DOI: 10.61782/fa.2023.0339

DEVELOPMENT AND VALIDATION OF A TEST PLATFORM FOR SPATIAL ECOACOUSTIC TECHNOLOGIES

N. P. Le Penru[1]* **A. Borrelli**[2] **B. E. Heath**[5]
R. M. Ewers[3] **S. S. Sethi**[4] **L. Picinali**[1]
[1] Dyson School of Design Engineering, Imperial College London, UK
[2] Department of Civil and Environmental Engineering, Imperial College London, UK
[3] Deparment of Life Sciences, Imperial College London, UK
[4] Deparment of Plant Sciences, University of Cambridge, UK
[5] Deparment of Zoology, University of Cambridge, UK

ABSTRACT

Microphone arrays and direction of arrival estimation algorithms have become increasingly capable and accessible, furthering EcoAcoustics / Passive Acoustic Monitoring (PAM) practitioners' capacity to analyse the spatial features of natural soundscapes and thereby yield richer insights into biodiversity and ecosystem health. However, there is a need for standardised, repeatable methods to comparatively evaluate these technologies. We developed a platform to this end, consisting of a 25-channel spherical loudspeaker array through which spatial natural soundscapes captured with a 19-capsule microphone (*Zylia* ZM-1) can be accurately reproduced and repeatedly re-recorded by spatial PAM devices under evaluation. Here, we first explore how well this lab-based platform can reproduce spatial natural soundscapes, and then present results from a trial of using the platform to evaluate a 6-microphone PAM device developed in our lab. We achieve this by comparing a range of typical ecoacoustic analyses between the field- and lab-based recordings. Further, we specifically use the platform to investigate how device orientation impacts the classification and localisation of avian calls with the software tools BirdNET and HARKBird, respectively. These initial outcomes suggest our test platform can provide key insights into the tuning and performance of spatial PAM devices and ecological analyses of their data. We aim for this platform and approach to be further validated and adopted, to inform the selection of spatial PAM technologies and ease collaboration between their users globally.

Keywords: *Spatial Ecoacoustics, Passive Acoustic Monitoring, Microphone Arrays, Bioacoustics*

1. INTRODUCTION

There are few standardised methods for testing PAM devices that are: (i) in a controlled lab environment free of natural soundscapes' inherent and continual variability, and (ii) tailored to the growing number of PAM devices that feature microphone arrays [1, 2] to capture more insightful, spatial insights into ecosystems [3].

In contrast to existing field recording simulations [4], our test platform is novel in the use of an Ambisonics playback system for replicating real spatial natural soundscapes. We first investigated the efficacy of the system's soundscape reproduction. Then, as a trial of example use, we employed the system to explore how the orientation of a 6-microphone array impacts the results of spatial ecoacoustic analyses, specifically, BirdNET [5] and HARKBird [6]. These applications are usually employed to classify and localise avian calls, respectively, and can be used together to estimate avian species abundance by revealing whether there are several birds of a particular species calling simultaneously from different locations around a mi-

---

***Corresponding author**: neel.le-penru16@imperial.ac.uk.
*

crophone array. However, they are expected to be rather susceptible to device orientation.

2. MATERIALS AND METHODS

2.1 Recording Device, Location and Schedule

We trialed our test platform on a 6-microphone spatial PAM device developed in-house ('6-mic device'). This builds on a monophonic, open-source PAM device [7], but features *Seeed Studio*'s commercially-available 'ReSpeaker 6-Mic Circular Array Kit', which contains 6 small omnidirectional microphones arranged at the vertices of a flat hexagonal board (Fig. 1). We used the intended 16 kHz sampling rate for the 6-mic device which limits file size while allowing a large range of ecoacoustic analyses.

Field recordings were made with the device oriented vertically in July 2022 at Imperial College's Silwood Park campus (Ascot, Berks., UK). We selected 6 sites that span a range of environments (e.g., various levels of forest density and proximity to a stream or human activity) and, as an initial trial, took 10 minute recordings at each site.

2.2 Lab-Based Soundscape Reproduction

We simultaneously captured high-quality spatial field recordings using the *Zylia* 'ZM-1' (sampling rate 48 kHz) mounted directly above the 6-mic device (Fig. 1). These 19-channel recordings were converted to third-order Ambisonics (SN3D normalisation, Furse Malham channel ordering) and decoded to our lab-based reproduction system (25 *Genelec* 8010A loudspeakers mounted around a sphere constructed from two hemispherical climbing frame domes) using the freely-available ICST Ambisonics externals for *Max/MSP* [8]. After calibrating the reproduction system, we re-recorded the lab-replicated 10 minute field recordings from each site with the 6-mic device positioned at the centre of the loudspeaker sphere – first with the 6-mic device oriented vertically (as in the field), then at 45° degrees about the horizontal axis parallel to the hexagonal board, and finally horizontally (microphones facing upwards in the latter two orientations).

2.3 Analyses

To first determine whether the 'virtual' replicated soundscapes could be a suitable substitute to field recordings for testing PAM devices, we looked at differences in typical analyses of the 6-mic device's field and vertically-oriented lab recordings. After examining spectral differences (spectrograms computed with 'pspectrum' in *MATLAB*), we used the *seewave* and *soundecology* packages in *R* to extract the following 7 widely-adopted Acoustic Indices [9–11] on 30 s windows [1]: Acoustic Complexity Index (ACI), Acoustic Diversity Index (ADI), Acoustic Evenness (AEve), Bioacoustic Index (Bio), Acoustic Entropy (H), Median of the Acoustic Envelope (M) and Normalised Difference Soundscape Index (NDSI). We also extracted the 128 feature embedding of the pretrained VGGish Convolutional Neural Network (CNN) [12], which has proven highly effective in various ecoacoustic analyses and classification tasks [11,13]. This was obtained using the 'vggishPreprocess' and 'predict' functions from *MATLAB*'s *Deep Learning Toolbox*.

Following Heath et al. [11], we used a modified version of Bland-Altman analysis [14] to look at the scaled like-for-like differences in Acoustic Indices and VGGish features between the field and vertical lab recordings as a percentage of the range of the corresponding index/feature for the field recording. As in past work, we set a threshold of ±5% difference between recordings within which the data are not considered to have been altered [11]. Since some field recordings' VGGish features had zero range, we set those features' differences to zero to avoid dividing by zero. We then averaged the 128 VGGish features' differences so as to convert them to a single-dimension difference set akin to the 7 other Acoustic Indices (Fig. 1).

To explore the impact of device orientation, we extended the above analyses to the lab-based re-recordings with the 6-mic device oriented at 45° and horizontally, revealing the differences in indices' values as device orientation changed (Fig. 1). We also ran all recordings through BirdNET [5] and HARKBird [6], to classify and localise avian calls, respectively. We ran the former on monophonic data from microphone 1 (in the upper left vertex), while the latter examines differences in time-of-arrival in the 6-channel data to determine the direction of bird calls. For each site, we then found the percentage of classifications and localisations in the field recordings that appeared in the lab re-recordings.

3. RESULTS AND DISCUSSION

Results from the spectral differences, VGGish, and BirdNET, indicate our test platform can reproduce soundscapes for the purpose of simulating certain real-world analyses.

---

[1] 60 s or more is a more common window size [10], but would lead to too few data points given the brevity of our recordings.

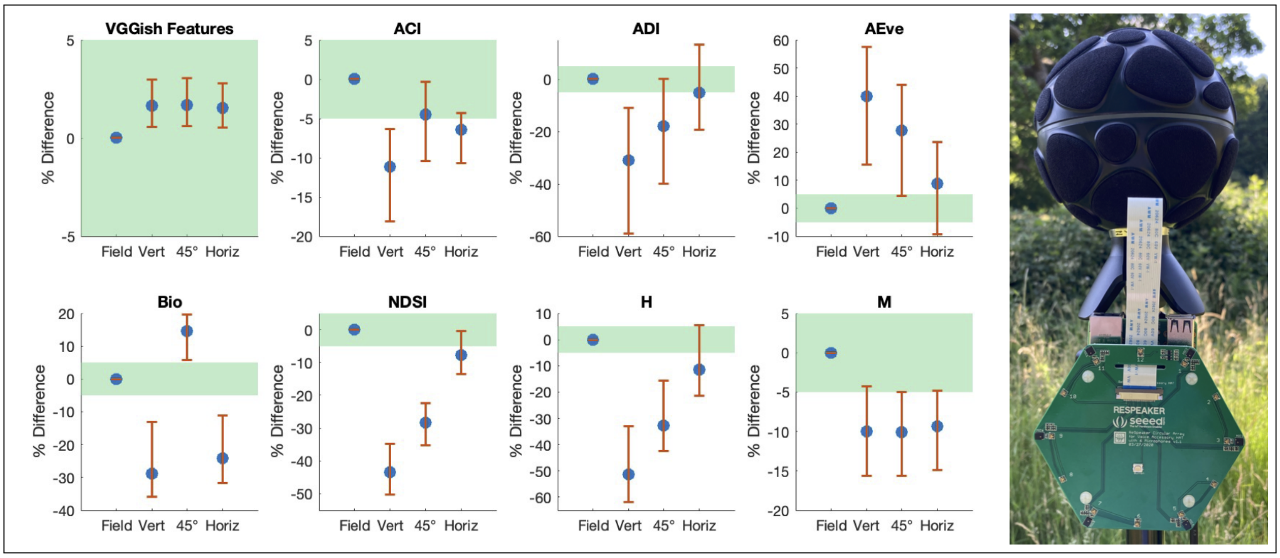

Figure 1. **Left:** Plots of indices' scaled like-for-like differences averaged across all sites for each recording condition (the reference vertical Field recordings, and lab re-recordings under Vertical, 45° and Horizontal orientation). Blue dots indicate the means of median percentage differences per site for each recording condition, whiskers show the mean upper and lower quartiles, and the shaded area denotes ±5%. Indices' abbreviations are defined as follows: ACI - Acoustic Complexity Index, ADI - Acoustic Diversity Index, AEve - Acoustic Evenness, Bio - Bioacoustic Index, NDSI - Normalised Difference Soundscape Index, H - Acoustic Entropy, M - Median of the Acoustic Envelope. **Right:** Field recording setup with the *Zylia* ZM-1 above our 6-mic device.

Spectral differences between field and lab vertical recordings were minimal (excluding an overall level increase due to the non-ideal room housing the system and limitations with the calibration process). The VGGish features also showed little difference across all recordings, while the Acoustic Indices varied considerably: interestingly, absolute percentage differences never increased as re-recording orientation progressively changed, but rather decreased or showed no trend. Fig. 1 shows the like-for-like percentage differences averaged for all sites. The variability in the Acoustic Indices' differences could be due to the limited audio data or window sizes used. Equally, this variability could reflect limitations of the Acoustic Indices themselves: the VGGish features' robustness to altered recording conditions and Acoustic Indices' lack thereof mirrors results from Heath et al.'s study of the effect of compression [11] and Sethi et al.'s findings that VGGish features outperform Acoustic Indices at various classification tasks [13].

Device orientation impacted the performance of BirdNET and HARKBird. For BirdNET, lab re-recordings had average overlaps to the field recordings of 60.5% (vertical), 49.4% (45°), and 68.7% (horizontal). Allowing ±45° in localisation error, the average overlaps for HARKBird were 4.1% (vertical), 7.2% (45°) and 7.7% (horizontal). Whilst it is surprising that a larger proportion of localisations from HARKBird were accurate in the non-vertical re-recordings (whose orientations differ from the field recordings), the overall low percentage overlaps suggest sound source direction is generally inaccurately reproduced in our Ambisonics playback system. Indeed, on average 39.2% of HARKBird field localisations appeared in the vertical re-recordings, but most had more than a 45° error. This is likely because Ambisonics playback systems are more tailored to preserving the human perception of a sound source's direction rather than its precise location. Moreover, the increased similarity of the horizontal recordings' results to their field counterparts may be due to the larger number of loudspeakers in the horizontal plane. This is also typical of Ambisonics playback systems because of humans' particular sensitivity to localising sound sources horizontally. Future work might therefore require exploring other spatial audio techniques for rendering of the virtual scene, for example those based on soundfield reconstruction rather than human perception.

4. CONCLUSION

Overall, our results indicate that our test platform can be used to replicate spatial natural soundscapes for the purpose of testing spatial PAM devices. Additional work is required to further validate this platform and in particular understand: (i) the variability in the Acoustic Indices' differences; (ii) the low proportion of accurate HARK-Bird localisations in the re-recordings; and (iii) whether an alternative method to Ambisonics would provide more spatially-accurate soundscape reproduction. However, the trial run of the system nonetheless yielded core insights into how device orientation can impact certain spatial ecoacoustic analyses. It is now hoped the platform's accessible hardware and software (including several free and/or open source components) may encourage additional researchers to develop and validate platforms like it, towards a more standardised, controlled approach for evaluating spatial ecoacoustic technologies.

5. ACKNOWLEDGMENTS

Thanks to Dr David Orme for advice on editing and the analyses undertaken. This work was supported by the Natural Environment Research Council [NE/S007415/1] (to N.P.L.P via the SSCP DTP).

6. REFERENCES

[1] M. Wijers, A. Loveridge, D. W. Macdonald, and A. Markham, "CARACAL: a versatile passive acoustic monitoring tool for wildlife research and conservation," *Bioacoustics*, vol. 30, pp. 41–57, Nov. 2019.

[2] D. Payo, L. González, P. Kogan, and M. Eguia, "Sentinel: Versatile real-time acoustic autonomous monitoring system for studying natural soundscapes," in *Euronoise 2021*, (Madeira, Portugal), Oct. 2021.

[3] D. T. Blumstein *et al.*, "Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus," *Journal of Applied Ecology*, vol. 48, no. 3, pp. 758–767, 2011.

[4] S. Kaneko and H. Gamper, "Large-scale simulation of bird localization systems in forests with distributed microphone arrays," *JASA Express Letters*, vol. 2, pp. 101201–1 – 101201–8, Oct. 2022.

[5] S. Kahl, C. M. Wood, M. Eibl, and H. Klinck, "BirdNET: A deep learning solution for avian diversity monitoring," *Ecological Informatics*, vol. 61, p. 101236, Mar. 2021.

[6] S. Sumitani *et al.*, "An Integrated Framework for Field Recording, Localization, Classification and Annotation of Birdsongs Using Robot Audition Techniques — Harkbird 2.0," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (Brighton, Unilted Kingdom), pp. 8246–8250, May 2019.

[7] S. S. Sethi, R. M. Ewers, N. S. Jones, C. D. L. Orme, and L. Picinali, "Robust, real-time and autonomous monitoring of ecosystems with an open, low-cost, networked device," *MEE*, vol. 9, pp. 2383–2387, Dec. 2018.

[8] Institute for Computer Music and Sound Technology, "Ambisonics Externals for MaxMSP," Nov. 2021.

[9] R. B. Machado, L. Aguiar, and G. Jones, "Do acoustic indices reflect the characteristics of bird communities in the savannas of Central Brazil?," *Landscape and Urban Planning*, vol. 162, pp. 36–43, June 2017.

[10] T. Bradfer-Lawrence, N. Gardner, L. Bunnefeld, N. Bunnefeld, S. G. Willis, and D. H. Dent, "Guidelines for the use of acoustic indices in environmental research," *MEE*, vol. 10, no. 10, pp. 1796–1807, 2019.

[11] B. E. Heath, S. S. Sethi, C. D. L. Orme, R. M. Ewers, and L. Picinali, "How index selection, compression, and recording schedule impact the description of ecological soundscapes," *Ecology and Evolution*, vol. 11, no. 19, pp. 13206–13217, 2021.

[12] S. Hershey *et al.*, "CNN architectures for large-scale audio classification," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, (New Orleans, LA, USA), pp. 131–135, Mar. 2017.

[13] S. S. Sethi *et al.*, "Characterizing soundscapes across diverse ecosystems using a universal acoustic feature set," *Proc Natl Acad Sci USA*, vol. 117, pp. 17049–17055, July 2020.

[14] D. Giavarina, "Understanding Bland Altman analysis," *Biochem Med*, vol. 25, pp. 141–151, June 2015.