A Dataset of Meeting Room Impulse Responses for Distributed Sound Source Localization

Bilgesu Çakmak*, Mike Vantorre†, Thomas Dietzen‡, Toon van Waterschoot*
*Dept. of Electrical Engineering (ESAT), STADIUS, KU Leuven, Leuven, Belgium
†Televic Conference, Izegem, Belgium
‡Dept. of Electrical Engineering (ESAT), PSI, EAVISE, Sint-Katelijne-Waver, Belgium

***Abstract*—This report describes a dataset of room impulse response (RIR) measurements acquired in a realistic U-shaped conference room using a distributed microphone system. The measurements were performed with conference microphones and multiple loudspeaker source positions under different microphone orientations and acoustic conditions. Room impulse responses were obtained using exponential sine sweep excitation, followed by post-processing of the recorded signals to extract the acoustic responses. The dataset is intended to support research on sound source localization and signal processing in wireless acoustic sensor networks.**

I. Introduction

Distributed microphone systems are increasingly used in conference rooms, teleconferencing, and smart meeting environments. These systems are often modeled as wireless acoustic sensor networks (WASNs) composed of spatially distributed microphones. However, publicly available room impulse response (RIR) datasets suitable for WASN research are limited, especially for realistic conference room scenarios involving a large number of microphones. Existing datasets, such as the HOMULA dataset [1], are typically recorded in classroom-like configurations. Although these datasets are valuable, they do not fully reflect the geometry and usage patterns of real conference rooms.

The motivation of this work is to provide RIR measurements acquired under realistic U-shaped conference room conditions that are representative of distributed microphone systems used in practice.

II. Methodology

The overall measurement chain is illustrated in Fig. 1. The excitation signal $y(t)$ is generated on the PC and transmitted over the Dante network to the Biamp Tesira unit. The Biamp routes the signal to the loudspeaker, which emits the sweep into the room. The acoustic propagation from loudspeaker to microphones, including direct sound, reflections, and reverberation, is represented by the transfer function $\mathbf{H}(\omega)$. The microphones capture the resulting sound field, and the recorded channels are collected by the Televic Multimedia Engine (MME) [1] and sent back to the PC as the multichannel signal $\mathbf{u}(t)$.

[1]https://www.televic.com/en/conference/products/central-equipment/plixus-multimedia-engine

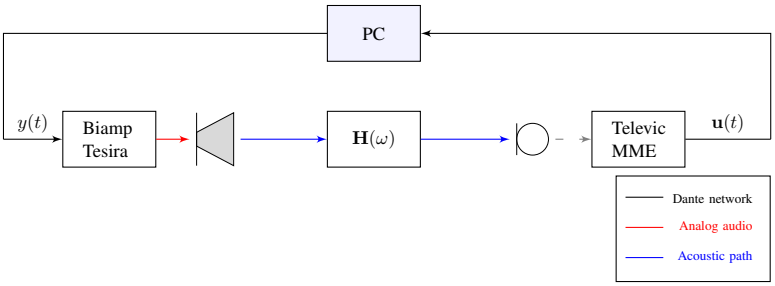

Fig. 1: Block diagram of the measurement and recording setup.

A. Excitation Playback

The excitation signal consisted of an exponential sine sweep, following the method proposed by [2], generated on the recording PC and transmitted via Dante Virtual Soundcard (DVS), configured as an ASIO audio interface. The Dante network was used to transport multichannel digital audio between the recording PC and the conference system over standard Ethernet. The sweep had a duration of 10 s, was sampled at 48 kHz, and covered the frequency range from 100 Hz to 24 kHz. To avoid truncation of the impulse response, 1 s of silence was added before and after the sweep, and the playback level was set to -12 dBFS. The digital signal was routed to a Biamp TesiraFORTÉ DAN VT DSP, which converted the digital signal into analogue audio required by the loudspeaker chain. A static gain block was configured inside the DSP to ensure a sufficiently high excitation level. Playback was performed using an HBK OmniPower Sound Source (Type 4292-L), placed sequentially at five predefined source locations.

B. Microphone Recording

Microphone signals were captured using Televic Confidea FLEX delegate units,[2] each equipped with a directional D-Mic 30 SL (30 cm) gooseneck microphone.[3] All channels were collected by the Televic MME, which exposed the multichannel audio stream on the Dante network and forwarded it to the recording PC. The signals were recorded at a sampling rate of 48 kHz with a bit depth of 32 bits across 23 channels.

[2]https://www.televic.com/en/conference/products/wired/flexible-wired-tabletop

[3]https://documents.televic.digital/conference/index.php/s/pzSymwd8iNntWBD

C. Signal Processing and RIR Extraction

The recordings were processed using scripts based on the open-source PyRIRTool framework,[4] implementing swept-sine deconvolution via Farina's method [2].

D. Coordinate Reconstruction

The spatial coordinates of the microphones and loudspeaker source were reconstructed from measured distances to four fixed anchor points with known coordinates. The anchor positions were measured manually and assumed to be known with sufficient accuracy. The distances $\hat{d}_k$ were measured using a laser distance meter between each microphone (and source) position and the four anchor points, and subsequently used as inputs to the coordinate reconstruction procedure.

Let $\mathbf{a}_k \in \mathbb{R}^3$, $k = 1, \ldots, 4$, denote the anchor coordinates, and let $\hat{d}_k$ denote the measured distance between a point of interest (microphone or source) and anchor k. Coordinate reconstruction was formulated as a nonlinear least-squares problem that minimizes the discrepancy between the measured distances and the Euclidean distances predicted by a candidate position $\mathbf{p}$:

$$\min_{\mathbf{p}} \sum_{k=1}^{4} \left(\|\mathbf{p} - \mathbf{a}_k\|_2 - \hat{d}_k \right)^2 . \tag{1}$$

To reduce ambiguity and improve robustness, a fixed height was assumed for all microphones and the loudspeaker source. Consequently, only the horizontal coordinates (x, y) were estimated, while the vertical coordinate z was fixed to a predefined value based on the measurement setup. The resulting two-dimensional optimization problem was solved independently for each microphone and source position.

The nonlinear least-squares problems were solved using a robust optimization procedure with a Huber loss function [3] to reduce the influence of inaccurate distance measurements. A top-down view of the reconstructed geometry for the case in which the source is located at the center of the room, including the four anchor positions, reconstructed microphone locations, and the reconstructed loudspeaker source position, is shown in Fig. 2.

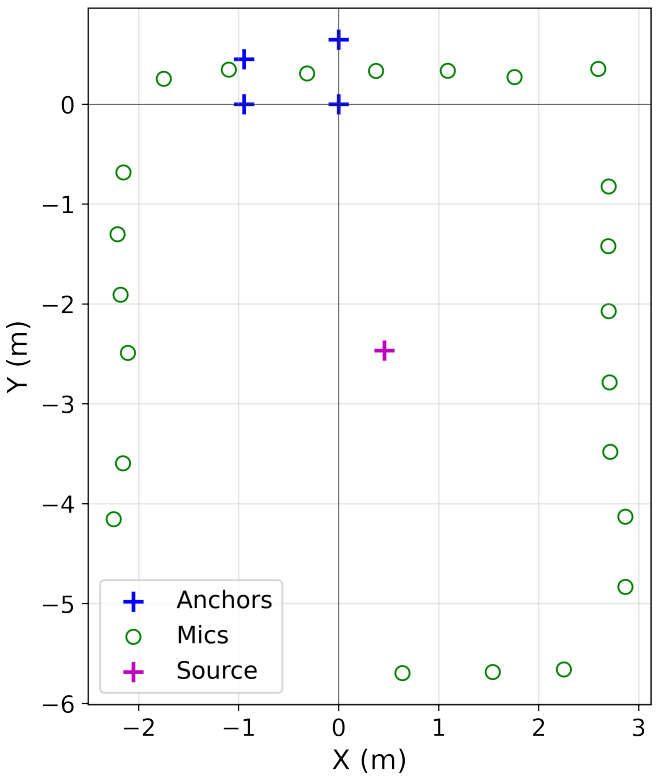

Fig. 2: Top-down view (X-Y plane) of the reconstructed geometry for the case in which the source is located at the center of the room. The anchor points, microphone positions, and the source location are shown.

III. Measurement Setup

RIRs were measured in an indoor U-shaped conference room of size $10.30\,\mathrm{m} \times 10.13\,\mathrm{m} \times 3.48\,\mathrm{m}$ using 23 microphones and a single loudspeaker source. Four fixed anchor positions are defined in the room and used as reference points for later coordinate reconstruction. The measurements were conducted for 5 different source positions. The microphones used in the measurements were gooseneck microphones as commonly used in real conference systems whose directional response depends on their orientation. Consequently, 3 microphone orientations were considered:

- upward-facing
- oriented toward the center of the room,
- oriented toward a typical seated speaker position at the desk.

The room consisted of two wooden-paneled walls and one concrete wall. The remaining wall was a glass surface, which could either be covered by a curtain or left uncovered. Measurements were therefore carried out under two room conditions: with curtain and without curtain. As a result, for each source position, a total of 6 RIRs were recorded, corresponding to the combination of three microphone orientations and two curtain configurations. The reverberation time (RT60) was estimated using the T30 method, based on linear regression of the Schroeder energy decay curve between $-5\,\mathrm{dB}$ and $-35\,\mathrm{dB}$, in accordance with ISO 3382-1 [4]. The resulting reverberation time was approximately $0.7\,\mathrm{s}$ with the curtain and $0.9\,\mathrm{s}$ without the curtain. Table I provides an overview of the parameters of the measured RIR dataset.

An example measured RIR corresponding to the configuration without curtains is shown in Fig. 3. The direct sound, early reflections, and the subsequent reverberation tail are clearly visible.

Examples of measurement configurations for the case without the curtain are shown in Fig. 4 and Fig. 5. In Fig. 4, the microphones are oriented toward the center of the room, whereas Fig. 5 illustrates the configuration in which the microphones are oriented toward a typical seated speaker position, with the loudspeaker located at the central source position.

[4] https://github.com/maj4e/pyrirtool

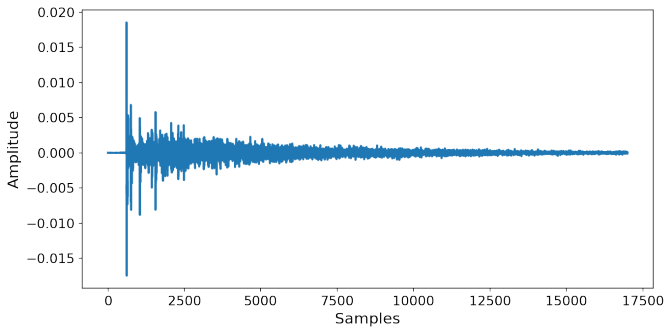

Fig. 3: Example measured room impulse response for one microphone channel.



Fig. 4: Measurement setup in the U-shaped conference room with microphones oriented toward the center of the room. The loudspeaker used for excitation playback is placed at one of the predefined source positions (here, at the center), while microphones are distributed around the tables. The room includes a glass wall on one side, which could be either covered by a curtain or left uncovered during measurements.

IV. Dataset Availability

The measured room impulse responses and reconstructed geometry are publicly available on Zenodo:

https://zenodo.org/records/19068723

The repository contains the excitation sweep signals, the recorded sweep responses, and the extracted room impulse responses for all measurement configurations. It also includes the reconstructed microphone and source coordinates, metadata describing the microphone orientations and room conditions, the coordinate reconstruction script, and top-down plots of the reconstructed geometry for each source position.



Fig. 5: Zoomed-in view of the measurement setup with microphones oriented toward a typical seated speaker position, showing the loudspeaker and microphone placement in greater detail.

TABLE I: Summary of the parameters of the measured RIR dataset.

Room size	$10.30\,\mathrm{m} \times 10.13\,\mathrm{m} \times 3.48\,\mathrm{m}$
Microphones	23
Source positions	5
Mic orientations	3
Curtain conditions	2
Configurations	$5 \times 3 \times 2 = 30$
Total RIRs	$23 \times 5 \times 3 \times 2 = 690$
Excitation	ESS, 10 s
Sampling rate	48 kHz
Sweep range	100 Hz – 24 kHz
Reverberation time	0.7 s / 0.9 s

Acknowledgment

The authors thank Televic for their time and energy spent on this project.

This research work was carried out at the ESAT Laboratory of KU Leuven, and has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 956369 'SOUNDS' and from the European Research Council under the European Union's Horizon 2020 research and innovation program / ERC Consolidator Grant: SONORA (no. 773268). This report reflects only the authors' views and the Union is not liable for any use that may be made of the contained information.

References

[1] F. Miotello, P. Ostan, M. Pezzoli, L. Comanducci, A. Bernardini, F. Antonacci, and A. Sarti, "HOMULA-RIR: A room impulse response dataset for teleconferencing and spatial audio applications acquired through higher-order microphones and uniform linear microphone arrays," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. Workshops (ICASSPW)*, 2024, pp. 795–799.

[2] A. Farina *et al.*, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," *Preprints-Audio Engineering Society*, 2000.

[3] P. J. Huber, "Robust Estimation of a Location Parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73 – 101, 1964. [Online]. Available: https://doi.org/10.1214/aoms/1177703732

[4] International Organization for Standardization, “Acoustics — measurement of room acoustic parameters — part 1: Performance spaces,” ISO Standard, 2009, iSO 3382-1:2009.