SECTOR-BASED FILTERING AND CLEAN-SC BEAMFORMING FOR NATURAL SOUNDSCAPE ANALYSIS : THE CASE OF THE LITTLE OWL

M. Biscarat[1], P. Lecomte[1], J. Rouch[2], D. Dragna[1], J. Sueur[3]
[1] *Ecole Centrale de Lyon, CNRS, Universite Claude Bernard Lyon 1, INSA Lyon, LMFA, UMR5509, 69130, Ecully, France*
[2] *CNRS, Inserm, Université Jean Monnet, Centre de Recherche en Neurosciences de Lyon, ENES, UMR5292, 42023, Saint-Étienne, France*
[3] *CNRS, Muséum national d'Histoire naturelle, Sorbonne Université, CESCO, UMR7204, 75005, Paris, France*

Abstract

Acoustic monitoring has become a widely used approach to study ecosystems. However, the large amount of acquired data requires fast processing techniques to efficiently analyze soundscapes. For that, sound detection and classification algorithms, such as BirdNET, have been developed. However, the performance of BirdNET, quantified by the confidence score, is closely linked to the Signal-to-Noise Ratio (SNR) of the audio recordings. Besides, the use of Spherical Microphone Arrays (SMA), which enable rotationally invariant Spherical Harmonics (SH) beamforming, allows for directional filtering of the soundscape and sound source localization in any direction. In this work, natural soundscape analysis is performed using a SMA to capture the soundscape. In particular, the night call of the Little Owl (*Athene noctua*) was automatically searched for. Firstly, sector-based filtering is carried out to provide a set of spatially filtered signals. These signals are then analyzed using BirdNET to provide a list of sound events. Secondly, for each sound event, the CLEAN-SC deconvolution beamforming algorithm is applied to the whole 3D sound scene captured by the SMA. Directional signals are finally extracted in the direction of the sound sources and analyzed through BirdNET. An improvement of over 200% in the number of Little Owl's detections in a natural environment has been achieved by means of sector-based filtering in comparison to a monophonic signal. After extracting directional signals, the confidence score of BirdNET has been raised by more than 40%. These results suggest that the proposed method improves the robustness of BirdNET.

Keywords: *spherical harmonics, beamforming, directional filtering, ecoacoustics, birdnet*

1. Introduction

Acoustic monitoring of natural soundscapes presents numerous advantages: it is an inexpensive and non-invasive method to assess the state of biodiversity, or to track changes in an ecosystem over long temporal scales [1], [2]. This acoustic monitoring is usually performed with Autonomous Recording Units (ARU), deployed in the environment to record monophonic omnidirectional signals of the soundscape. However, long-term measurements result in substantial amounts of data, making fast processing essential for efficient analysis. To address this issue, detection and classification tools have been developed. Among them, BirdNET[1] is a neural network trained to identify more than 6000 species, most of them birds [3]. Its identification accuracy is given by a confidence score, between 0 and 1 [4]. Its ability to detect and classify species is related to the Signal-to-Noise Ratio (SNR) and BirdNET becomes unefficient in noisy environments [5]. Natural soundscapes often present multiple sound sources. In monophonic recordings, sound sources of interest can be completly masked by other sound sources, such as geophony (wind, rain), anthropophony (cars, planes) or other species.

To overcome this limitation, we investigated the captation of the soundscape with a Spherical Microphone Array (SMA) paired with beamforming techniques to perform spatial filtering and to unmask sound sources

[1] `https://birdnet.cornell.edu/?lang=fr`

Corresponding author: *maud.biscarat@ec-lyon.fr.*

of the soundscape. Spatial filtering improves the SNR of the recordings, and should thus enhance BirdNET's performance in the number of detections and their confidence score. Sound source localization with beamforming is conventionally performed by steering a beamformer across a dense grid of directions, providing a power map. The maxima of the power map correspond to the Direction of Arrival (DOA) of the sources. However, as shown in the Fig. 5(a), in natural soundscapes, sound source localization can be compromised. Indeed, the dominant sound source might mask the other ones, especially if they are at a lower level. Numerous sound sources can lead to a diffuse sound field with no clearly identifiable DOA. Another issue with generating heatmaps to localize sound sources is that steering beamformers in a dense grid of directions results in a significant computational burden when a large volume of data is collected. We therefore propose to steer in a limited and minimal number of directions with a distribution of points on the sphere that neither favors nor discriminates no direction. This criteria is met with the condition of conservation of energy. The nodes of a spherical t-design [6] satisfy this condition [7] and are adopted here. To adress the issue of the secondary sound sources masking, CLEAN-SC deconvolution algorithm [8], [9] is performed to generate a cleaned power map (see Fig. 5(b)) and localize sources. This algorithm is robust to noise, as it iteratively cleans the noisy map, unmasking secondary sources.

The paper is organized as follows: in Sec. 2, the recording and post-treatment process of the proposed method are detailed. In Sec. 3, the method is tested on a field recording of two Little Owls made in the Parc Naturel Régional du Pilat in France, during Little Owl's census. Finally, in Sec. 4, a discussion of the obtained results is given.

2. Method

A block diagram of the proposed method is shown Fig. 1. The method is composed of five steps.

1. The soundscape is decomposed onto the basis of Spherical Harmonics (SH).
2. The sphere is partitioned into non-overlapping beampatterns following a t-design, and sector-based signals are obtained by steering beampatterns toward the center of each sector.
3. Sector-based signals are analyzed through BirdNET to detect sound events, and the timestamps of each Little Owl vocalization are collected.
4. For each timestamp, the CLEAN-SC algorithm is applied to the sound scene described by the SH components, allowing the localization of multiple sound sources.
5. Directional signals are extracted in the direction of each identified sound source and analyzed through a second BirdNET processing stage.

Those steps are discussed in the following subsections.

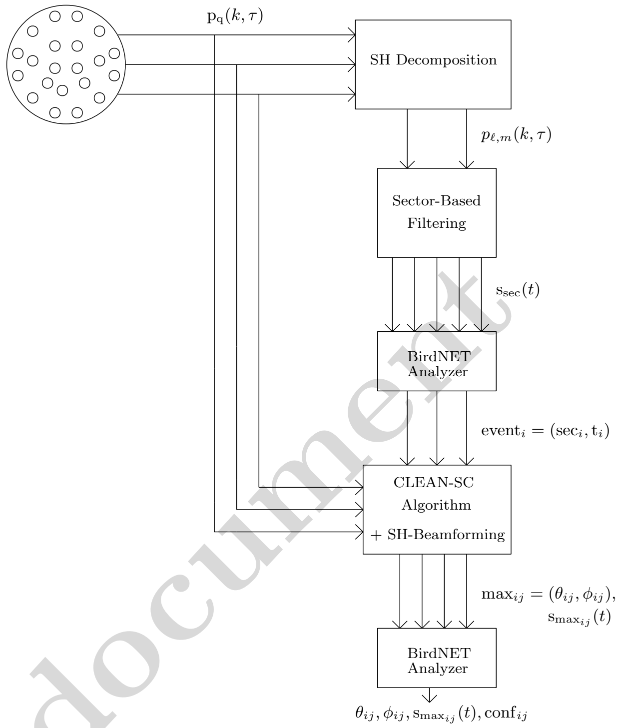

Figure 1: Block diagram of the proposed method, with $p_q(k,\tau)$ the microphone signals, $p_{\ell,m}(k,\tau)$ their SH decomposition, $s_{\text{sec}}(t)$ the sector-based signals, sec_i and t_i the sector and timestamp of detection for each detection instant i, (θ_{ij}, ϕ_{ij}) the location of each maximum j for a detection instant i, with the directional signal associated $s_{\max_{ij}}(t)$, and its final confidence score conf_{ij}

2.1. Capturing the sound scene

In this study, all recordings were carried out with the em64 Eigenmike®, a SMA composed of 64 microphones, allowing a decomposition on SH up to order $L = 6$. The field recordings were made on the 17$^{\text{th}}$ of May 2026, between 7 PM and 7:30 PM at the position N 45°23′24.5”, E 004°42′25.9”.

2.2. SH-Beamforming

SH-Beamforming allows for the localization of sound sources and the extraction of directional signals [10]. It is based on the decomposition of the soundfield captured by Q microphones on the Spherical Harmonics (SHs), which form a basis of $L^2(\mathbb{S})$, the space of square-integrable functions defined on the unit sphere. The sound field is expressed in the time-frequency domain following short-time Fourier transform processing. The SH decomposition is performed with the discrete spherical Fourier transform:

$$p_{\ell,m}(k,\tau) \approx \sum_{q=1}^{Q} c_q \, p_q(k,\tau) \, Y_{\ell,m}(\theta_q, \phi_q) \qquad (1)$$

with ℓ the order and m the degree of the decomposition on the SHs, c_q the integration weights which depend on the position of the microphones on the sphere, $p_q(k,\tau)$ the sound pressure for a time-frequency tile, with $k = 2\pi f/c$ the wavenumber, f the frequency and c the sound speed. The sound pressure is captured at a microphone located at (θ_q, ϕ_q), θ_q the colatitude and ϕ_q the azimuth. k is the wavenumber and $Y_{\ell,m}$ the vector composed of the SHs evaluated in the direction of the microphone up to a maximum order L, linked to the number of microphones of the SMA.

The output of the beamformer for a steering angle (θ_s, ϕ_s) is given by

$$y(k,\tau,\theta_s,\phi_s) = \sum_{\ell=0}^{L} \sum_{m=-\ell}^{\ell} B_\ell(k,a) Y_{\ell,m}(\theta_s,\phi_s)\, p_{\ell,m}(k,\tau) \tag{2}$$

with $B_\ell(k,a)$ the radial filters compensating for the scattering of the rigid array of radius a [11]:

$$B_\ell(k,a) = \mathrm{i}^{-(\ell-1)}(ka)^2 h_\ell^{(2)'}(ka) \tag{3}$$

where $h_\ell^{(2)'}(ka)$ denotes the derivative of the spherical Hankel function of the second kind.
$B_\ell(k,a)Y_{\ell,m}(\theta_s,\phi_s)$ is called the steering vector and corresponds to the SHs evaluated at a steering angle. The output of the beamformer is a directional signal, allowing for a listening with an enhanced directivity. For a decomposition order L, the SNR gain in the listening direction is described in dB by the Directivity Factor (DF) defined as $DF = 10\log((L+1)^2)$ [12]. In our case, the DF is $DF = 10\log((7+1)^2) \approx 17$ dB. The directivity pattern of the beampattern is shown in Fig. 2.

Figure 2: Beampattern at the order $L = 6$

The Sound Pressure Level (SPL) of the beamformer output for each direction can also be computed to generate a SPL heatmap and localize sources (see Fig. 5(a)).

2.3. Sector-Based Filtering

The chosen spherical sampling scheme should neither favors nor discriminates against any direction, which relies on the conservation of energy for the sum of all beampatterns. This can be expressed as [7]:

$$\sum_{n=1}^{N} C w_n^2(\theta,\phi) = 1 \tag{4}$$

with N the number of sectors of the sampling scheme, $w_n^2(\theta,\phi)$ a beampattern steering at the center of the sector and C a normalization factor. Some sampling schemes satisfying this condition for axis-symmetric beampatterns are the t-designs [6]. A t-design allows for a maximum order of decomposition L for each sector with $t = 2L$. For this study, a 84-node t-design is chosen, enabling sector base beamforming up to order L=6. In [7], the minimal t-design is chosen, with the minimal numbers of sectors for computational efficiency. Here, the minimal number of sectors is selected for a 6^{th} order SHs decomposition, the maximal order allowed by the SMA chosen. The 84 sector-based signals are then computed and analyzed using BirdNET.

2.4. Detection of the Little Owl's vocalizations through BirdNET

Each directional signal is processed through BirdNET by segments of 3 seconds. The output of this analysis is a list of species with a confidence score between 0 and 1. These confidence scores are related to the prediction accuracy of the neural network but are not probabilities and should not be considered as such [4]. These confidence scores are not replicable inter-species and highly depend of the recording's features (sample rate, meteorological conditions, SNR...). As the prediction accuracy increases with the confidence score, we choose a medium confidence score of 0.6 as a threshold of detection for the Little Owl. In the reported case, 84 signals are processed this way giving a sound event timestamp log for each sector.

2.5. CLEAN-SC deconvolution algorithm for sound source localization

For every 3-seconds segment marked, CLEAN-SC algorithm is processed to localize precisely the different sound sources. CLEAN-SC is an iterative beamforming algorithm based on source coherence. It was initially developed for source localization in aeroacoustics with planar arrays [8] and then adapted to SMA [9]. For aeroacoustics application in a wind-tunnel environment, the sound sources are in the near-field, and the SNR is relatively high. In the case of natural sound sources, they are located in the far-field with lower SNR. Thus, the parameters of the algorithm have to be modified to analyze natural soundscapes. At each iteration, a fraction ϑ of the contribution of the principal sound source is removed from the so-called "dirty" map which is the power map of the regular beamforming 5(a), cleaning it. The DOA of the principal sound source is added to the clean map. After I iterations, the local maxima can be observed on the clean map (see Fig. 5(b)). By choosing a low ϑ value, the algorithm is able to better localize sources in cases where there are multiple sources at different levels. However, more iterations are required to clean up the "dirty" map. As in natural environments, the soundscape is complex, $\vartheta = 0.1$ and $I = 300$ are chosen.

2.6. Localization of the maxima and 2nd BirdNET analysis

Once the clean map for the segment is computed, the coordinates (θ_j, ϕ_j), $j = 1, 2, \ldots, J$ of the J maximums corresponding to the J sources of the soundscape are acquired. The output of the beamformer for each maximum (θ_j, ϕ_j) is computed (see Eq. (2)) up to order $L = 6$. The output of the different maxima are analyzed through BirdNET again, enhancing the confidence score. The outputs with a high confidence score correspond to the location of the Little Owl.

3. Results

3.1. Full recording

To test the proposed method, a recording in a natural environment of two Little Owls is processed. The soundscape was captured in the Parc Naturel Régional du Pilat, France, in an open environment, composed of fields and surrounded with trees, see Fig. 3.

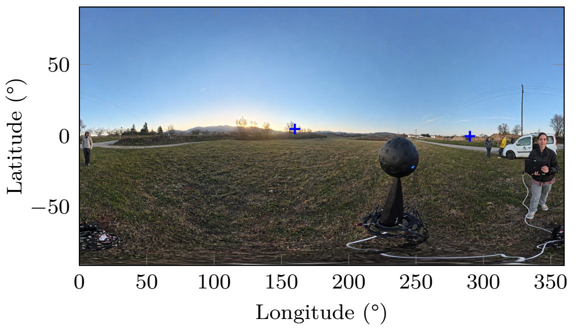

Figure 3: Panoramic view of the recording environment. The apparent size, shape, and distance of objects may be distorted. The crosses show the localization of Little Owls

Acoustic playback of a Little Owl was used to elicit a response of Little Owls in the environment. A loudspeaker played Little Owl vocalizations for 90 seconds, followed by a one-minute pause, before broadcasting the vocalizations again for another 90 seconds. As a territorial bird, a Little Owl is likely to respond to the call of another Little Owl. The playback sound was shut down immediately if a Little Owl started vocalizing to minimize disturbance due to playback, or if a predatory species responded.

After the playback, two apparent Little Owls were observed vocalizing. One was located closer to the SMA, in the trees located at $(\lambda_{o1}, \phi_{o1}) = (4°, 160°)$, and the further one was at $(\lambda_{o2}, \phi_{o2}) = (-5°, 290°)$ where $\lambda = 90 - \theta$ is the latitude. After 60 seconds of recording, the owl at the first location flew to join the second one. The spectrogram of the monophonic signal recorded as the sum of all microphone signals is shown Fig. 4.

The omnidirectional recording signal is analyzed through Birdnet as well as the 84 sector-based signals. The number of detections (confidence score > 0.6) for 3-second segments in each case is indicated in Table 1, and compared to that obtained by manual annotation.

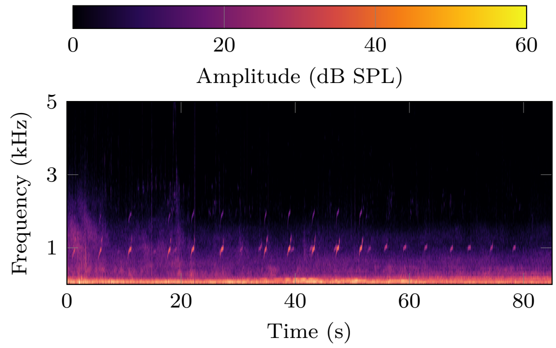

Figure 4: Spectrogram of the omnidirectional recording (sum of the SMA signals)

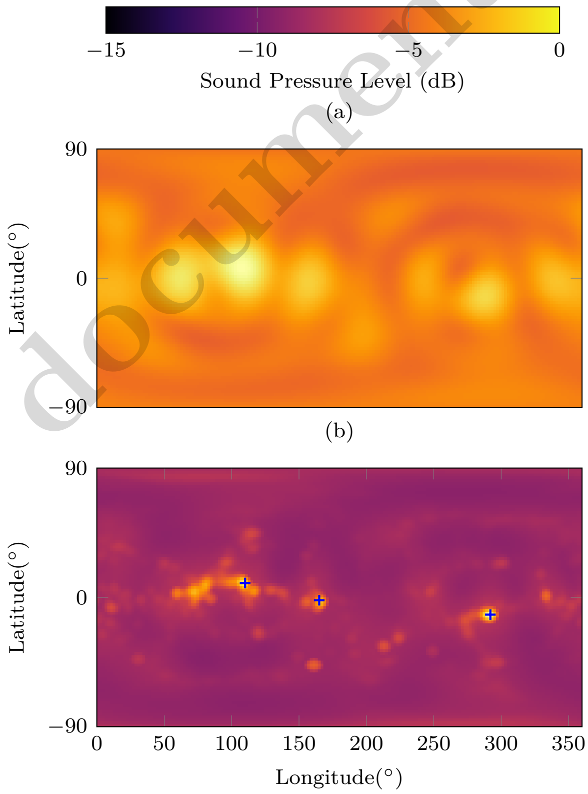

Figure 5: Heatmap of the soundscape for the 15 – 18-s segment with (a) The dirty map before CLEAN-SC and (b) The clean map after CLEAN-SC, with local maximums indicated by blue crosses

Table 1: Number of Little Owl's detections for omnidirectional recording (OR), sector-based signals (SB) and manual annotation (MA)

OR	SB	MA
7	23	24

For each 3-seconds segment where the Little Owl is detected in the Sector-Based signals, CLEAN-SC is performed, with an overlapping of 0.5 second for each segment, to improve the localization of Little Owls at the beginning and end of the recordings. The $15-18$ seconds segment is studied.

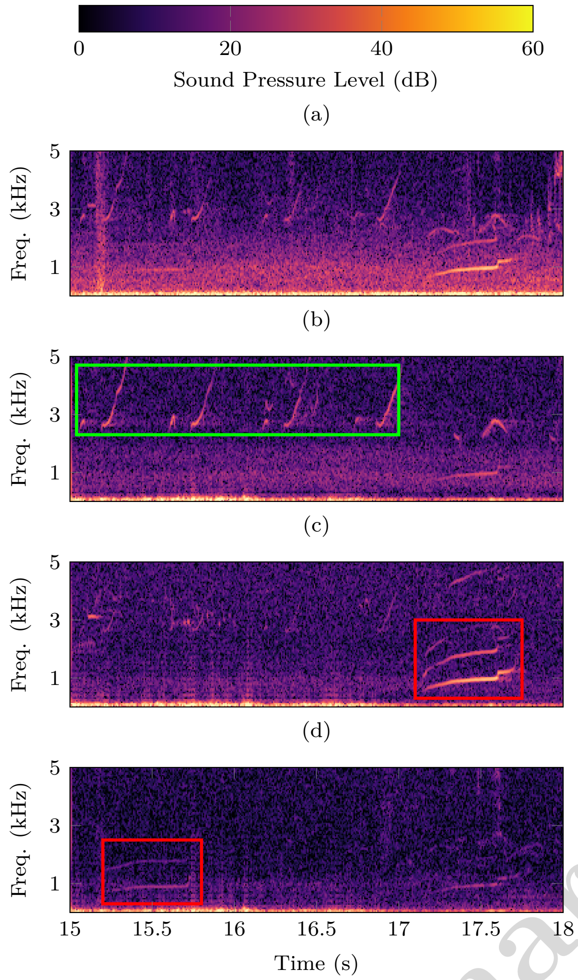

Figure 6: Spectrograms of the 15-18 seconds segment of the (a) omnidirectional recording, and the directional signals at (b) $(10°, 110°)$, (c) $(-2°, 165°)$, (d) $(-12°, 292°)$, with a bird contribution (probably a Song Thrush, *Turdus philomelos*) enclosed in a green box in (b), with the nearest Little Owl's contribution enclosed in a red box in (c) and the furthest Little Owl's contribution enclosed in a red box in (d)

3.2. 15-18 seconds segment analysis

The output map of the CLEAN-SC algorithm is shown Fig. 5. Several local maxima can be observed on the output map, higlighting the complexity of the studied soundscape. The three main maximums, associated with the highest SPL, are indicated by the blue crosses. Directional signals in direction of the local maximums are extracted. Their spectrogramms are shown Fig. 6. These directional signals isolate the different sound sources of the soundscape during this segment. On Fig. 6(b), vocalizations of a bird are enhanced with frequency contents between 3 and 5 kHz, whereas on Fig. 6(c),(d), the Little Owl's contributions are enhanced, meaning the direction of extraction of the signals coincides with the owls locations. On this time segment, the confidence score of the omnidirectional signal is 0.107 for the Little Owl's detection. After extracting the directional signal of the second maximum, the confidence score is 0.9565.

3.3. BirdNET 2$^{\text{nd}}$ analysis

After analyzing directional signals for all maxima through BirdNET for each time segment, the mean confidence score of the detections (confidence score above 0.6) is 0.963, whereas for omnidirectional recording, the mean confidence score was 0.656. The $18-21$ s segment and the $39-42$ s segment were not included in this part of the analysis, as the Little Owl was at the same angular location as an other sound source or mainly associated with the previous segment.

4. Discussion and Conclusion

In this study, we proposed a method to enhance BirdNET's performance in natural soundscapes, based on the use of a SMA, paired with Sector-Based filtering and source localization with CLEAN-SC. We focused on the case of a recording of the Little Owl, and investigated the detectability and the ability of BirdNET to classify the vocalizations with confidence. By summing the signals recorded by the SMA, we reconstructed an omnidirectional recording that could have been recorded with a monophonic device as an ARU. We analyzed this omnidirectional recording as well as sector-based recordings through BirdNET. The results showed that 7 3-seconds time segments are detected with the Little Owl's vocalizations for a confidence score over 0.6 for the omnidirectional recording, and 23 segments for the Sector-Based signals. Consequently, sector-based analysis detected 220% more vocalizations than could be detected with a monophonic recorder. At each time-segment with Little Owl's detection, CLEAN-SC was performed and directional signals were extracted in the direction of the maxima. On the 15-18 seconds time segment, the confidence score of the omnidirectional signal was 0.107. After extracting directional signals in direction of 3 local maxima, results showed the ability of beamforming to isolate the different sound sources of the soundscape. Thus, the presence of a bird and two Little Owls was highlighted. By analyzing the directional signals for the three main maximums through BirdNET, the confidence score of Little Owl's detection was raised to 0.956. Not only did the proposed method enable the detection of sound events that could not be detected using monophonic recordings, but it also greatly increased BirdNET's confidence scores, thereby improving the reliability of the detections. Finally, for all time-segments with Sector-Based detections, the mean confidence score after the analysis of omnidirectional signal is 0.656 and for maximum directional signals 0.947, so a raise of over 45%.

The segments containing both owls do not allow for strict filtering, due to the beam geometry and its side lobes. Thus, even when directional signals are

extracted in directions other than those of the nearest owl, the owl's contribution remains visible on the spectrograms and detectable by BirdNET. Consequently, two individuals vocalizing at the same time cannot currently be correctly distinguished only by post-processing with BirdNET.
Some segments with the end of the acoustic signature of the Little Owl, as well as when the Little Owl is closer to another sound source failed to be classified efficiently by BirdNET. However, these cases are specific and do not represent the majority of the vocalizations. Moreover, the detection confidence threshold chosen should be investigated and tuned as a species-specific threshold for the Little Owl, for instance using a precision–recall analysis to achieve an appropriate trade-off between sensitivity and false detections.
As the proposed method was only tested in one recording environment, additional tests should be conducted to confirm this efficiency in other Little Owl recordings and also for other nocturnal and diurnal species evolving in different soundscapes.
To conclude, the approach introduced in this paper showed a raise of over 220% in the number of Little Owl's detection and an improvement of more than 40% of the confidence score, improving BirdNET's detectability and robustness in the case of the Little Owl.

5. Acknowledgments

The authors would like to thank the Parc Naturel Régional du Pilat for their collaboration and for providing access to the park for fieldwork. The authors would like to thank Marien Eydant for organizing the Little Owl monitoring field sessions.

References

[1] S. R.-J. Ross et al., "Passive acoustic monitoring provides a fresh perspective on fundamental ecological questions," *Functional Ecology*, vol. 37, no. 4, pp. 959–975, 2023.

[2] M.-L. Ng, N. Butler, and N. Woods, "Soundscapes as a surrogate measure of vegetation condition for biodiversity values: A pilot study," *Ecological Indicators*, vol. 93, pp. 1070–1080, 2018.

[3] S. Kahl, C. M. Wood, M. Eibl, and H. Klinck, "Birdnet: A deep learning solution for avian diversity monitoring," *Ecological Informatics*, vol. 61, p. 101 236, 2021.

[4] C. M. Wood and S. Kahl, "Guidelines for appropriate use of birdnet scores and other detector outputs," *Journal of Ornithology*, vol. 165, no. 3, pp. 777–782, 2024.

[5] F. Michaud, J. Sueur, F. Sèbe, M. Le Cesne, and S. Haupert, "Acoustic detection of a nocturnal bird with deep learning: The challenge of low signal-to-noise ratio," *Ecological Indicators*, vol. 181, p. 114 475, 2025.

[6] R. H. Hardin and N. J. Sloane, "Mclaren's improved snub cube and other new spherical designs in three dimensions," *Discrete & Computational Geometry*, vol. 15, no. 4, pp. 429–441, 1996.

[7] A. Politis, J. Vilkamo, and V. Pulkki, "Sector-based parametric sound field reproduction in the spherical harmonic domain," *IEEE Journal of Selected Topics in Signal Processing*, vol. 9, no. 5, pp. 852–866, 2015.

[8] P. Sijtsma, "Clean based on spatial source coherence," *International journal of aeroacoustics*, vol. 6, no. 4, pp. 357–374, 2007.

[9] Z. Chu, S. Zhao, Y. Yang, and Y. Yang, "Deconvolution using clean-sc for acoustic source identification with spherical microphone arrays," *Journal of Sound and Vibration*, vol. 440, pp. 161–173, 2019.

[10] B. Rafaely, *Fundamentals of spherical array processing*. Springer, 2015, vol. 8.

[11] S. Moreau, J. Daniel, and S. Bertet, "3d sound field recording with higher order ambisonics–objective measurements and validation of a 4th order spherical microphone," in *120th Convention of the AES*, 2006, pp. 20–23.

[12] J. Meyer and G. Elko, "A highly scalable spherical microphone array based on an orthonormal decomposition of the soundfield," in *2002 IEEE International Conference on Acoustics, Speech, and Signal Processing*, IEEE, vol. 2, 2002, pp. II–1781.