SURIR: A DATASET OF SUCCESSIVELY MEASURED ROOM IMPULSE RESPONSES FOR LOCALIZING SILENT HUMANS

Jeremy Lawrence[1], Andreas Walther[2], Emanuël A. P. Habets[1, 2], Nils Peters[3]

[1] *International Audio Laboratories Erlangen, Germany* *
[2] *Fraunhofer Institute for Integrated Circuits (IIS), Erlangen, Germany*
[3] *Trinity College, The University of Dublin, Ireland*

Abstract

Acoustic localization of humans traditionally focuses on scenarios in which they emit sound. In certain settings, however, a human may remain silent while another sound source is active. Recent work has demonstrated that silent humans can be localized by analyzing the subtle perturbations their bodies introduce into room impulse responses (RIRs). The existing SoundCam dataset includes RIRs estimated with a silent human present, but provides only single-snapshot RIRs per human position. In contrast, successively measured RIR estimates enable the analysis of subtle temporal variations caused by involuntary human micromotion, e.g., breathing, which can lead to more robust localization performance than single-snapshot human-present versus human-absent comparisons. To advance research in this field, we introduce SuRIR, the first human localization dataset that provides four successive RIR estimates for each configuration, measured a few seconds apart. SuRIR comprises 15,360 RIR estimates, acquired at 48 kHz using exponential sine sweeps. Measurements were conducted using six microphones and two loudspeakers across two rooms, with three humans present under five conditions that varied posture, movement, and orientation. To provide a static reference condition, SuRIR includes 960 RIRs measured with a mannequin present. Besides human localization, SuRIR provides a controlled testbed for evaluating RIR-based algorithms under subtle environmental variations.

Keywords: *room impulse response, acoustic localization, listener tracking, microphone array*

***Corresponding author**: jeremy.lawrence@fau.de.*

*The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg and Fraunhofer IIS.

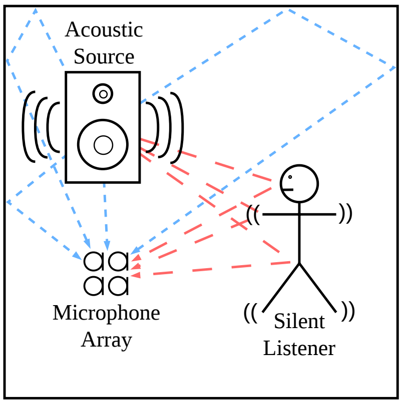

Figure 1: Schematic illustration of selected acoustic paths between an acoustic source and a microphone array. In successive RIR measurements, human micromotion alters the red acoustic paths, whereas the blue paths are affected only by environmental variations.

1. Introduction

Acoustic localization of silent humans is an emerging research area with applications such as listener-position-adaptive audio playback without requiring any external tracking devices beyond the components already integrated into modern audio systems. Several approaches have been proposed that estimate a human's position from room impulse responses (RIRs) [1], [2], [3], [4], [5]. Most of these methods compare a single human-present RIR to a human-absent baseline RIR to isolate the perturbation introduced by the human body. Although theoretically sound, these approaches can degrade in practice when additional differences arise between the two measurements, e.g. due to furniture displacement, which have been shown to substantially alter the reverberation time of a room [6]. Even in static environments, room air temperature is never constant [7], causing measurable temporal variations in estimated RIRs that become more pronounced with increasing time between measurements [8], [9],

Table 1: Comparison of SoundCam and SuRIR.

	SoundCam	**SuRIR**
Rooms	3	2
Human subjects	5	3
(Unique) human positions	5,000	10
Movement conditions	1	5
Scenarios	**5,000**	**150**
Loudspeaker positions	1	2
Microphone positions	10	6
Sweep lengths	1	2
Sweep repetitions	1	4
RIRs per scenario	**10**	**96**
Human-present RIRs	**50,000**	**14,400**
Human-absent RIRs	**1,300**	**960**

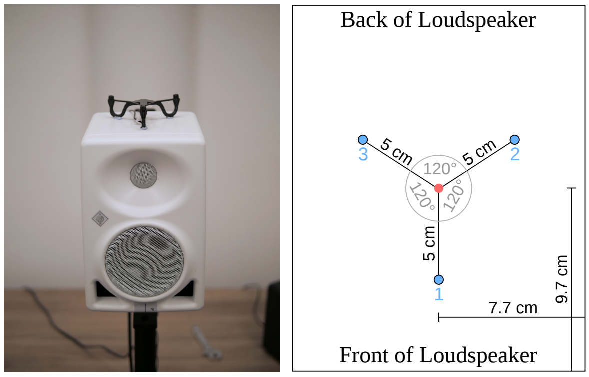

Figure 2: Photograph of one loudspeaker with the three top-mounted microphones (left) and a schematic of the resulting microphone layout (right). The blue labels denote the microphone channel order.

[10]. The approach presented in [5] mitigates these limitations by eliminating the need for the human-absent baseline RIR and instead comparing *successive* RIR estimates using differential RIRs, exploiting involuntary human movement [11]. This principle is related to approaches exploiting Wi-Fi channel state information measurements for human-motion sensing [12], but is applied here to audible acoustic signals. As illustrated in Fig. 1, human movement causes slight changes in the lengths of the red acoustic paths, resulting in corresponding differences between successive RIR estimates. To enable further study of micromotion-based localization, an appropriate dataset is required. To the best of our knowledge, the only existing dataset for acoustic human localization is SoundCam, which provides 5,000 ten-channel RIR estimates captured at unique human positions in three rooms [3]. However, only one measurement was performed for each position, providing no successive RIR measurements that could capture temporal variations caused by human micromotion. In contrast, the Arni dataset contains five successive RIR measurements for each of 5,342 absorption configurations, but was recorded without a human present [13].

In this paper, we introduce SuRIR, a dataset comprising 15,360 RIR estimates. Unlike both SoundCam and the Arni dataset, SuRIR combines successive RIR measurements with the presence of a human by providing four consecutive RIR estimates for each measurement configuration. In addition, SuRIR considers five movement conditions that vary in movement amount, posture, and orientation, whereas SoundCam provides measurements only under a single movement condition. SuRIR deliberately trades positional diversity for temporal resolution: It covers 150 room-participant-position-condition combinations, compared with 5,000 room-participant-position combinations in SoundCam, but captures the within-position temporal variations that micromotion-based localization relies on. The dataset also includes 960 RIRs measured with a life-sized mannequin in place of a human, providing a baseline condition without human micromotion. A detailed comparison of the SuRIR and SoundCam datasets is provided in Table 1. Note that unlike SoundCam, where the participants contributed an irregular number of human positions, all three participants in SuRIR were located at the same five human positions in each of the two rooms, resulting in $2 \times 3 \times 5 \times 5 = 150$ measurement scenarios across rooms, participants, positions, and movement conditions. The dataset is publicly available at `https://doi.org/10.5281/zenodo.21494888`.

2. Dataset Acquisition

Two Neumann KH 80 DSP loudspeakers were employed, each equipped with three top-mounted AKG C417 PP microphones as shown in Fig. 2. The microphones were routed through an RME Micstasy 8-channel preamplifier into an RME MADIface audio interface, while the loudspeakers were connected directly to the MADIface. The heights of the acoustic centers of the loudspeakers and the microphones were approximately 1.13 m and 1.26 m, respectively. The 6-channel RIRs were estimated using Farina's method [14], with exponential sine sweeps spanning 20 Hz–24 kHz, recorded by the microphones at a sampling rate of 48 kHz and a bit depth of 24 bit. The RIR estimates were subsequently bandpass filtered using both a low-pass and a high-pass zero-phase second-order-sections fifth-order Butterworth filter with cutoff frequencies 50 Hz and 20 kHz, respectively. The filtered RIRs were cropped to 1 s, and corrected for initial offsets introduced by the recording chain. All RIRs were subsequently normalized relative to each other, such that the peak amplitude of the loudest RIR was scaled to 1 separately for each room.

Two sweep lengths were employed for RIR estimation: a short sweep of 1 s and a long sweep of 5 s, each followed by 1 s of silence. The sweeps were played back in contiguous blocks, one block per loudspeaker and sweep length. First, Loudspeaker 1 (LS1) emitted four repetitions of the short sweep, followed by four repetitions from Loudspeaker 2 (LS2). Both sequences were captured simultaneously by the two

Table 2: Relevant room parameters.

Parameter	**Room A**	**Room B**
T_{60}	0.38 s	0.68 s
Temperature	28 °C	22 °C
Noise floor	46 dB(Z)	44 dB(Z)
Excitation level	92 dB(Z)	80 dB(Z)
Room height	2.87 m	3.15 m
Room volume	$\approx 63\,\mathrm{m}^3$	$\approx 55\,\mathrm{m}^3$

circular microphone arrays, with Microphone Channels 1–3 corresponding to the three microphones attached to LS1, and Microphone Channels 4–6 corresponding to those attached to LS2. Immediately thereafter, the same sequence was repeated using the long sweeps. For each measurement scenario, this yielded 6 (microphones) $\times$ 8 (sweeps) $\times$ 2 (loudspeakers) = 96 RIRs with a total measurement time of 64 s.

Measurements were conducted in two acoustic environments. Room A resembles a furnished living room, featuring three solid walls and one wall with recessed windows covered by curtains. The floor is covered with a thin carpet, a couch was placed near the center of the room, and corner absorbers were installed at the wall-ceiling junction along two walls. Room B represents an unfurnished space with four solid walls, with a thin carpet laid on the floor, and the ceiling partially featuring acoustic ceiling treatment. Table 2 lists the main room parameters and photographs of the rooms are shown in Fig. 3. Note that T_{60} was obtained by backward integration [15] followed by a linear fit to the energy decay curve, where a 30 dB decay range was employed for the regression. The final T_{60} value is the average of 12 individual estimates recorded with the microphones attached to LS1 and LS2 (6 microphones $\times$ 2 loudspeakers). The excitation levels were determined using an NTi Audio XL2 sound level meter, measuring white-noise playback from each loudspeaker at a distance of 1 m. Two different excitation levels were used, resulting from the respective recording settings for each room. The room temperatures were measured using a consumer-grade thermometer and reflect the value observed throughout all acoustic measurements.

Three participants took part in the measurements, with one of the participants being one of the authors

Table 3: Utilized movement conditions, where *moving* denotes subtle whole-body sway and *facing* refers to the orientation of the participant's entire body rather than that of the head alone. *Left* indicates facing the left wall, while *down* denotes facing parallel to the left wall toward the lower part of the floor plan in Fig. 4.

Condition	**Description**
1	Standing while moving, facing left
2	Standing still, facing left
3	Standing still, facing down
4	Sitting while moving, facing left
5	Sitting still, facing left

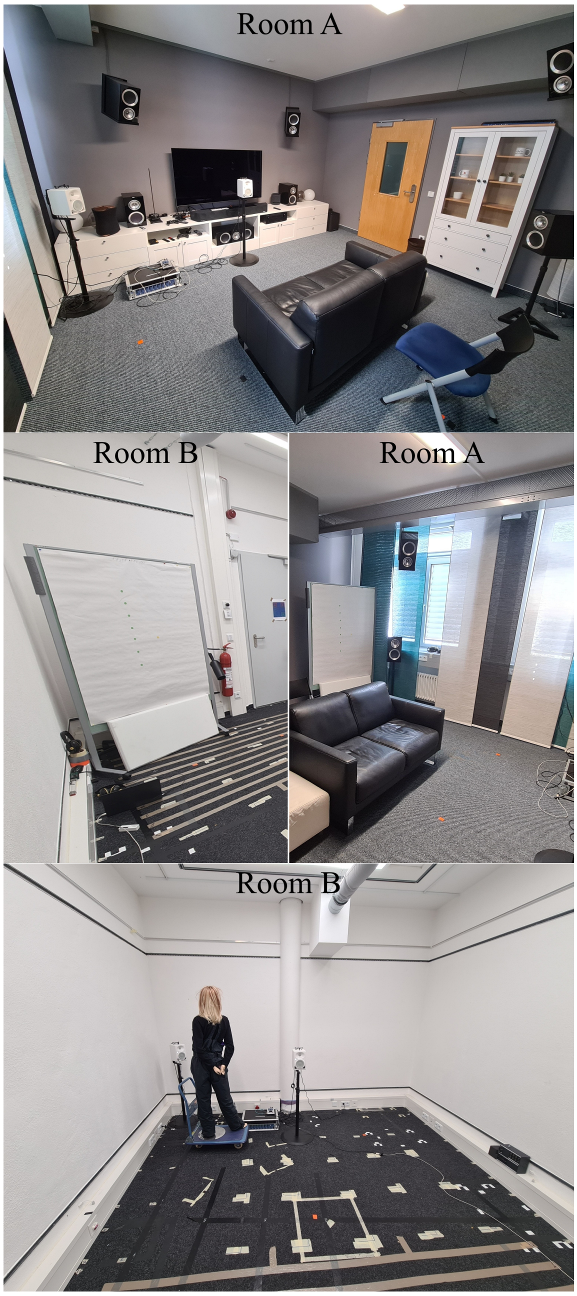

Figure 3: Top: furnished living room (Room A). Middle: poster panel placements in Room A and Room B. Bottom: unfurnished room (Room B).

of this work. All participants wore T-shirts and shorts throughout the measurements, providing comparable clothing coverage across participants, as well as earmuff hearing protection. In each room, five numbered measurement positions were marked on the floor as shown in the floor plans in Fig. 4. All positions and room dimensions were measured using a laser rangefinder. Each participant was measured individually at each of the five positions under five experimental conditions. The conditions varied the participant's posture, movement, and orientation as shown in Table 3. For Conditions 1–3 the participant's feet were positioned symmetrically around the

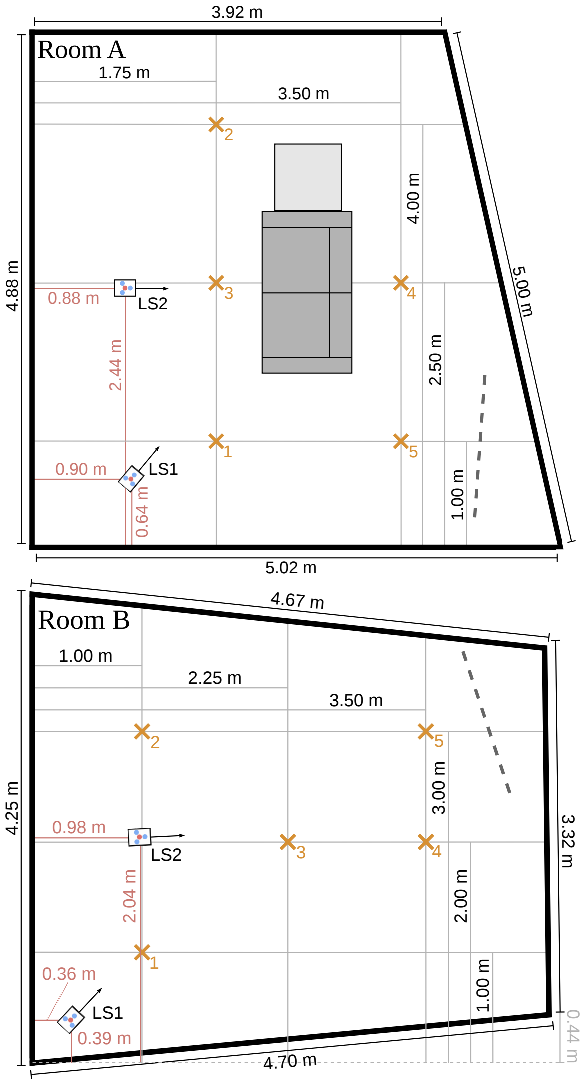

Figure 4: Floor plans of Room A (top) and Room B (bottom). The dashed lines represent the poster panel.

floor markings. The chair used in Conditions 4 and 5 is visible behind the couch in the top image of Fig. 3, and it was relocated for each participant position by aligning the center of the seat cushion with the floor markings. For each combination of room, participant, position, and condition, the complete measurement sequence described above was recorded, resulting in 96 (RIRs) $\times$ 2 (rooms) $\times$ 3 (participants) $\times$ 5 (positions) $\times$ 5 (conditions) = 14,400 RIRs.
Additionally, a rigid, fully clothed life-sized standing mannequin was placed at each of the five positions using a platform trolley and measured under Condition 2, as shown in the bottom image in Fig. 3. These measurements provide a non-moving reference and can therefore be used to investigate changes in the RIRs that are unrelated to human movement. The mannequin measurements contribute an additional 960 RIRs to the dataset. For illustration, Fig. 5 shows the beginning of the first 6-channel RIR measured in Room A with the mannequin at Position 5 using LS1 and the 1 s sweep.

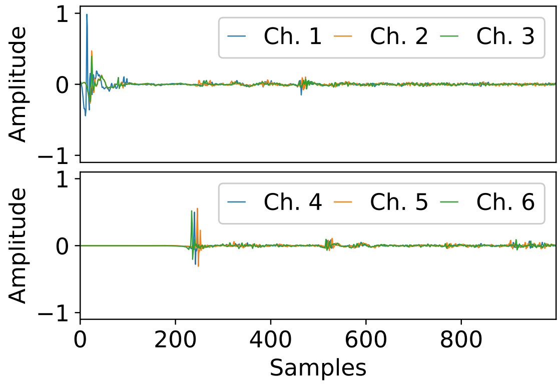

Figure 5: First 6-channel RIR recorded in Room A with the mannequin at Position 5, using LS1 and the 1 s sweep.

For all measurements, a 2 m-high poster panel was placed in the corner of each room, as shown in Fig. 3 and indicated by the dashed lines in Fig. 4. For measurements involving participants 2 and 3, an experiment supervisor was positioned behind the poster panel; no additional person was present in the room during the measurements involving participant 1 or the mannequin.

3. Dataset Description

The dataset is provided in the standardized Spatially Oriented Format for Acoustics (SOFA) format[1] utilizing the `SingleRoomSRIR` convention. The ground-truth human positions and body orientations are stored in the `ListenerPosition` and `ListenerView` attributes, respectively. The human-present RIR estimates are organized into individual SOFA files following the structure illustrated by the filename `RoomA_Participant01_Condition01_Sweep1s.sofa`, where the room, participant, condition, and sweep length are encoded in the filename. RIRs obtained with the 1 s and 5 s sweeps are stored in separate SOFA files, allowing the two sets of measurements to be evaluated independently. In total, the human measurements comprise 60 SOFA files.
Each SOFA file contains 40 $\times$ 6 RIRs, corresponding to five human positions, two loudspeakers, four repetitions and six microphone channels. For each human position, the first four RIRs correspond to the four repetitions recorded with LS1, followed by four RIRs recorded with LS2. This ordering is repeated for all five human positions. Thus, each file contains the following sequence:

$$\underbrace{\begin{matrix}\mathrm{LS1}_{R1},\ \mathrm{LS1}_{R2},\ \mathrm{LS1}_{R3},\ \mathrm{LS1}_{R4},\\ \mathrm{LS2}_{R1},\ \mathrm{LS2}_{R2},\ \mathrm{LS2}_{R3},\ \mathrm{LS2}_{R4},\end{matrix}}_{\text{Human Position 1}} \\ \vdots \\ \underbrace{\begin{matrix}\mathrm{LS1}_{R1},\ \mathrm{LS1}_{R2},\ \mathrm{LS1}_{R3},\ \mathrm{LS1}_{R4},\\ \mathrm{LS2}_{R1},\ \mathrm{LS2}_{R2},\ \mathrm{LS2}_{R3},\ \mathrm{LS2}_{R4}\end{matrix}}_{\text{Human Position 5}},$$

[1] See https://www.sofaconventions.org/ and [16] for the formal definition.

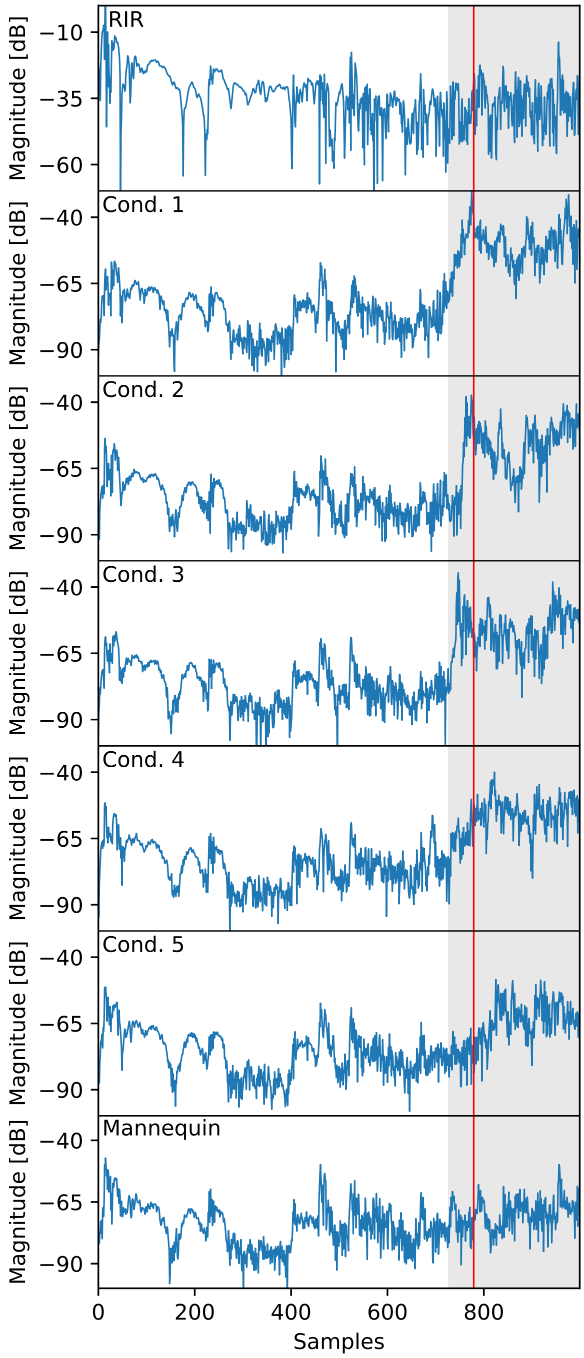

Figure 6: Mean absolute RIR of Condition 1 (top) and mean absolute differential RIRs for Conditions 1–5 and the mannequin (bottom), computed from four successive RIR estimates for Room A, Participant 1, LS1, Microphone Channel 1, and human/mannequin Position 5, using the 1 s sweeps. The red line indicates the expected arrival time of the human-induced reflection. Prior to the reflection, all differential RIRs exhibit a similar temporal structure and reflect the underlying RIR shape. Afterward, their energy increases progressively from Condition 5 to 1 and becomes indistinguishable from the differential noise floor for the mannequin (see highlighted areas). Elevated energy persists beyond the human-induced reflection.

where $\mathrm{LS}i_{\mathrm{R}j}$ denotes a 6-channel measurement obtained using Loudspeaker i during repetition j.
The motionless mannequin recordings are provided in four SOFA files following the naming structure `RoomA_Mannequin_Sweep1s.sofa`. Each file contains 40×6 RIRs with the same sequence as the human-present SOFA files, with the filename encoding the room and sweep length. All SOFA files additionally encode the spatial metadata for the human, microphones, loudspeakers, and room geometry, with all positions specified in Cartesian coordinates in meters. The room geometry is specified as a non-rectangular four-sided floor plan, and the `GLOBAL_RoomGeometry` attribute stores the coordinates of the room corners and the room height. Further metadata document the measurement environment, room volume and temperature. The files were generated using the `sofar`[2] Python package.

4. Example: Differential RIR Analysis

As an example of how the SuRIR dataset can be used to investigate the impact of human micromotion on successive RIR estimates, we computed all pairwise differences between the four successive RIR estimates for Room A, Participant 1, LS1, Microphone Channel 1 and human/mannequin Position 5, using the 1 s sweeps. This yielded six differential RIRs per condition. The absolute values of these differential RIRs were subsequently averaged, whereas the absolute values of the four original RIRs for Condition 1 were averaged separately. The log-scaled magnitudes of the resulting RIRs and differential RIRs are shown in Fig. 6, where the expected arrival time of the human-induced reflection was calculated from the loudspeaker–human–microphone path length using an approximate speed of sound of 348 m/s, based on the room temperature of 28 °C. It can be observed that prior to the human-induced reflection, the differential RIRs exhibit a similar temporal structure and reflect the underlying RIR shape. This structure has previously been attributed to local temperature fluctuations in the acoustic environment [10]. Afterward, the energy level increases progressively from Condition 5 to Condition 1 (see highlighted areas), reflecting the combined effects of human movement and effective reflection area. In particular, the moving conditions exhibit higher energy levels than their respective stationary conditions, while the smaller effective reflection area in the seated conditions further reduces the reflected energy, which becomes indiscernible from the differential noise floor in the mannequin (no-movement) condition. Moreover, it can be observed that elevated energy levels persist beyond the human-induced reflection, indicating that human micromotion influences the RIR over an extended temporal region.

5. Conclusion

SuRIR comprises 15,360 RIR estimates that enable the study of subtle temporal variations arising from

[2] https://pypi.org/project/sofar/

the presence of a silent human. The inclusion of successive measurements enables differential RIR-based analysis, making micromotion-induced changes observable while also capturing variations arising from the acoustic environment. Beyond its intended application to micromotion-based localization, SuRIR supports the study of short-term RIR variability, since the mannequin condition isolates environmental drift from human-induced change. Future work will extend the dataset to a larger participant pool and add ground-truth motion capture, enabling the movement conditions to be related to quantified physical displacement.

6. Ethics Statement

All participants agreed to the public release of their anonymized data. Two participants additionally provided explicit consent to be acknowledged by name in this publication. No personally identifiable information is included in the dataset itself.

7. Acknowledgments

We thank Yaqiong Hou and Yijiong Wang for posing as silent humans during the dataset acquisition. We also thank Felix Knauff, who designed and 3D-printed the microphone array mounts.

References

[1] K. Kim, S. Wang, H. Ryu, and S. Q. Lee, "Acoustic-based position estimation of an object and a person using active localization and sound field analysis," *Applied Sciences*, vol. 10, no. 24, 2020. DOI: 10.3390/app10249090.

[2] Z. Yang, X. Fan, V. Isler, and H. S. Park, "PoseKernelLifter: Metric lifting of 3D human pose using sound," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, US, 2022, pp. 13 169–13 179. DOI: 10.1109/CVPR52688.2022.01283.

[3] M. Wang, S. Clarke, J.-H. Wang, R. Gao, and J. Wu, "SoundCam: A dataset for finding humans using room acoustics," in *Advances in Neural Information Processing Systems 36*, New Orleans, Louisiana, USA: Neural Information Processing Systems Foundation, Inc., 2023, pp. 52 238–52 264. DOI: 10.52202/075280-2274.

[4] S. S. Bhattacharjee, A. Jonas Fuglsig, F. Christensen, J. R. Jensen, and M. Græsbøll Christensen, "Robust fixed-filter sound zone control with audio-based position tracking," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025. DOI: 10.1109/ICASSP49660.2025.10889419.

[5] J. Lawrence, C. Tuna, A. Walther, and N. Peters, "Acoustic position estimation of a silent listener," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025. DOI: 10.1109/ICASSP49660.2025.10888595.

[6] G. Götz, S. J. Schlecht, and V. Pulkki, "A dataset of higher-order ambisonic room impulse responses and 3D models measured in a room with varying furniture," in *Immersive and 3D Audio: from Architecture to Automotive (I3DA)*, 2021. DOI: 10.1109/I3DA48870.2021.9610933.

[7] A. Melikov, U. Krüger, G. Zhou, T. Madsen, and G. Langkilde, "Air temperature fluctuations in rooms," *Building and Environment*, vol. 32, no. 2, pp. 101–114, 1997. DOI: https://doi.org/10.1016/S0360-1323(96)00034-0.

[8] K. Prawda, S. J. Schlecht, and V. Välimäki, "Short-time coherence between repeated room impulse response measurements," *The Journal of the Acoustical Society of America*, vol. 156, no. 2, pp. 1017–1028, 2024. DOI: 10.1121/10.0028172.

[9] K. Prawda, "Sensitivity of room impulse responses in changing acoustic environment," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025. DOI: 10.1109/ICASSP49660.2025.10889026.

[10] J. Lawrence, Z. Xu, A. Walther, N. Peters, and E. A. P. Habets, "Integrating atmospheric fluctuations into the image source method," *JASA Express Letters*, vol. 6, no. 7, p. 072 401, 2026. DOI: 10.1121/10.0044473.

[11] V. E. Gonzalez-Sanchez, A. Zelechowska, and A. R. Jensenius, "Correspondences between music and involuntary human micromotion during standstill," *Frontiers in Psychology*, vol. 9, 2018. DOI: 10.3389/fpsyg.2018.01382.

[12] Y. Ma, G. Zhou, and S. Wang, "Wifi sensing with channel state information: A survey," *ACM Comput. Surv.*, vol. 52, no. 3, 2019. DOI: 10.1145/3310194.

[13] K. Prawda, S. J. Schlecht, and V. Välimäki, *Dataset of impulse responses from variable acoustics room Arni at Aalto Acoustic Labs*, Version v1, Zenodo, 2022. DOI: 10.5281/zenodo.6985104. Accessed: Jul. 31, 2026.

[14] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Proc. of the 108th Audio Engineering Society Convention*, Paris, France, 2000.

[15] M. R. Schroeder, "New Method of Measuring Reverberation Time," *The Journal of the Acoustical Society of America*, vol. 37, no. 3, pp. 409–412, 1965. DOI: 10.1121/1.1909343.

[16] P. Majdak, F. Zotter, F. Brinkmann, J. De Muynke, M. Mihocic, and M. Noisternig, "Spatially Oriented Format for Acoustics 2.1: Introduction and Recent Advances," *Journal of the Audio Engineering Society*, vol. 70, pp. 565–584, 2022.