DOI: 10.61782/fa.2025.0696

FORUM ACUSTICUM EURONOISE 2025

PRELIMINARY EVALUATION OF A LOW-COST DEVICE FOR BINAURAL RECORDING AND SOUNDSCAPE MONITORING

Juan Miguel Navarro Ruiz[1*] **Massimiliano Masullo**[2]
[1] Research Group in Advanced Telecommunications (GRITA),
UCAM Universidad Católica de Murcia, 30107 Guadalupe, Spain
[2] Department of Architecture and Industrial Design,
Università degli Studi della Campania "Luigi Vanvitelli", 81031 Aversa, Italy

ABSTRACT

The growing interest in soundscape analysis has highlighted the need for accessible tools capable of capturing and processing audio with high spatial sound fidelity. This study presents a low-cost device designed for binaural recording and soundscape monitoring. The device employs a microphone array to capture sound signals and subsequently applies signal processing techniques to synthesize a binaural output.

A preliminary evaluation was conducted to assess the device's performance in a controlled urban environment characterized by road traffic noise. The evaluation compared recordings obtained with the prototype device against reference recordings captured using in-ear microphones in an artificial head and worn by human. Key metrics, including spatial accuracy, and psychoacoustic parameters, were analyzed to determine the device's suitability for soundscape studies.

Results suggest that the proposed device provides an effective approximation of binaural audio, with certain limitations inherent to its low-cost design. These findings support the potential of the device as a scalable solution for soundscape monitoring, particularly in applications where traditional binaural equipment may be cost-prohibitive. Future work will focus on refining the signal processing algorithms and expanding the evaluation to diverse soundscapes.

**Corresponding author*: *jmnavarror@ucam.edu.*

Keywords: *soundscapes, binaural, traffic noise, IoT, monitoring*

1. INTRODUCTION

The study of soundscapes has emerged as a crucial area in acoustic research, particularly in the context of urban planning, environmental monitoring, and public health. Soundscape analysis aims to understand how people perceive and interact with their acoustic environment, providing valuable insights to design more sustainable and livable urban spaces [1, 2]. A fundamental aspect of soundscape studies is the ecological capture of spatial auditory information, which requires high-fidelity binaural recordings. These recordings preserve the spatial cues necessary for human auditory perception, including interaural level and time differences, which are essential for evaluating the subjective experience of sound environments [3].

Traditional binaural recording techniques employ artificial heads, binaural headsets or in-ear binaural microphones worn by human subjects. These devices offer high spatial sound accuracy but are often expensive, equipment and/or human resources, limiting their widespread use in large-scale or long-term soundscape monitoring [4, 5]. Additionally, logistical challenges arise when deploying multiple binaural recording units in diverse environments, further restricting their applicability. To address these limitations, recent advancements in the Internet of Things (IoT) and digital signal processing have facilitated the development of alternative low-cost solutions. Even though low-cost IoT solutions have made significant progress in the soundscape monitoring field, challenges are still open in capturing perceptually accurate data. One approach has focused on integrating low-cost acoustic sensors with

AI-based analysis for urban soundscape research, but the system relies on monaural recording and does not incorporate psychoacoustic parameter extraction [6]. Another initiative has introduced a FIWARE-based wireless acoustic sensor network designed for urban sound classification, utilizing machine learning models; however, it lacks binaural capabilities and the ability to assess perceptual sound attributes [7]. A different solution has explored a multipurpose IoT sensor for biologging and soundscape applications, offering flexibility in data collection but still relying on monaural recording without advanced psychoacoustic analysis [8]. While these efforts contribute to the accessibility and scalability of soundscape monitoring, they also reveal gaps in binaural recording and perceptual sound analysis, highlighting the need for enhanced solutions that improve fidelity and psychoacoustic processing.

In this study, a preliminary evaluation of a low-cost IoT device designed for binaural recording and soundscape monitoring is presented. The proposed system employs a microphone array to capture sound signals and implements signal processing algorithms to synthesize binaural output [9], [10]. A controlled experiment was conducted in an urban environment dominated by road traffic noise to assess the device's performance against a reference system consisting of human-worn binaural headphones. Key performance metrics, including including spatial accuracy, and psychoacoustic parameters, were analyzed to determine the suitability of the proposed device for soundscape research.

The remainder of this paper is structured as follows: Section 2 describes the materials and methods, including the system architecture, experimental setup, and evaluation methodology. Section 3 presents the results and discussion, highlighting the device's performance in comparison with the reference system. Finally, Section 4 concludes the study and outlines potential directions for future work.

2. MATERIALS AND METHODS

The following section provides a comprehensive description of the methodology employed in this study. In the first part, the evaluated device is described in terms of its hardware and software components. Subsequently, the experimental setup is presented, detailing the environment and reference system used for evaluation. The third part outlines the data collection procedure, including the recording protocols and evaluation criteria. Finally, data processing and analysis methods are described, emphasizing the techniques used to assess the performance of the device.

2.1 Description of the Device

The low-cost IoT device was designed to approximate the spatial fidelity of conventional binaural recording systems. The system is structured into two main components: the audio acquisition subsystem and the processing core, which together facilitate real-time data capture, processing, and transmission.

The audio acquisition is based on a Sony PlayStation Eye camera, which features a four-microphone array. This array captures sound with 16-bit depth at a sampling rate of 16 kHz, offering a signal-to-noise ratio of 90 dB. The array configuration is specifically designed to replicate spatial hearing characteristics, with an inter-microphone distance of approximately 62 mm and reversed polarity in the central microphones to enhance spatial audio capture. Calibration procedures were conducted in a semi-anechoic chamber comparing the results with those of a Rion NL-05 sound level meter to ensure reliable and accurate measurements.

The processing core is based on a Raspberry Pi 3 Model B, which serves as the central processing and control unit. This platform is equipped with a 1.2 GHz 64-bit quad-core ARMv8 CPU, 1 GB of RAM, four USB ports, and integrated 802.11n Wireless LAN.

The system captures audio in configurable time-window segments and utilizes a signal processing pipeline to synthesize binaural output. The stages of the pipeline are shown in Fig. 1.

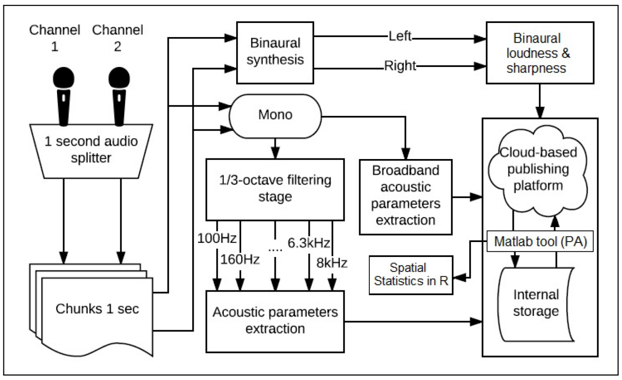

Figure 1. Diagram of the signal processing pipeline to synthesize binaural output. [9]

The system estimates the direction of incoming sounds using time-frequency domain analysis, employing the Short-Term Fourier Transform and assuming sparsity of audio sources to calculate the Direction-Of-Arrival (DOA). Based on the estimated DOA, the system applies Head-Related Transfer Functions (HRTFs) to synthesize binaural audio. The selected HRTF corresponds to the estimated azimuth, ensuring an accurate spatial representation. The device computes objective psychoacoustic metrics, following Zwicker's and Moore's loudness models. The processed data, including acoustic parameters and raw recordings, are stored locally on the Raspberry Pi and transmitted to a cloud platform for further analysis and archiving.

The entire processing pipeline is implemented in Python, using the Advanced Linux Sound Architecture (ALSA) for audio acquisition. The modular software architecture allows independent operations of each processing stage, enhancing robustness and facilitating updates or modifications. The system periodically uploads processed data to an IoT cloud platform, enabling real-time monitoring and remote data retrieval.

2.2 Experimental Setup

As a preliminary experiment, the performance of the proposed device was evaluated in an indoor environment, specifically an open-plan office of the Universidad Católica de Murcia, Spain. This setting is characterized by stable and consistent acoustic conditions, with predominant noise sources originating from vocal interactions among staff members. The environment was chosen due to its relevance in soundscape studies, as speech-related acoustic fields are common in office and educational settings, making it an ideal scenario to assess the device's performance in capturing binaural audio under realistic indoor conditions. The recording area was situated at the center of the right-hand row of desks within the open-plan office, which is composed of three parallel rows of desks. This positioning was chosen to minimize reflections from nearby furniture and maintain consistency throughout the data collection process.

To assess the accuracy of the proposed device, the recordings were compared against two reference systems. The first reference system consisted of a Roland CS-10EM binaural headphones connected to a Zoom H4n Pro 4-channel digital recorder. The Roland CS-10EM headphones, featuring built-in omnidirectional microphones, were placed in the listener's ear canals. The second reference system consists of the a OKM II Classic Binaural Microphones mounted on a Soundman Dummy Head, also connected to a Zoom H4n Pro. The Zoom H4n Pro recorder provided high-quality audio capture with 24-bit/96 kHz resolution, ensuring accurate documentation of the acoustic environment.

Each recording session lasted 10 minutes, ensuring adequate signal representation for analysis. Both systems were synchronized to ensure simultaneous recording sessions under identical acoustic conditions. Synchronization was achieved using a common time reference and initiating recordings simultaneously to eliminate temporal discrepancies between datasets. Calibration procedures were performed before the experiments to ensure that the signal levels recorded by both systems were comparable, thereby minimizing potential bias due to differences in microphone sensitivity or frequency response.

The experimental setup was designed to allow both the proposed device and the reference system to capture the same acoustic events from identical spatial perspectives. This approach enabled a direct comparison between the binaural outputs of the low-cost device and the reference microphones, focusing on spatial accuracy and perceptual fidelity.

2.3 Data Processing and Analysis

All recorded signals were processed and analyzed using MATLAB. The analysis workflow included signal segmentation, feature extraction, and statistical comparison between the proposed device and the reference system. The 10-minute recording session was segmented into 10-second windows, and for each window, several key performance metrics were calculated, followed by the extraction of statistical descriptors from the entire session.

Spatial accuracy was assessed by comparing the interaural level differences (ILD) in dB, the interaural time differences (ITD) in ms, and the interaural cross-correlation (IACC). Perceptual fidelity was evaluated through an objective analysis of psychoacoustic metrics, including loudness (LO) in sone, sharpness (SH) in acum, roughness (RO) in asper, and fluctuation strength (FS) in vacil, using MATLAB's psychoacoustic modeling tools.

3. RESULTS AND DISCUSSION

This section presents the preliminary results obtained from the evaluation of the proposed low-cost binaural recording device analyzing a single 10-minute recording session, with particular emphasis on a detailed examination of a 10-second window within that session to assess the suitability of the system. The performance of the device was compared with two reference systems consisting of human-worn binaural microphones and a dummy head equipped with binaural microphones. Moreover, to evaluate the impact of the binaural synthesis algorithm on the recorded signals, the results include the metrics calculated both before (denoted as pre in tables and figures) and after (denoted as post in tables and figures) the signal processing implemented in the device. The results are presented in terms of average values, highlighting the comparative performance of the proposed device relative to the reference system.

Tab. 1 presents the spatial accuracy metrics obtained for the open-plan office environment, including values recorded before and after the signal processing applied in the device.

Table 1. Spatial accuracy analysis metrics for the open-plan office

Capture	ILD (dB)	ITD (ms)	IACC
Device (pre)	-0.18	0.0000	0.97
Device (post)	3.24	-0.5000	0.24
Dummy	0.88	-0.5575	0.50
Human-worn	0.28	-0.5568	0.36

The proposed device demonstrated significant changes in spatial accuracy after the binaural synthesis process. Notably, the ILD increased from -0.18 dB to 3.24 dB, indicating an adjustment in interaural level representation. Similarly, the ITD shifted from 0 ms to -0.5 ms, suggesting an alteration in the interaural time alignment. The IACC also decreased from 0.97 to 0.24, reflecting a reduction in spatial coherence after processing.

When compared with the reference systems, the device's post-processed ITD was closer to the values observed in the dummy and human-worn setups. However, the ILD values remained less consistent, highlighting a limitation in accurately reproducing level-based spatial cues. The reduction in IACC after processing indicates a potential loss of spatial coherence, which could affect the perceptual quality of the binaural output.

These findings suggest that, while the proposed system is capable of approximating spatial cues, improvements in the signal processing algorithm and microphone array configuration are necessary to enhance spatial fidelity, particularly regarding ILD accuracy and spatial coherence.

Tab. 2 presents the perceptual fidelity analysis metrics obtained for the open-plan office environment, including values recorded before and after the signal processing applied in the device.

Table 2. Perceptual fidelity analysis metrics for the open-plan office

Recording	LO (sone)	SH (acum)	RO (asper)	FS (vacil)
Device (pre)	16.32	1.10	0.08	0.16
Device (post)	12.79	1.07	0.10	0.04
Dummy	13.36	1.11	0.10	0.16
Human-worn	8.50	1.15	0.07	0.08

The analysis revealed that the proposed device's loudness decreased from 16.32 sone to 12.79 sone, aligning more closely with the dummy reference, 13.36 sone, but remaining notably higher than the human-worn microphones, 8.50 sone. Sharpness values remained relatively consistent across all setups, with a slight decrease after processing. Moreover, sharpness values around 1 acum show a balanced sound, with a certain brightness, characteristic of a speech signal.

Roughness showed a small increase from 0.08 asper to 0.10 asper, which corresponds with the dummy reference value but is slightly higher than the human-worn system, 0.07 asper. Fluctuation strength exhibited a more pronounced reduction from 0.16 vacil to 0.04 vacil, diverging from both the dummy and human-worn references, which recorded values of 0.16 and 0.08 vacil, respectively.

These results suggest that the proposed device, after binaural synthesis, more accurately replicates loudness and roughness when compared to the dummy reference. However, the reduced fluctuation strength indicates a potential loss in the dynamic variability of the signal, which could impact the perceived naturalness of the soundscape. Adjustments in the signal processing algorithm could help preserve fluctuation cues while maintaining perceptual fidelity.

To further analyze the frequency response of the proposed device and the effect of the binaural synthesis algorithm, the specific loudness versus frequency for a representative 10-second recording is examined. Fig. 2 shows the specific loudness spectra for both the device, before a) and after processing b), and the reference systems, dummy head c) and human-worn binaural microphones d).

In terms of frequency distribution, both device graphs, Fig. 2a and Fig. 2b, display a concentration of loud-

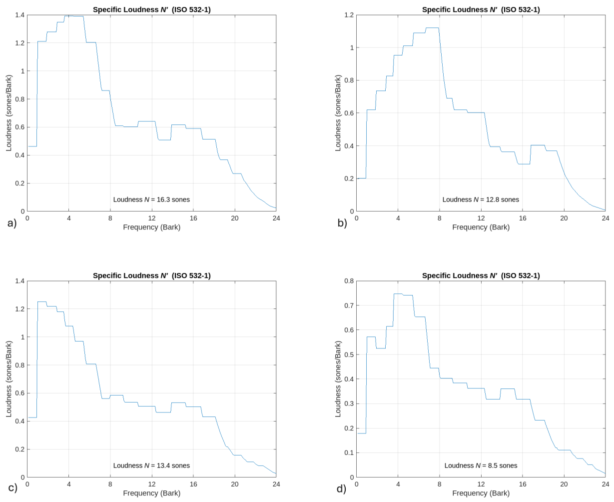

Figure 2. Specific loudness versus frequency for an example of 10 second-recording. a) Device (pre) b) Device (post) c) Dummy d) Human-worn

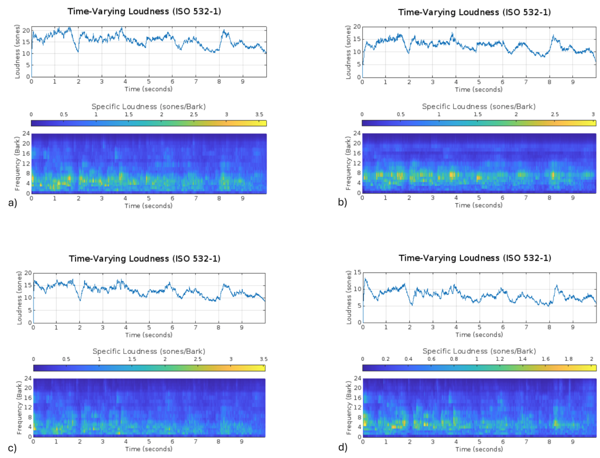

Figure 3. Time-varying loudness for an example of 10 second-recording. a) Device (pre) b) Device (post) c) Dummy d) Human-worn

ness predominantly in the low-to-mid frequency range (0-8 Bark), followed by a gradual decline up to approximately 16 Bark. This pattern is consistent with the human-worn microphone response, Fig. 2d, though with generally lower peak values. In contrast, the dummy head response in Fig. 2c shows a concentration in the lower frequencies, which also exhibits a more uniform distribution across the mid-frequency range, maintaining higher levels up to 16 Bark before decreasing.

Overall, the specific loudness analysis indicates that while the binaural synthesis algorithm can adjust the overall intensity, it does not fully compensate for spectral discrepancies inherent to the device's hardware configuration. Future improvements could include the integration of higher-bandwidth microphones to better capture high-frequency content and enhance perceptual accuracy.

To assess the dynamic behavior of the proposed device, the time-varying loudness for a representative 10-second recording is analyzed. Fig. 3 shows the time-varying loudness (ISO 532-1) for both the device, before and after processing, and the reference systems. The graphs illustrate the variation of loudness over time and its distribution across frequency bands.

Overall, the pattern of level variations, in the time-domain representation in Fig. 3 is similar across the four recordings, indicating that despite differences in loudness values, the temporal dynamics of the sound are consistently captured by both the proposed device and the reference systems. In Fig. 3a, the device's pre-processed signal exhibits noticeable fluctuations in loudness, with values ranging between approximately 10 and 20 sones. After processing, Figure 3b, the device shows a reduction in overall loudness variation, stabilizing around 12 to 18 sones. This decrease suggests that the binaural synthesis algorithm reduces the intensity dynamics, potentially due to signal averaging or noise reduction mechanisms. Comparatively, the dummy head recording, Fig. 3c, presents a loudness range similar to the post-processed device, with values between 10 and 18 sones. On the other hand, the human-worn microphone setup, Fig. 3d, exhibits consistently lower loudness values, ranging from 8 to 15 sones.

The spectrograms in Fig. 3 reveal that the loudness distribution across frequency bands differs notably between the device and both reference systems. Both the pre-processed and post-processed device signals show dominant loudness contributions concentrated in the lower frequency range (below 8 Bark), whereas both reference systems (dummy head and human-worn) exhibit a more evenly distributed loudness profile, particularly across mid-to-high frequencies. This discrepancy can be attributed to the limited frequency response of the PS3 Eye microphones, which do not capture higher frequencies as effectively as the reference binaural microphones. Consequently, the device may not fully replicate the time-varying perceptual attributes of the soundscape, especially in scenarios with prominent high-frequency components.

In summary, the time-varying loudness analysis indicates that while the binaural synthesis algorithm smooths the loudness profile, it does not fully compensate for the hardware limitations of the proposed device. Further improvements could involve using higher-bandwidth microphones to capture a more accurate representation of dynamic acoustic environments.

4. CONCLUSIONS

This study evaluated a low-cost IoT device for binaural recording and soundscape monitoring in an indoor environment with vocal noise. The device demonstrated reasonable spatial cue approximation, showing potential as a cost-effective alternative for soundscape studies. However, differences in interaural cues and psychoacoustic metrics indicate areas for improvement.

The spatial accuracy analysis showed that after processing, the device achieved interaural time differences similar to reference systems, but interaural level differences remained inconsistent. This suggests limitations in reproducing level-based spatial cues. The reduced interaural cross-correlation after processing indicates a possible loss of spatial coherence, affecting binaural output quality.

After processing, the device's psychoacoustic metrics were closer to the dummy head, especially for sharpness and loudness. However, discrepancies were noted in roughness and fluctuation strength, likely due to the PS3 Eye microphone's limited frequency bandwidth, which may hinder accurate high-frequency representation.

The time-varying loudness analysis showed consistent temporal dynamics across the signals, though the device's processed signal displayed reduced variability, likely due to the synthesis algorithm. The device signals concentrated loudness in lower frequencies, while reference systems had a more even distribution, particularly at higher frequencies.

Despite some limitations, the device demonstrates the potential for affordable binaural recording, useful for soundscape monitoring. Future work should improve spatial accuracy, loudness estimation, and frequency capture, focusing on diverse environments and refining algorithms for better accuracy.

5. ACKNOWLEDGMENTS

This work is the result of the stay 22208/EE/23 funded by the Fundación Séneca - Science and Technology Agency of the Region of Murcia, under the Regional Program for Mobility, International Collaboration, and Knowledge Exchange "Jiménez de la Espada" (Action Plan 2023).

6. REFERENCES

[1] J. Kang, "Soundscape in city and built environment: current developments and design potentials," *City and Built Environment*, vol. 1, no. 1, p. 1, 2023.

[2] A. Fiebig and B. Schulte-Fortkamp, "How to put soundscape into practice," in *Soundscapes: Humans and Their Acoustic Environment*, pp. 313–330, Springer, 2023.

[3] M. Yang and M. Masullo, "Combining binaural psychoacoustic characteristics for emotional evaluations of acoustic environments," *Applied Acoustics*, vol. 210, p. 109433, 2023.

[4] G. Brambilla and A. Fiebig, "Measurements and techniques in soundscape research," in *Soundscapes: humans and their acoustic environment*, pp. 185–214, Springer, 2023.

[5] O. Acosta, L. Hermida, M. Herrera, C. Montenegro, E. Gaona, M. Bejarano, K. Gordillo, I. Pavón, and C. Asensio, "Remote binaural system (rbs) for noise acoustic monitoring," *Journal of Sensor and Actuator Networks*, vol. 12, no. 4, p. 63, 2023.

[6] J. Wang, C. Li, Y. Lin, C. Weng, and Y. Jiao, "Smart soundscape sensing: A low-cost and integrated sensing system for urban soundscape ecology research," *Environmental Technology & Innovation*, vol. 29, p. 102965, 2023.

[7] P. Arce, D. Salvo, G. Piñero, and A. Gonzalez, "Fiware based low-cost wireless acoustic sensor network for monitoring and classification of urban soundscape," *Computer Networks*, vol. 196, p. 108199, 2021.

[8] D. Vasconcelos and N. J. Nunes, "A low-cost multipurpose iot sensor for biologging and soundscape activities," *Sensors*, vol. 22, no. 19, p. 7100, 2022.

[9] J. E. Noriega-Linares, A. Rodriguez-Mayol, M. Cobos, J. Segura-García, S. Felici-Castell, and J. M.

Navarro, "A wireless acoustic array system for binaural loudness evaluation in cities," *IEEE Sensors Journal*, vol. 17, no. 21, pp. 7043–7052, 2017.

[10] J. Segura-Garcia, J. M. Navarro-Ruiz, J. J. Perez-Solano, J. Montoya-Belmonte, S. Felici-Castell, M. Cobos, and A. M. Torres-Aranda, "Spatio-temporal analysis of urban acoustic environments with binaural psycho-acoustical considerations for iot-based applications," *Sensors*, vol. 18, no. 3, p. 690, 2018.