876, 2019.

[12] Elior Hadad, Florian Heese, Peter Vary, and Sharon Gannot, "Multi-channel audio database in various acoustic environments," in *2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC)*. IEEE, 2014, pp. 313–317.

[13] Marco Jeub, Magnus Schäfer, and Peter Vary, "A binaural room impulse response database for the evaluation of dereverberation algorithms," in *Proceedings of International Conference on Digital Signal Processing (DSP)*. IEEE, 2009, pp. 1–4.

[14] V Ralph Algazi, Richard O Duda, Dennis M Thompson, and Carlos Avendano, "The cipic hrtf database," in *Proceedings of the 2001 IEEE Workshop on the Applications of Signal Processing to Audio and Acoustics (Cat. No. 01TH8575)*. IEEE, 2001, pp. 99–102.

[15] Kanji Watanabe, Yukio Iwaya, Yôiti Suzuki, Shouichi Takane, and Sojun Sato, "Dataset of head-related transfer functions measured with a circular loudspeaker array," *Acoustical science and technology*, vol. 35, no. 3, pp. 159–165, 2014.

[16] Olivier Warusfel, "Listen hrtf database," http://recherche.ircam.fr/equipes/salles/listen/, Accessed: 26.02.2021.

[17] "UMA-8 user manual," https://www.minidsp.com/images/documents/UMA-8\20v2\%20User\%20manual.pdf, Accessed: 26.02.2021.

[18] "Amazon Echo," https://www.amazon.com/gp/help/customer/display.html?nodeId=201601770, Accessed: 26.02.2021.

[19] Douglas D Rife and John Vanderkooy, "Transfer-function measurement with maximum-length sequences," *Journal of the Audio Engineering Society*, vol. 37, no. 6, pp. 419–444, 1989.

[20] "Room Eq Wizard," https://www.roomeqwizard.com/, May, Accessed: 26.05.2021.

[21] Angelo Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Audio Engineering Society Convention 108*. Audio Engineering Society, 2000.

[22] Charles Knapp and Glifford Carter, "The generalized correlation method for estimation of time delay," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 24, no. 4, pp. 320–327, 1976.

[23] Joseph Hector DiBiase, *A high-accuracy, low-latency technique for talker localization in reverberant environments using microphone arrays*, Ph.D. thesis, Providence, RI: Brown University, 2000.