

Short Communications

A FIRST ASSESSMENT OF BIRDNET PERFORMANCE AT VARYING DISTANCES: A PLAYBACK EXPERIMENT

UN PRIMER ANÁLISIS DEL RENDIMIENTO DE BIRDNET A DISTANCIAS VARIABLES: UN EXPERIMENTO CON *PLAYBACK*

Cristian PÉREZ-GRANADOS^{1, 2} *

SUMMARY.—Bird vocalisations, like any other acoustic signals, attenuate over distance, and therefore their structure degrades progressively. Such degradation may have an impact on the ability of automated signal recognition software, to detect and correctly identify bird vocalisations. BirdNET is a recently launched automated bird song recogniser commonly employed by researchers and the public. However, few studies have assessed its performance and our current knowledge about how BirdNET performance may vary over distance or with species is very limited. I aimed to evaluate whether BirdNET's ability to correctly identify bird vocalisations of three bird species varied over distance or according to recorder type and target species using a playback broadcast from 10 to 150m away. BirdNET's ability to correctly identify bird songs varied among species and generally decreased over distance but did not vary among recorder types. Overall BirdNET recall rate, defined as the percentage of vocalisations detected, and correctly identified, by the software, was 59.9% (499 vocalisations correctly identified of 840 vocalisations broadcast). A significantly higher number of vocalisations were correctly identified when broadcast at 50m or closer (mean recall rate of 92.2%), when compared to vocalisations broadcast farther than that distance (mean recall rate of 34.9%). Recall rate was also significantly higher for the Grasshopper Sparrow and the Hooded Warbler, when compared to the Gray Vireo. The number of misclassifications varied over distances and did not follow a linear pattern. This study provides valuable information that may contribute to improved surveys and for expanding the use of BirdNET for surveying bird communities using passive acoustic monitoring. —Pérez-Granados, C. (2023). A first assessment of BirdNET performance at varying distances: a playback experiment. *Ardeola*, 70: 221-233.

Key words: automated detection, automated software, autonomous recording unit, bird monitoring, bird recognition, Convolutional Neural Network, passive acoustic monitoring.

¹ Ecology Department, Universidad de Alicante, 03080 Alicante, Spain.

² Biodiversity and Animal Conservation Lab, Landscape Dynamics and Biodiversity Programme, Forest Science and Technology Center of Catalonia (CTFC), Solsona, Catalonia, Spain.

* Corresponding author: cristian.perez@ua.es

RESUMEN.—Las vocalizaciones de las aves, como cualquier otra señal acústica, se atenúan con la distancia y, por lo tanto, la estructura de las vocalizaciones de las aves se degrada progresivamente. Tal degradación puede tener un impacto en la capacidad de programas automatizados de reconocimiento de señales a la hora detectar e identificar correctamente las vocalizaciones de las aves. BirdNET es un reconocedor automatizado de cantos de pájaros de reciente creación y comúnmente empleado por investigadores y el público. Sin embargo, pocos estudios han evaluado su rendimiento y nuestro conocimiento actual sobre cómo el rendimiento de BirdNET puede variar en función de la distancia o entre especies es muy limitado. Aquí, mi objetivo era evaluar si la habilidad de BirdNET para identificar correctamente las vocalizaciones de tres especies de aves variaba según la distancia, el tipo de grabadora empleada y entre especies, utilizando una grabación reproducida de 10 a 150 m. La habilidad de BirdNET para identificar correctamente los cantos de las aves varió entre las especies y, en general, disminuyó con la distancia, pero no varió entre los dos tipos de grabadores testados. La tasa de detección de BirdNET, definida como el porcentaje de vocalizaciones detectadas e identificadas correctamente por el *software*, fue del 59,9% (499 vocalizaciones identificadas correctamente de las 840 vocalizaciones reproducidas). Se identificó de manera correcta y significativa un mayor número de vocalizaciones cuando se emitieron a 50 m o más cerca (tasa media de detección del 92,2%), en comparación con las vocalizaciones emitidas a más de esa distancia (tasa media de detección del 34,9%). La tasa de detección también fue significativamente más alta para el chingolo saltamontes y la reinita encapuchada, en comparación con el vireo gris. El número de clasificaciones erróneas varió con las distancias y no siguió un patrón lineal. Ese estudio proporciona información valiosa que puede contribuir a mejorar futuros muestreos y a expandir el uso de BirdNET para censar comunidades de aves usando monitoreo acústico pasivo.—Pérez-Granados, C. (2023). Un primer análisis del rendimiento de BirdNET a distancias variables: un experimento con *playback*. *Ardeola*, 70: 221-233.

Palabras clave: detección automatizada, reconocimiento de aves, Red Neuronal Convolutiva, seguimiento acústico pasivo, seguimiento de aves, *software* automatizado, unidad de grabación autónoma.

The development of new technologies has changed ecological monitoring in the recent decades (reviewed by Lahoz-Monfort & Magrath, 2021). Some of these new technologies require the use of computer programs and automated software to deal with the vast amounts of data collected. For example, managing and visualising sound recordings collected using passive acoustic monitoring (PAM) can be a very resource-consuming task and a wide range of algorithms have been developed to alleviate this burden, including automated species identification (e.g. Zhao *et al.*, 2017). Among recently developed algorithms for automated bird identification on sound recordings is BirdNET (reviewed by Pérez-Granados, 2023). BirdNET has already proved its utility for bird monitoring purposes and

as a reliable research resource (Cole *et al.*, 2022; Manzano-Rubio *et al.*, 2022; Wood *et al.*, 2022).

BirdNET is a free automated bird song recogniser able to identify over 3,000 bird species (see full description in Kahl *et al.*, 2021). It is a convolutional neural network that identifies bird vocalisations in three-second fragments within longer audio recordings. Recordings can be directly collected with a smartphone using the BirdNET APP (Wood *et al.*, 2022), uploaded into a web-based platform (<https://birdnet.cornell.edu/api>), or run as a Python script (Manzano-Rubio *et al.*, 2022, <https://github.com/kahst/BirdNET-Analyzer>). For each three-second fragment of an audio recording BirdNET provides a bird species identification with a quantitative confidence score, which ranges

from 1 (perfect match, high confidence in the identification) to 0 (no match).

Kahl *et al.* (2021) assessed BirdNET's ability (precision) to correctly classify bird vocalisations of 984 species in focal recordings and found an overall precision of 0.791 (see an extended review in Pérez-Granados, 2023). However, bird vocalisations, as with any other acoustic signal, attenuate over distance and therefore inevitably degrade progressively during propagation from the signaller to the receiver (Naguib & Wiley, 2001). The distances at which bird vocalisations can travel and be recorded depend on several factors, such as frequency of bird vocalisations (Rempel *et al.*, 2013), habitat type (Yip *et al.*, 2017), or microphone quality (usually measured as signal-to-noise ratio, SNR, Darras *et al.*, 2021), among others. Identifying and measuring the area within which bird vocalisations can be recorded is important to adjustment of acoustic estimates derived from sound recordings, such as species richness, activity, abundance or density (Darras *et al.*, 2016). Likewise, signal degradation over distance also has an impact on the ability of automated signal recognition software, such as BirdNET, to detect and correctly classify bird vocalisations (see Knight *et al.*, 2020 for a relationship between recognizer score and the energy of acoustic signals). Therefore, it is expected that BirdNET performance will decrease with distance, but to date, there is no information available about the impact of bird distance from the recorder on BirdNET performance (Pérez-Granados, 2023).

I investigated the relationship between bird distance from the recorder and BirdNET precision and recall rate for three bird species and whether these relationships vary with the recorder type employed, since recorders usually differ in microphone quality (see extended recorder comparisons for bird monitoring in Rempel *et al.*, 2013 and Pérez-Granados *et al.*, 2019, see a microphone

comparison for wildlife monitoring in Darras *et al.*, 2021). My approach relied on using vocalisations broadcast at known distances from four acoustic recorders (two different types) deployed in a semi-open habitat. Although based on a limited number of species and recorders, understanding how BirdNET performance varies according to distance, devices and species might be useful, for example, to define the sampling area of surveys using PAM and BirdNET, which is needed to model detection accuracies, develop site occupancy models or to estimate bird richness or bird density around recorders.

Field tests were carried out in the hour before dawn (6:00 a.m. local time, UTC +2) on 5 September 2022 in the "Pinada Villamontes" (San Vicent del Raspeig, Alicante, eastern Spain, 38.26° N, 0.31° W). This is a flat and homogeneous area with a low-density of pines *Pinus* spp. and no shrub or grass cover (see Supplementary Electronic Material, Image S1). Although the recorders were mounted inside the pine forest, the area could be considered as a semi-open habitat given the low tree density and absence of any direct blocking of sound between the recorders and the loudspeaker. The mean air temperature during field tests was around 22°C. As wind speed was constant and lower than 2m/s, I did not consider it in my analyses. In addition, no natural bird vocal activity was detected during the field tests. The recorders used were two Song Meter Micro (ca. 249\$) and two Song Meter Mini devices (499\$, Wildlife Acoustics Inc.). Both recorders are equipped with built-in omnidirectional microphones, but while the Song Meter Micro microphone is integrated within the device and has a SNR of 73dB at 1kHz, the Song Meter Mini can have one (as here) or two external microphones with a higher SNR of 78dB at 1kHz (See Supplementary Electronic Material, Image S2). The four recorders were acquired during the study year and only used during a few weeks in spring 2022. The

acoustic devices were attached to a 70-90cm wooden stick and remained in the same position, below the forest canopy and with microphones oriented towards the loudspeaker, during the tests (Supplementary Electronic Material, Image S2). Recordings were made using a sampling rate of 22.050kHz, +18dB gain, and 16-bit “.wav” format.

I broadcast a digital standardised recording (playback hereinafter) that consisted of ten vocalisations of each of the three following species: the Hooded Warbler *Setophaga citrina*, the Grasshopper Sparrow *Ammodramus savannarum*, and the Gray Vireo *Vireo*

vicinior (the vocalisations of the three species can be found in Supplemental Audio S1). Species selection was based on previous studies that proved the ability of BirdNET to correctly identify these species (Malamut *et al.*, 2022), which makes them good study models for evaluating how BirdNET performs over different distances. Furthermore, the three are American species whose vocalisations would be unlikely to provoke a response from, or to disturb, local birds. There was enough separation (three seconds minimum) between vocalisations for them to be within different fragments analysed

Assessment of BirdNET performance

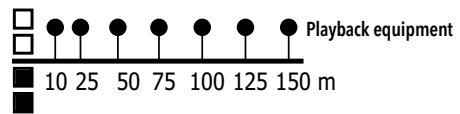
Two recorder types (2x)

- Song Meter Micro
- Song Meter Mini

Three bird species

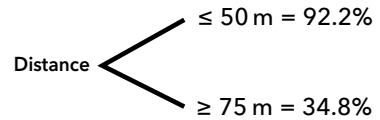
- Grasshopper Sparrow
- Gray Vireo
- Hooded Warbler

Seven distance categories



Overall BirdNET recall rate: 59.9% (499 of 840)

- Song Meter Micro (58.1%)
- Song Meter Mini (60.7%)
- Grasshopper Sparrow (62.5%)
- Gray Vireo (47.9%)
- Hooded Warbler (67.9%)



Number of BirdNET mislabelled detections: 31

- Song Meter Micro (17)
- Song Meter Mini (14)
- Grasshopper Sparrow (5)
- Gray Vireo (26)
- Hooded Warbler (0)

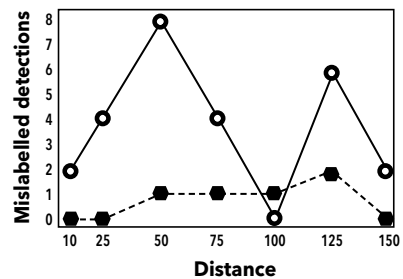


FIG. 1.—Summary graph showing the methodology and main results of the assessment of BirdNET performance at variable bird distances to the recorder and using different types of recorders and bird species.

[Resumen gráfico mostrando la metodología y principales resultados del análisis del rendimiento de BirdNET a distancias variables del ave al grabador y usando diferentes tipos de grabadores y especies de aves.]

by BirdNET and so that every vocalisation had the chance to be independently identified by BirdNET. Indeed, before performing the field tests I proved that BirdNET was able to correctly identify the 30 vocalisations included in the playback, by obtaining a recall rate and a precision of 1 for the source recordings (see BirdNET analyses section for recall rate and precision description). Such assessment was made by uploading the playback to the BirdNET web-based platform (see below) and verifying that BirdNET's detections agreed with the target species. For each species, the playback included a total of ten vocalisations, with five vocalisations extracted from each of two high-quality recordings downloaded from the Xeno-canto public library (<https://xeno-canto.org/>). The loudspeaker consisted of a digital Samsung mp3 coupled with a RadioShack amplifier (frequency response 100Hz-10KHz, distortion 1KHz < 2%) situated one metre above the ground throughout the tests. Playback volume was around 76dB, trying to emulate the normal singing volume of birds (see similar approaches using playbacks in Rempel *et al.*, 2013; Pérez-Granados *et al.*, 2019). Sound level was measured with a Volcraft SL 400 sound level meter placed one metre from the loudspeaker. The playback was broadcast once from each of seven fixed locations: at distances of 10, 25, 50, 75, 100, 125 and 150m (Figure 1).

Collected recordings were split into seven segments (one for each distance category) and uploaded into the BirdNET web-based platform (<https://birdnet.cornell.edu/api>). I selected that version of BirdNET because it offers an easy way to verify the detected vocalisations by looking at the spectrogram of the uploaded recording and BirdNET output simultaneously. Therefore, the validation of whether the labels were correctly identified or incorrect was confirmed visually and acoustically, if needed. Sensitivity was set as 1.0 (default value). I did not use any con-

fidence score as threshold and so a vocalisation was considered correctly identified when the target species was labelled as such by BirdNET, regardless of the quantitative confidence score (minimum quantitative score for detections is 0.01). For each distance, species, and recorder I noted whether or not the broadcast vocalisation was correctly identified by BirdNET. For non-identified vocalisations I noted whether the vocalisation was just undetected (missed detection, false negative) or labelled as a different bird species (mislabelled detection, false positive). I evaluated whether the precision (probability of correctly identifying a bird vocalisation) and the recall rate (percentage of songs of a target species correctly identified by BirdNET in relation to the total number of songs broadcast) varied over distance.

I applied a mixed logistic regression model in R 3.6.2 (R Development Core Team, 2019) with family binomial to evaluate whether the probability of correctly identifying a vocalisation varies over distance (log transformed), between two recorder types (two levels), and species (three levels). These were used as fixed variables, and probability of identification (identified/non-identified) was the dependent variable. Recorder ID was considered as random effect. When a fixed variable was found to be significant, I ran Tukey *post-hoc* comparisons to test for differences among the levels of the variables by using the multcomp package (Hothorn, 2008).

BirdNET correctly identified 499 of the 840 vocalisations broadcast (59.4% of the total). However, the recall rate greatly varied between distances, with 332 of the 360 vocalisations broadcast ≤ 50 m being correctly identified (92.2%), while only 167 of the 480 vocalisations broadcast ≥ 75 m were correctly identified (34.8%, see Table 1 for a detailed table showing the number of vocalisations correctly identified at each distance for each recorder and target species). BirdNET precision varied over distances, with just 1-2

TABLE 1

Numbers of vocalisations correctly identified by BirdNET per recorder and distance. Results are shown separately for each of the three target species. Ten vocalisations of each species were broadcast at each distance.

[Número de vocalizaciones identificadas correctamente por BirdNET en función del tipo de grabador y distancia. Los resultados se muestran por separado para cada una de las tres especies analizadas. A cada distancia un total de 10 vocalizaciones fueron reproducidas.]

Recorder	Species	10m	25m	50m	75m	100m	125m	150m	TOTAL
Song Meter Micro A	Grasshopper Sparrow	10	10	10	10	0	0	0	40
	Gray Vireo	9	7	8	2	3	1	0	30
	Hooded Warbler	10	10	10	10	8	7	2	57
	TOTAL	29	27	28	22	11	8	2	127
Song Meter Micro B	Grasshopper Sparrow	10	10	9	8	0	0	0	37
	Gray Vireo	10	10	9	2	2	0	0	33
	Hooded Warbler	10	10	9	9	6	2	1	47
	TOTAL	30	30	27	19	8	2	1	117
Song Meter Mini A	Grasshopper Sparrow	10	10	10	10	7	3	0	50
	Gray Vireo	9	9	6	5	2	1	2	34
	Hooded Warbler	9	9	10	8	2	1	0	39
	TOTAL	28	28	26	23	11	5	2	123
Song Meter Mini B	Grasshopper Sparrow	10	10	10	10	8	0	0	48
	Gray Vireo	8	8	6	6	5	3	1	37
	Hooded Warbler	9	9	9	9	7	3	1	47
	TOTAL	27	27	25	25	20	6	2	132

vocalisations mislabelled at 10m, 100m and 150m, but up to 8-9 vocalisations mislabelled at 50m and 125m (see Figure 1 and Supplementary Electronic Material, Table S1, for detailed table). BirdNET precision differed among species, with no mislabelled detections for the Hooded Warbler, five vocalisations mislabelled for the Grasshopper Sparrow and 26 mislabelled detections for the Gray Vireo. Four of the five mislabelled detections of the Grasshopper Sparrow were identified as vocalisations of closely-related

species, such as Cassin's Sparrow *Peucaea cassinii* (n = 2) and Baird's Sparrow *Centronyx bairdii* (n = 2), but in one case it was misclassified as the Ruddy Shelduck *Tadorna ferruginea*. Similarly, among the 18 mislabelled detections of the Gray Vireo there were also several vocalisations identified as vocalisations of a close-relative, the Plumbeous Vireo *Vireo plumbeus* (n = 5), but the species most often mislabelled as Gray Vireo was the Spotted Redshank *Tringa erythropus* (n = 8). The other species mislabelled as

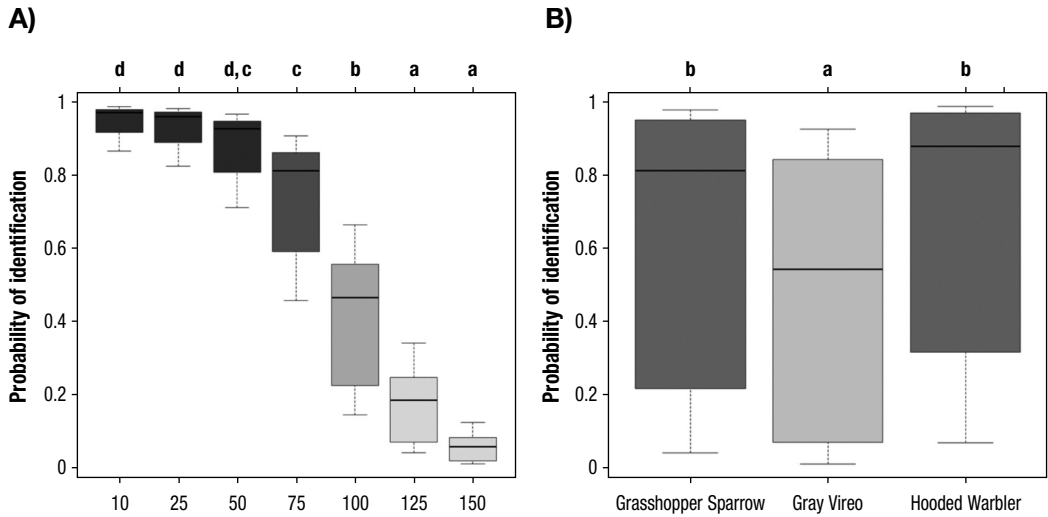


FIG. 2.—Boxplots showing the probability of BirdNET correctly identifying a bird vocalisation as a function of (A) bird distance to the recorder and (B) bird species. Calculations of the graphs are based on a logistic regression using the correct identification/no-identification of vocalisations as the dependent variable and distance of playback to the recorder, bird species and recorder type as the predictor variables. Recorder ID was fitted as random effect. Similar colours and different letters mean significant differences in the identification success according to the Tukey test.

[Diagramas de cajas mostrando la probabilidad de BirdNET en identificar correctamente una vocalización de un ave en función de (A) la distancia del ave al grabador y (B) especie de ave. Los cálculos de las gráficas se basan en una regresión logística usando como la correcta identificación/no identificación de las vocalizaciones como variable dependiente y la distancia del altavoz al grabador, especie de ave y tipo de grabador como variables predictoras. El identificador del grabador fue considerado como efecto aleatorio. Los colores similares y letras diferentes muestran diferencias significativas en la probabilidad de identificación según el test de Tukey.]

Gray Vireo were the Ruddy Shelduck ($n = 5$), the Hepatic Tanager *Piranga flava* ($n = 2$), the Little Ringed Plover *Charadrius dubius* ($n = 2$), the Pine Grosbeak *Pinicola enucleator* ($n = 1$), the Booted Eagle (*Hieraaetus pennatus*, ($n = 1$), the Yellow-breasted Chat *Icteria virens* ($n = 1$), and the Thekla's Lark *Galerida theklae*, ($n = 1$).

According to the logistic regression the probability of BirdNET correctly identifying a bird vocalisation varied over distance and the species considered (Figure 2; Table 2). The probability of correctly identifying a bird was similar among the first three distance categories, from 10m to 50m, which was

significantly higher than those obtained at greater distances (Figure 2). The probability of correctly identifying a bird vocalisation broadcast at 75m was significantly lower when compared to those broadcast ≤ 50 m, but significantly higher in comparison to those when the playback was broadcast at ≥ 100 m. Similarly, vocalisations emitted at 125m and 150m showed significantly lower probabilities of being correctly identified than those broadcast at ≤ 100 m (Figure 2 and Table 2).

According to the logistic regression, there were no significant differences in BirdNET's probability of correctly identifying a bird

TABLE 2

Summary table of the results of a logistic regression testing the relationship between BirdNET probability of correctly identifying a broadcast vocalisation and Distance (seven levels), Recorder type (two levels) and Species (three levels). Recorder ID was fitted as random effect. Estimates are expressed as the differences from the intercept, which was estimated using the recordings made at 10m, using the Song Meter Micro A recorder and the Grasshopper Sparrow as reference values. Graph results of the Tukey *post-hoc* tests to assess differences among distance categories and species can be found in Figure 2.

[Tabla resumen de los resultados de una regresión logística analizando la relación entre la probabilidad de BirdNET de identificar correctamente una vocalización reproducida con un altavoz y la distancia (siete niveles), tipo de grabador (dos niveles) y especie (tres niveles). El identificador del grabador fue considerado como efecto aleatorio. Las estimas están representadas como las diferencias respecto al intercepto, el cual se ha estimado considerando las grabaciones realizadas a 10 m, el tipo de grabador Song Meter Micro A y el chingolo saltamontes como valores de referencia. Resultados gráficos del test de Tukey para analizar las diferencias en función de las categorías de distancias y especies pueden encontrarse en la figura 2.]

	<i>Estimate</i>	<i>Std. Error</i>	<i>Z-value</i>	<i>P</i>
(Intercept)	3.355	0.474	7.068	<0.0001
Distance (25m)	-0.316	0.565	-0.560	0.576
Distance (50m)	-0.963	0.517	-1.863	0.062
Distance (75m)	-2.025	0.481	-4.211	<0.0001
Distance (100m)	-3.614	0.479	-7.546	<0.0001
Distance (125m)	-4.943	0.508	-9.726	<0.0001
Distance (150m)	-6.220	0.595	-10.444	<0.0001
Song Meter Mini	0.236	0.207	1.136	0.226
Gray Vireo	-1.284	0.263	-4.879	<0.0001
Hooded Warbler	0.509	0.263	1.938	0.053

vocalisation between recorder types (see Table 2). In terms of species, 134 Gray Vireo vocalisations 175 Grasshopper Sparrow and 190 Hooded Warbler vocalisations were correctly identified (Table 1). Therefore, the Gray Vireo was less often correctly identified than the other species (Figure 1). Indeed, the Tukey test showed that the Hooded Warbler and the Grasshopper Sparrow were significantly more often correctly identified by BirdNET than the Gray Vireo (Figure 2).

The maximum distance at which BirdNET was able to correctly identify a Gray Vireo and Hooded Warbler vocalisation was 150m, while the maximum distance for the Grasshopper Sparrow was 125m (Table 1).

The data collected here provide a first assessment of how BirdNET performance varies over distance. Nonetheless, the experiment considered a limited number of species and was carried out under the most favourable recording conditions i.e. with

no background noise and the loudspeaker facing the recorders (see Pérez-Granados *et al.*, 2019 for the effect of loudspeaker direction on bird song detection by an automated recogniser). In these circumstances, the results found should be considered with caution. The total number of mislabelled detections was very low (31 of 840 vocalisations broadcast, 3.7%). BirdNET precision was relatively high at all considered distances (10–150m), and was maximal at 10m, 100m and 150m. The high recall rate at 10m can be explained because vocalisations broadcast at that distance were recorded at high quality (i.e. high sound pressure level), and therefore more easily identified by BirdNET. Conversely, vocalisations recorded at 150m were weak sounds and very few were detected by BirdNET. The high precision at 100m is explained because there were no Gray Vireo vocalisations mislabelled, this being the species most often mislabelled and so contributing most to that index. However, up to four and six Gray Vireo vocalisations were mislabelled at 75m and 125m, respectively (see Supplementary Electronic Material, Figure S1). The low number of mislabelled detections at 100m is difficult to explain, since there was no patent variation in ambient noise. Indeed, the number of Gray Vireo vocalisations correctly identified at 100m (12) was relatively close to that obtained at 75m (15), which corroborates the assumption that recording quality was as expected. The recall rate was very high within the first 50m, but after that distance the probability of identifying a bird vocalisation significantly decreased, which is particularly true for the Gray Vireo. The vocalisations broadcast at ≥ 100 m were sounds that a human observer would have difficulty detecting owing to the poor quality of the recorded vocalisations, and thus explaining the low recall rate at the most distant categories. These results are in agreement with previous studies assessing the probability of human or other automated

software detecting bird vocalisations at increasing distances (see Rempel *et al.*, 2013; Yip *et al.*, 2017; Pérez-Granados *et al.*, 2019). A low recall rate although might be a problem for some specific research goals, for example when conducting detailed behavioural studies or employing cue counting). However, a low recall rate would have a minor impact in most monitoring programmes, since it could be largely mitigated by using extended recording schedules and be accounted for with imperfect detection (see Cole *et al.*, 2022).

The main finding of this study is the identification of a 50m radius around recorders within which BirdNET has a high recall rate and makes accurate identifications. That finding may be useful to fix the detection radius of the recorder for a bird species or to model detection accuracies, which will improve future monitoring programmes aiming to develop site occupancy models, estimate bird richness, or bird density, using passive acoustic monitoring and BirdNET (Darras *et al.*, 2018; Pérez-Granados & Traba, 2021). Nonetheless, the identified radius cannot be directly exported to other studies, and specific estimates should be carried out for each species, recorder type and habitat type. Initially, I would have predicted that the quantitative confidence score provided by BirdNET to the vocalisations recorded at variable distances could have been used as a threshold to delimit a fixed radius around recorders (see Knight *et al.*, 2020 for a relationship between recogniser score and the energy of an acoustic signal). However, over 90% of the identified vocalisations had a confidence score of 1 (see Supplementary Electronic Material, Table S2), and thus this parameter does not seem to be useful to make any inference about bird distance from the recorder. Nonetheless, I would again like to highlight that the experiment was carried out under almost perfect recording conditions (i.e. low wind speed, no birds vocalising). However, when recording in noisier conditions, such

as during the dawn chorus or with higher wind speed, the relationship between bird distance to the recorder and confidence score would probably be weaker. e.g. owing to overlapping foreground sounds. The same reasoning would be valid for the recall rate, so recording in more challenging conditions, such as the dawn chorus, would probably decrease the recall rate of BirdNET. The distribution of confidence scores provided by BirdNET can be also influenced by the “sensitivity” setting. Therefore, the number of high- and low-scoring predictions could be also adjusted. Sensitivity cannot be modified in the BirdNET web-based platform, where I ran the analyses, but it can be modified in the desktop version of BirdNET (<https://github.com/kahst/BirdNET-Analyzer>). Future studies aiming to further evaluate the relationship between bird distance from the recorder and confidence score may evaluate the use of variable values of sensitivity. A viable approach to estimate the approximate distance at which birds vocalise to the recorder might be to measure the sound pressure level (SPL) from BirdNET detections and estimate the relationship between SPL and bird distance to the recorder (see similar approaches in Yip *et al.*, 2017; Hedley *et al.*, 2021). It might be also useful to assess the relationship between SPL of BirdNET detections and the confidence score provided by BirdNET (see similar approach in Knight *et al.*, 2020).

The analyses did not show differences between the two types of recorders employed. Although I did not aim to achieve a comprehensive comparison between all acoustic devices available I recognise that the assessment was based upon a small sample size. Moreover, both recorders had a similar SNR, which may partly explain their similar performance (Darras *et al.*, 2021). Therefore, it would be desirable to perform similar studies using a larger set of recorders commonly employed for bird monitoring (e.g. Audio-

Moth, Song Meter SM4, SwiftOne) to provide fuller recommendations for choosing the most effective device for a monitoring program (see extended comparisons of acoustic devices for bird monitoring in Rempel *et al.*, 2013; Yip *et al.*, 2017; Pérez-Granados *et al.*, 2019; Toenies & Rich, 2021). More specifically, it would be desirable to assess the relationship between the microphone SNRs of a larger set of recorders and BirdNET performance, since SNR is one of the few factors that researchers can control by design and that largely determine the sound detection space of recorders (Darras *et al.*, 2021).

The results manifest large differences in BirdNET performance among the three species considered, with variable numbers of vocalisations correctly identified, mislabelled detections or maximum detection distances. This study shows that the Gray Vireo was the species most often mislabelled and the one with the lowest recall rate, while the Hooded Warbler was the species with the largest number of detections and with a precision of 1 (all BirdNET’s detections were true positives). These results are in agreement with previous studies assessing BirdNET performance (reviewed by Pérez-Granados, 2023). For example, Cole *et al.* (2022) found that BirdNET precision for detecting the California Quail *Callipepla californica* and the Common Raven *Corvus corax* was 0.96 and 0.29, respectively; while the recall rate for these two species was 0.68 and 0.11, respectively. To the best of my knowledge, the causes influencing variable BirdNET performance among species have not yet been evaluated (reviewed by Pérez-Granados, 2023). A possible explanation might be on the training data sample size used for creating species recognisers in BirdNET, since usually the larger the training sample the better the recogniser performance (Knight *et al.*, 2020). However, it is not possible to know (at least not immediately) the size of the training data sample employed for each

species in BirdNET. Further research should aim to evaluate whether BirdNET precision and recall rates vary among species according to their vocalisation characteristics, such as mean duration, main frequency or degree of complexity, but also with the size of training data employed.

BirdNET has a lot of potential and currently is commonly being used by the scientific community and the general public (Cole *et al.*, 2022; Wood *et al.*, 2022). Nonetheless, we still need more effort towards making passive acoustic surveys more efficient and accurate and should be cautious about the circumstances under which BirdNET can be used. The results of this study provides useful information that will likely contribute to expanding the use of BirdNET for surveying bird communities. The field tests were based upon a small sample size and the bird species were selected because of their known high detectability by BirdNET (Malamut *et al.*, 2022). I therefore encourage continued evaluation of BirdNET with a larger number of species: including randomly selected ones and with different recorder models, microphone qualities (Darras *et al.*, 2021), and recording schedule settings (e.g. sample rate). In particular, evaluation in different habitat types is even more important, since BirdNET's ability to identify bird vocalisations over distance will probably decrease within more closed habitats, such as rainforest (Yip *et al.*, 2017).

ACKNOWLEDGEMENTS.—I thank Wildlife Acoustics and Weboryx providing me with the devices. This study is a contribution to the project “Illuminating the nocturnal bird migration on islands: assessing the relationship between soundscape indices and migration diversity and intensity”. I also thank Kevin Darras, Connor Wood, Jesús M. Avilés and two anonymous referees whose comments helped to improve the manuscript. I am grateful to Ernest Garcia whose comments helped to improve the English quality of the manuscript.

FUNDING.—The author acknowledges support from Ministerio de Educación y Formación Profesional through the Beatriz Galindo Fellowship (Beatriz Galindo-Convocatoria 2020).

REFERENCES

- Cole, J.S., Michel, N.L., Emerson, S.A. & Siegel, R.B. (2022). Automated bird sound classifications of long-duration recordings produce occupancy model outputs similar to manually annotated data. *Ornithological Applications*, 124: duac003.
- Darras, K., Pütz, P., Rembold, K. & Tschardt, T. (2016). Measuring sound detection spaces for acoustic animal sampling and monitoring. *Biological Conservation*, 201: 29-37.
- Darras, K., Batáry, P., Furnas, B.J., Grass, I., Mulyani, Y.A. & Tschardt, T. (2019). Autonomous sound recording outperforms human observation for sampling birds: a systematic map and user guide. *Ecological Applications*, 29: e01954.
- Darras, K.F., Deppe, F., Fabian, Y., Kartono, A.P., Angulo, A., Kolbrek, B., Mulyani, Y.A. & Prawiradilaga, D.M. (2020). High microphone signal-to-noise ratio enhances acoustic sampling of wildlife. *PeerJ*, 8: e9955.
- Hedley, R.W., Wilson, S.J., Yip, D.A., Li, K. & Bayne, E.M. (2021). Distance truncation via sound level for bioacoustic surveys in patchy habitat. *Bioacoustics*, 30: 303-323.
- Hothorn, T., Bretz, F. & Westfall, P. (2008). Simultaneous Inference in General Parametric Models. *Biometrical Journal*, 50: 346-363.
- Kahl, S., Wood, C.M., Eibl, M. & Klinck, H. (2021). BirdNET: A deep learning solution for avian diversity monitoring. *Ecological Informatics*, 61: 101236.
- Knight, E.C., Sólomos, P., Scott, C. & Bayne, E.M. (2020). Validation prediction: a flexible protocol to increase efficiency of automated acoustic processing for wildlife research. *Ecological Applications*, 30: e02140.
- Lahoz-Monfort, J.J. & Magrath, M.J. (2021). A comprehensive overview of technologies for species and habitat monitoring and conservation. *BioScience*, 71: 1038-1062.

- Malamut, E.J. (2022). *Using autonomous recording units and image processing to investigate patterns in avian singing activity and nesting phenology* (Doctoral dissertation. University of California, Los Angeles).
- Manzano-Rubio, R., Bota, G., Brotons, L., Soto-Largo, E. & Pérez-Granados, C. (2022). Low-cost open-source recorders and ready-to-use machine learning approaches provide effective monitoring of threatened species. *Ecological Informatics*: 101910.
- Naguib, M. & Wiley, R.H. (2001). Estimating the distance to a source of sound: mechanisms and adaptations for long-range communication. *Animal behaviour*, 62: 825-837.
- Pérez-Granados, C. (2023). BirdNET: applications, performance, pitfalls and future opportunities. *Ibis*. DOI: 10.1111/ibi.13193
- Pérez-Granados, C. & Traba, J. (2021). Estimating bird density using passive acoustic monitoring: a review of methods and suggestions for further research. *Ibis*, 163: 765-783.
- Pérez-Granados, C., Bota, G., Giralt, D., Albarracín, J. & Traba, J. (2019). Cost-effectiveness assessment of five audio recording systems for wildlife monitoring: differences between recording distances and singing direction. *Ardeola*, 66: 311-325.
- R Core Team (2019). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <http://www.R-project.org>.
- Rempel, R.S., Francis, C.M., Robinson, J.N. & Campbell, M. (2013). Comparison of audio recording system performance for detecting and monitoring songbirds. *Journal of Field Ornithology*, 84: 86-97.
- Toenies, M. & Rich, L.N. (2021). Advancing bird survey efforts through novel recorder technology and automated species identification. *California Fish and Wildlife*, 107: 56-70.
- Wood, C.M., Kahl, S., Chaon, P., Peery, M.Z. & Klinck, H. (2021). Survey coverage, recording duration and community composition affect observed species richness in passive acoustic surveys. *Methods in Ecology and Evolution*, 12: 885-896.
- Wood, C.M., Kahl, S., Rahaman, A. & Klinck, H. (2022). The machine learning-powered BirdNET App reduces barriers to global bird research by enabling citizen science participation. *Plos Biology*, 20: e3001670.
- Yip, D., Leston, L., Bayne, E., Sólymos, P. & Grover, A. (2017). Experimentally derived detection distances from audio recordings and human observers enable integrated analysis of point count data. *Avian Conservation and Ecology*, 12: 11.
- Zhao, Z., Zhang, S.H., Xu, Z.Y., Bellisario, K., Dai, N.H., Omrani, H. & Pijanowski, B.C. (2017). Automated bird acoustic event detection and robust species classification. *Ecological Informatics*, 39: 99-108.

SUPPLEMENTARY ELECTRONIC MATERIAL

Additional supporting information may be found in the online version of this paper. See volume 70(2) on www.ardeola.org

[Información adicional sobre este artículo en su versión en línea en www.ardeola.org, volumen 70(2).]

Table S1. Number of vocalisations mislabelled by BirdNET. Results are shown separately per distance, recorder and species. The Hooded Warbler is not included since there were no mislabelled detections for that species.

[Número de vocalizaciones mal etiquetadas por BirdNET. Los resultados se muestran por separado por distancia, grabadora y especie. No se incluye la reinita encapuchada, ya que no hubo detecciones mal etiquetadas para esa especie.]

Table S2. Mean qualitative score provided by BirdNET for those vocalisations correctly identified. Results are shown separately per distance, recorder and species. * No vocalisations detected.

[Puntuación cualitativa media proporcionada por BirdNET para las vocalizaciones correctamente identificadas. Los resultados se muestran por separado por distancia, grabadora y especie. * No se detectaron vocalizaciones.]

Image S1. Semi-open habitat type monitored. The field tests were carried out inside the pine

copse, which has a limited tree density and no shrub or grass cover.

[Tipo de hábitat semiabierto supervisado. Las pruebas de campo se realizaron en el interior del pinar, que tiene una densidad arbórea limitada y carece de cubierta arbustiva o herbácea.]

Image S2. Setup of the playback experiment.

The four recorders were oriented towards the playback and with no obstacles blocking sound between them.

[Montaje del experimento de reproducción sonora. Las cuatro grabadoras estaban orientadas hacia la reproducción y sin obstáculos que bloquearan el sonido entre ellas.]

Audio S1. The link provides the playback employed during the field tests. The recording comprises ten vocalisations of each of the three target species.

[El enlace proporciona la reproducción sonora empleada durante las pruebas de campo. La grabación comprende diez vocalizaciones de cada una de las tres especies objetivo.]

Submitted: February 02, 2023

Major revision: March 16, 2023

Accepted: March 31, 2023

Editor: Jesús Avilés

SUPPLEMENTARY ELECTRONIC MATERIAL

ARDEOLA, 70(2)

A FIRST ASSESSMENT OF BIRDNET PERFORMANCE AT VARYING DISTANCES: A PLAYBACK EXPERIMENT

UN PRIMER ANÁLISIS DEL RENDIMIENTO DE BIRDNET A
DISTANCIAS VARIABLES: UN EXPERIMENTO CON *PLAYBACK*

Cristian PÉREZ-GRANADOS^{1,2*}

¹ Ecology Department, Universidad de Alicante, 03080 Alicante, Spain.

² Biodiversity and Animal Conservation Lab, Landscape Dynamics and Biodiversity Programme, Forest Science and Technology Center of Catalonia (CTFC), Solsona, Catalonia, Spain.

* Corresponding author: cristian.perez@ua.es

TABLE S1

Number of vocalisations mislabelled by BirdNET. Results are shown separately per distance, recorder and species. The Hooded Warbler is not included since there were no mislabelled detections for that species.

[Número de vocalizaciones mal etiquetadas por BirdNET. Los resultados se muestran por separado por distancia, grabadora y especie. No se incluye la reinita encapuchada, ya que no hubo detecciones mal etiquetadas para esa especie.]

Recorder	Species	10 m	25 m	50 m	75 m	100 m	125 m	150 m	TOTAL
	Grasshopper Sparrow	0	0	0	0	0	1	0	1
Song Meter Micro A	Gray Vireo	1	3	2	1	0	3	1	11
	TOTAL	1	3	2	1	0	4	1	12
	Grasshopper Sparrow	0	0	1	1	0	1	0	3
Song Meter Micro B	Gray Vireo	0	0	0	1	0	1	0	2
	TOTAL	0	0	1	2	0	2	0	5
	Grasshopper Sparrow	0	0	0	0	0	0	0	0
Song Meter Mini A	Gray Vireo	0	0	3	1	0	0	0	4
	TOTAL	0	0	3	1	0	0	0	4
	Grasshopper Sparrow	0	0	0	0	1	0	0	1
Song Meter Mini B	Gray Vireo	1	1	3	1	0	2	1	9
	TOTAL	1	1	3	1	1	2	1	10

TABLE S2

Mean qualitative score provided by BirdNET for those vocalisations correctly identified. Results are shown separately per distance, recorder and species. * No vocalisations detected.

*[Puntuación cualitativa media proporcionada por BirdNET para las vocalizaciones correctamente identificadas. Los resultados se muestran por separado por distancia, grabadora y especie. * No se detectaron vocalizaciones.]*

Recorder	Species	10 m	25 m	50 m	75 m	100 m	125 m	150 m	TOTAL
Song Meter Micro A	Grasshopper Sparrow	1	1	0.92	0.96	*	*	*	0.97
	Gray Vireo	1	1	0.95	1	0.92	0.41	*	0.96
	Hooded Warbler	1	0.98	1	1	1	0.91	0.59	0.97
	TOTAL	1	0.99	0.95	0.98	0.98	0.85	0.59	0.97
Song Meter Micro B	Grasshopper Sparrow	1	1	1	1	*	*	*	1
	Gray Vireo	1	1	1	1	1	*	*	1
	Hooded Warbler	0.91	1	1	0.96	0.91	1	1	0.96
	TOTAL	0.97	1	1	0.98	0.93	1	1	0.98
Song Meter Mini A	Grasshopper Sparrow	1	1	1	1	0.87	0.63	*	0.96
	Gray Vireo	0.99	0.97	0.75	0.92	1	1	1	0.93
	Hooded Warbler	0.85	1	1	0.84	0.96	1	*	0.93
	TOTAL	0.95	0.99	0.94	0.93	0.91	0.78	1	0.94
Song Meter Mini B	Grasshopper Sparrow	1	1	1	1	0.89	*	*	0.98
	Gray Vireo	1	0.85	0.93	1	1	0.61	1	0.92
	Hooded Warbler	1	1	0.97	0.98	0.91	0.67	1	0.96
	TOTAL	1	0.95	0.97	0.99	0.92	0.64	1	0.96

IMAGE S1

Semi-open habitat type monitored. The field tests were carried out inside the pine copse, which has a limited tree density and no shrub or grass cover.

[Tipo de hábitat semiabierto supervisado. Las pruebas de campo se realizaron en el interior del pinar, que tiene una densidad arbórea limitada y carece de cubierta arbustiva o herbácea.]



IMAGE S2

Setup of the playback experiment. The four recorders were oriented towards the playback and with no obstacles blocking sound between them.

[Montaje del experimento de reproducción. Las cuatro grabadoras estaban orientadas hacia la reproducción y sin obstáculos que bloquearan el sonido entre ellas.]



AUDIO S1

The link provides the playback employed during the field tests. The recording comprises ten vocalisations of each of the three target species.

[El enlace proporciona la reproducción sonora empleada durante las pruebas de campo. La grabación comprende diez vocalizaciones de cada una de las tres especies objetivo.]

<https://figshare.com/s/998fb4813dac382c9090>