DOI: 10.61782/fa.2025.0339

FORUM ACUSTICUM
EURONOISE 2025

ENHANCED BIRDCALL DETECTION USING MULTIDIRECTIONAL BEAMFORMING AND AUTOMATED SOURCE SELECTION IN LOW SNR SOUNDSCAPES

Rifqi Ikhwanuddin[1,4*] **Emilie D'Olne**[2] **Sarab S Sethi**[3] **Lorenzo Picinali**[1]
[1] Dyson School of Design Engineering, Imperial College London, United Kingdom
[2] Electrical and Electronic Engineering, Imperial College London, United Kingdom
[3] Department of Life Sciences, Imperial College London, United Kingdom
[4] Engineering Physics, Institut Teknologi Sumatera, Indonesia

ABSTRACT

In forest ecosystems, detecting birdcalls is hindered by high environmental noise and multiple sound sources, complicating species identification. Beamforming techniques can enhance signal-to-noise ratio (SNR) by focusing on specific directions but face challenges with numerous, unpredictably located sources. This study proposes a novel methodology integrating multidirectional beamforming with automated source selection to address these issues. Our approach sequentially scans multiple directions and employs BirdNET (a machine learning-based platform for bird recognition) confidence scores to identify optimal detection sectors, improving accuracy. Using a Multichannel Acoustic Autonomous Recording Unit (MAARU), experiments were conducted in a controlled virtual sound environment with simulated Eurasian Blue Tit calls under varying SNR conditions (+30dB to -30dB). Results demonstrate a significant enhancement in species detection accuracy compared to single-channel recordings, especially at low signal-to-noise ratio, with beamforming achieving up to 588% improvement in detection counts at -20 dB ($p < 0.0001$) and consistently higher confidence scores ($p < 0.05$). The findings highlight the potential of combining multidirectional beamforming with AI-based detection for biodiversity monitoring in challenging acoustic environments. Future work will extend these methods to field deployments to validate their effectiveness in real-world conditions.

**Corresponding author*: *r.ikhwanuddin23@imperial.ac.uk.*

Keywords: *birdcall, passive acoustic monitoring, multidirectional beamforming, & ecoacoustics.*

1. INTRODUCTION

Passive Acoustic Monitoring (PAM) offers promising capabilities for biodiversity assessment through IoT sensors and AI analysis [1]. However, practical implementation faces challenges from overlapping sound sources and environmental noise [2, 3].

While autonomous recording units are increasingly used for monitoring bird occurrences in remote areas, multi-channel sensor devices remain uncommon due to data size constraints and their relative novelty in ecological applications [4]. Traditional single-channel systems struggle in low signal-to-noise ratio (SNR) environments, where target bird vocalisations are often obscured by background noise, making species identification difficult without spatial information [4, 5].

Multi-channel recording systems with beamforming techniques present an effective solution by enabling focused listening in specific directions. These techniques can be implemented through either dynamic steering, which continuously adjusts to sound source movements, or static steering, which maintains fixed directional focus points [6].

This research integrates multidirectional static steering beamforming with BirdNET, an AI-based detection system, aiming to enhance biodiversity monitoring accuracy in challenging acoustic environments with overlapping sound sources and significant background noise.

2. MATERIALS AND METHODS

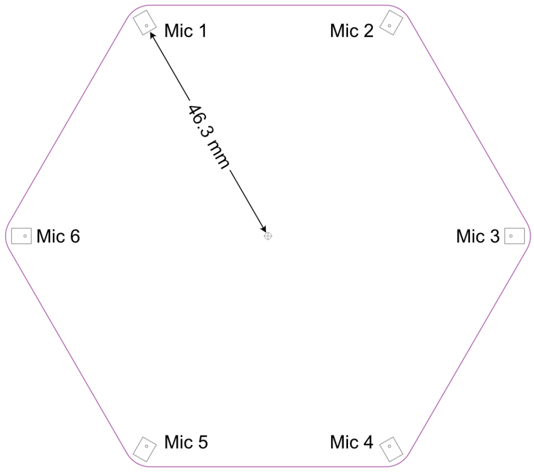

Figure 1. Circular microphone array configuration with six microphones (Mic 1-6) positioned at the vertices. Each microphone is positioned at an equal distance of 46.3 mm from the centre point of the array.

We used the Multichannel Acoustic Autonomous Recording Unit (MAARU) [4,7], which integrates a Raspberry Pi 3 Model B+ with a ReSpeaker 6-microphone circular array containing omnidirectional MEMS microphones (see Fig. 1) [8]. Complete instructions are available on GitHub [9]. The MAARU sampled audio at 16kHz during all experiments [10, 11].

Testing employed a Virtual Sound Environment (VSE) with third-order Ambisonics recordings reproduced through a 31-loudspeaker system [12]. Our experimental scenario combined an 8-minute background soundscape with Eurasian Blue Tit (EBT) calls (*Cyanistes caeruleus*) reproduced sequentially from 31 different positions at progressively decreasing sound pressure levels (70dB to 10dB in -10dB steps). The EBT call was selected for its frequency range (3-8kHz) which aligns with common bird vocalisations (see Fig. 2) [10].

For beamforming implementation, we measured 3D impulse responses using time-stretched pulse signals generated from 31 spatial positions at a fixed distance of 1.5 meters. These measurements incorporated azimuth angles

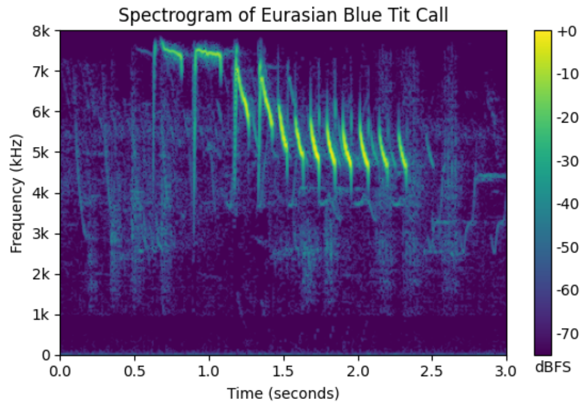

Figure 2. Spectrogram of a Eurasian Blue Tit (Cyanistes caeruleus, XC28208) call showing its distinctive frequency pattern. The characteristic high-frequency modulations (3-8kHz) are visible between 0.5-2.5 seconds.

from 0° to 360° (40° increments) and elevations of -45°, 0°, 45°, and 90°.

Our processing pipeline employed a Butterworth high-pass filter (500Hz cutoff) followed by an Impulse Response-based Filter-and-Sum beamformer, using measured impulse responses as steering vectors rather than analytical models. This data-driven approach captures complex acoustic environment characteristics including room reflections and array imperfections.

For automated source selection, we processed 3-second segments independently through beamforming, `birdnetlib` [13], and `BirdNET-Analyzer` model v2.4 [14]. For determining successful detections, we applied a confidence score threshold of 0.4, which represents a balance between detection sensitivity and false positive reduction as recommended in previous ecological monitoring studies [15, 16]. All detection events that exceeded this threshold were counted in our analysis of detection performance across varying SNR conditions. Our system automatically selects the direction yielding the highest BirdNET confidence score for each time segment, creating a comprehensive spatial acoustic map. For baseline comparison, we extracted single-channel audio from the same recordings and processed it directly with BirdNET without beamforming.

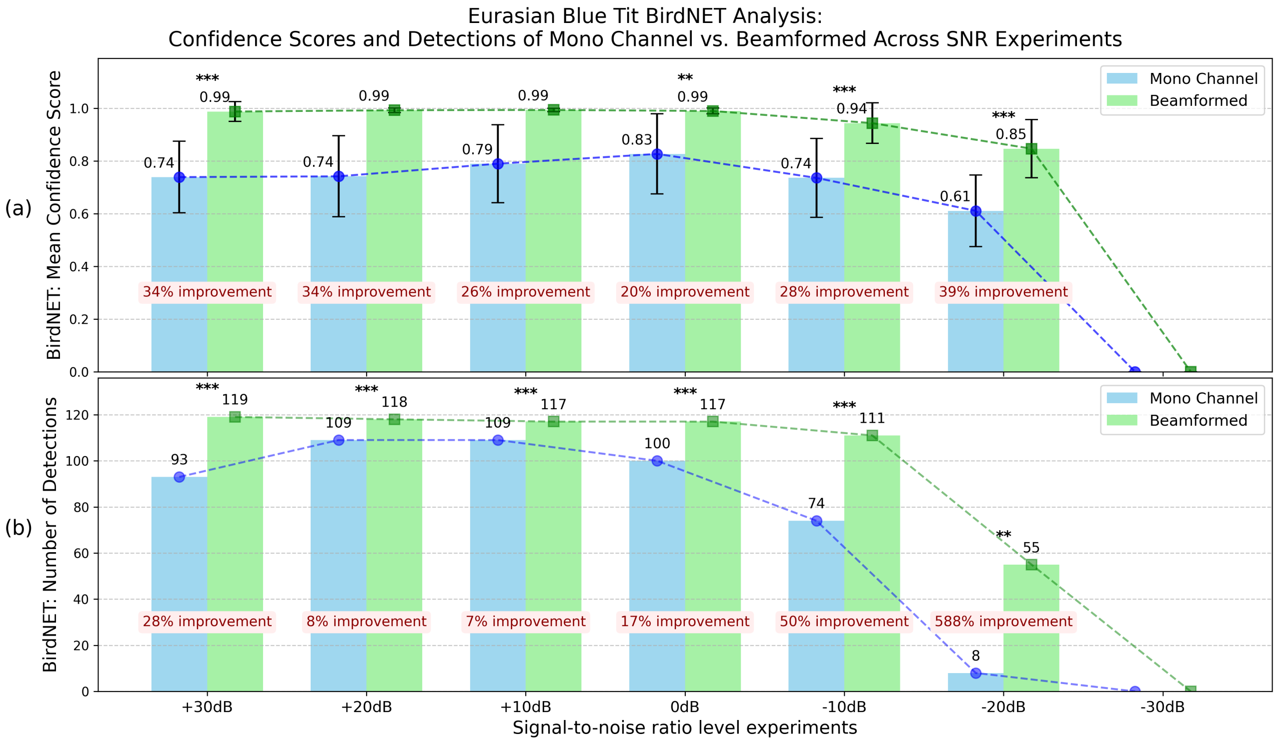

Figure 3. Performance comparison between mono channel recording and multidirectional beamforming for Eurasian Blue Tit detection across decreasing SNR levels. Panel (a) shows BirdNET confidence scores (0-1 scale), while panel (b) displays the number of successful detections (out of 124 possible events). Percentage values indicate the relative improvement of beamforming over mono channel recording. Error bars represent standard deviation.

3. RESULTS AND DISCUSSIONS

3.1 Detection Performance Comparison

Results demonstrate a clear superiority of multidirectional beamforming over conventional mono-channel recordings for birdcall detection. Higher BirdNET confidence scores were consistently achieved with the beamforming approach across various SNR levels (Fig. 3). Paired T-tests revealed significant differences in confidence scores at all SNR levels tested ($p < 0.05$), with Beamformed outperforming Mono Channel by 20-39% (Fig. 3). For instance, at +30dB, Beamformed achieved a mean confidence score of 0.9882 compared to 0.7392 for Mono Channel ($t = 17.0666$, $p < 0.0001$, ***; Fig. 3), with a large effect size (Cohen's $d = 2.6246$). This trend persisted even at lower SNRs, such as -20dB, where Beamformed maintained a confidence score of 0.8466 compared to 0.6107 for Mono Channel ($t = 4.3936$, $p = 0.0022$, **; Cohen's $d = 2.0474$). It is important to note that confidence scores are not direct probabilities and remain specific to the experimental setting and Eurasian Blue Tit species [16].

Both recording approaches showed declining confidence scores at extreme low SNR levels (-20dB and -30dB), with no detections recorded at -30dB where the birdcalls became virtually inaudible, aligning with our experimental design expectations.

3.2 Species Detection Accuracy at Varying SNR Levels

The experimental design incorporated 124 distinct birdcall events. At favourable SNR conditions (+30dB to -30dB), the beamforming approach detected up to 119 calls while mono-channel recordings detected a maximum of 109 calls (Fig. 3). Chi-square tests confirmed significant differences in detection rates at several SNR levels. For example, at +30dB, Beamformed detected 119 calls compared to 93 by Mono Channel (χ^2 = 20.3092, p < 0.0001, ***; Cramer's V = 0.2862), indicating a moderate effect size.

A pronounced performance gap emerged at lower SNRs. At -10dB, mono-channel detection dropped to 74 calls while beamforming maintained 111 detections (χ^2 = 27.5768, p < 0.0001, ***; Cramer's V = 0.3335). This disparity widened dramatically at -20dB, with beamforming detecting 55 calls compared to only 8 with the mono-channel approach (χ^2 = 45.0251, p < 0.0001, ***; Cramer's V = 0.4261), corresponding to a 588% improvement in detection counts (Fig. 3). At -30dB, neither method detected any calls, causing statistical analysis infeasible at this level.

These findings, supported by significance stars in Fig. 3, demonstrate the beamforming technique's capacity to maintain reliable detection in challenging acoustic environments, simulating scenarios where birds are at considerable distances from recording equipment.

3.3 Key Performance Trends

Analysis of the performance data reveals several important patterns that highlight the practical advantages of beamforming:

Increasing advantage at lower SNR levels: The most striking trend is how the performance gap between beamforming and mono-channel recording widens dramatically as SNR decreases. While beamforming shows modest improvements (7-28%) at high SNR (+30dB to 0dB), where differences in detection counts were sometimes not statistically significant (e.g., at +20dB, χ^2 = 3.3296, p = 0.0680), this advantage increases exponentially at challenging SNR levels. At -10dB, beamforming achieved 50% improvement in detection counts, which surged to a remarkable 588% improvement at -20dB, supported by a highly significant Chi-square test (χ^2 = 45.0251, p < 0.0001, ***; Fig. 3).

Robustness across varying SNR conditions: Beamforming maintains consistently high performance (confidence scores >0.94 and detection counts >110) across the +30dB to -10dB range, while mono-channel performance fluctuates more dramatically even at favourable SNR levels. For instance,, at 0dB, Beamformed detected 117 calls compared to 100 by Mono Channel (χ^2 = 9.4378, p = 0.0021, **; Cramer's V = 0.1951), highlighting greater stability and reliability regardless of source distance.

Extended detection threshold: The beamforming approach extends the practical detection threshold by approximately 10dB, maintaining useful detection capabilities (55 detections) at -20dB where mono-channel recording nearly fails completely (8 detections). This effectively increases the functional detection range of the recording system, as evidenced by the significant difference at -20dB (p < 0.0001, ***).

These trends suggest beamforming's greatest value lies in extending reliable detection to challenging acoustic environments, effectively increasing monitoring range and providing more consistent data quality regardless of source distance in field deployments.

3.4 Limitations and Practical Considerations

The multidirectional beamforming approach faces three primary challenges:

Computational requirements: Processing an 8-minute recording required approximately 14 minutes (2 minutes for beamforming across 31 directions plus 12 minutes for BirdNET classification), potentially limiting large-scale deployments.

Storage demands: The process generates 31 WAV beamforming files (15MB each) plus a mono file (3MB) from a single 45MB FLAC file, necessitating substantial storage capacity for extensive field deployments.

Complex Impulse Response Characterisation Requirements: An additional significant limitation is the complexity involved in measuring impulse responses for multichannel devices. This process requires specialised equipment (such as precision turntables and calibrated loudspeaker arrays), anechoic or controlled acoustic environments, and considerable technical expertise. The full characterisation of 3D impulse responses across multiple azimuth and elevation angles is time-consuming and must be repeated for each new recording device configuration. This represents a substantial barrier to implementation compared to single-microphone deployments, which require no such calibration procedures.

4. CONCLUSION

This study demonstrates that integrating multidirectional beamforming with AI-based species detection significantly enhances birdcall identification in challenging acoustic environments. Statistical analysis confirmed that Beamformed outperformed Mono Channel across all SNR levels, with highly significant differences in both confidence scores (p < 0.05) and detection counts (p < 0.1

at significant levels), as indicated by significance stars in Fig. 3. Our approach maintained high detection accuracy at SNRs as low as -20dB, far outperforming traditional mono-channel recordings.

The methodology offers a promising framework for developing more sophisticated PAM systems capable of providing valuable behavioral and ecological insights across various species. While focused on birds, the approach could be adapted for other species with vocalisation frequencies below 8kHz.

This initial study demonstrates promising results, but several important areas require further investigation. Field testing in real-life environments with unknown sound source positions is essential to validate performance under natural reverberation conditions and diverse forest soundscapes. A more comprehensive analysis examining how beamforming affects other ecological metrics beyond simple detection is also needed. Additionally, deeper analysis of how detection performance varies with source location could provide valuable insights for optimizing array designs. Parallel to these ecological validations, optimising the computational efficiency of the BirdNET classification process would enhance the practical applicability of this method for large-scale monitoring. These extensions would strengthen the ecological applications of the multidirectional beamforming approach and better quantify its advantages over conventional recording methods across diverse monitoring scenarios.

5. ACKNOWLEDGMENTS

This work is funded by the Indonesia Endowment Fund for Education / Lembaga Pengelola Dana Pendidikan (LPDP), Indonesia (Scholarship Identification Number: 20230522097975).

We are grateful to Neel Patrick Le Penru for granting us permission to use his field recordings in our experiment to recreate the forest soundscape in a virtual sound environment.

6. REFERENCES

[1] 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," *Methods in Ecology and Evolution*, vol. 9, pp. 2383–2387, Dec. 2018.

[2] S. R. P.-J. Ross, D. P. O'Connell, J. L. Deichmann, C. Desjonquères, A. Gasc, J. N. Phillips, S. S. Sethi, C. M. Wood, and Z. Burivalova, "Passive acoustic monitoring provides a fresh perspective on fundamental ecological questions," *Functional Ecology*, vol. 37, pp. 959–975, Apr. 2023.

[3] I. Tolkova and H. Klinck, "Source separation with an acoustic vector sensor for terrestrial bioacoustics," *The Journal of the Acoustical Society of America*, vol. 152, p. 1123, Aug. 2022.

[4] B. E. Heath, R. Suzuki, N. P. Le Penru, J. Skinner, C. D. L. Orme, R. M. Ewers, S. S. Sethi, and L. Picinali, "Spatial ecosystem monitoring with a Multichannel Acoustic Autonomous Recording Unit MAARU," *Methods in Ecology and Evolution*, vol. 15, pp. 1568–1579, Sept. 2024.

[5] J. P. Ethier and D. R. Wilson, "Using microphone arrays to investigate microhabitat selection by declining breeding birds," *Ibis*, vol. 162, no. 3, pp. 873–884, 2020.

[6] F. Grondin, D. Létourneau, C. Godin, J.-S. Lauzon, J. Vincent, S. Michaud, S. Faucher, and F. Michaud, "ODAS: Open embeddeD Audition System," *Frontiers in Robotics and AI*, vol. 9, p. 854444, May 2022.

[7] N. Le Penru, A. Borrelli, B. Heath, R. Ewers, S. Sethi, and L. Picinali, "Development and Validation of a Test Platform for Spatial EcoAcoustic Technologies," in *Proc. of the 10th Convention of the European Acoustics Association Forum Acusticum 2023*, (Turin, Italy), pp. 4671–4674, European Acoustics Association, Jan. 2023.

[8] Seeed Studio, "ReSpeaker 6-Mic Circular Array Kit for Raspberry Pi | Seeed Studio Wiki." https://wiki.seeedstudio.com/ReSpeaker_6-Mic_Circular_Array_kit_for_Raspberry_Pi/, Jan. 2023.

[9] B. E. Heath, "MAARU - Multichannel Acoustic Autonomous Recording Unit." https://beckyheath.github.io/MAARU/index.html, 2024.

[10] H. S. S. C. Sagar, A. Anand, M. E. Persche, A. M. Pidgeon, B. Zuckerberg, Ç. H. Şekercioğlu, and Z. Buřivalová, "Global analysis of acoustic frequency characteristics in birds," *Proc. of the Royal Society B: Biological Sciences*, vol. 291, p. 20241908, Nov. 2024.

[11] N. P. Le Penru, B. E. Heath, J. Dunning, L. Picinali, R. M. Ewers, and S. S. Sethi, "Towards using virtual acoustics for evaluating spatial ecoacoustic monitoring technologies," *Methods in Ecology and Evolution*, vol. 16, no. 1, pp. 108–125, 2025.

[12] D. G. Malham and A. Myatt, "3-D Sound Spatialization using Ambisonic Techniques," *Computer Music Journal*, vol. 19, no. 4, pp. 58–70, 1995.

[13] J. Weiss, "Birdnetlib v0.18.0 documentation." https://joeweiss.github.io/birdnetlib/, 2025.

[14] 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.

[15] D. Funosas, L. Barbaro, L. Schillé, A. Elger, B. Castagneyrol, and M. Cauchoix, "Assessing the potential of BirdNET to infer European bird communities from large-scale ecoacoustic data," *Ecological Indicators*, vol. 164, p. 112146, July 2024.

[16] C. M. Wood and S. Kahl, "Guidelines for appropriate use of BirdNET scores and other detector outputs," *Journal of Ornithology*, Feb. 2024.