3-8-2018

Bat detective—Deep learning tools for bat acoustic signal detection

Oisin Mac Aodha
University College London

Rory Gibb
University College London

Kate E. Barlow
Bat Conservation Trust

Michael Firman
University College London

Robin Freeman
Zoological Society of London

See next page for additional authors

Recommended Citation

Mac Aodha O, Gibb R, Barlow KE, BrowningE, FirmanM, FreemanR, et al. (2018) Bat detective—Deep learning tools for bat acoustic signal detection. PLoS Comput Biol 14(3): e1005995. https://doi.org/10.1371/journal. pcbi.1005995

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

Bat detective—Deep learning tools for bat acoustic signal detection

Oisin Mac Aodha[1]*, Rory Gibb[2], Kate E. Barlow[3], Ella Browning[2,4], Michael Firman[1], Robin Freeman[4], Briana Harder[5], Libby Kinsey[1], Gary R. Mead[6], Stuart E. Newson[7], Ivan Pandourski[8], Stuart Parsons[9], Jon Russ[10], Abigel Szodoray-Paradi[11], Farkas Szodoray-Paradi[11], Elena Tilova[12], Mark Girolami[13], Gabriel Brostow[1], Kate E. Jones[2,4]*

1 Department of Computer Science, University College London, London, United Kingdom, **2** Centre for Biodiversity and Environment Research, Department of Genetics, Evolution and Environment, University College London, London, United Kingdom, **3** Bat Conservation Trust, Quadrant House, London, United Kingdom, **4** Institute of Zoology, Zoological Society of London, Regent's Park, London, United Kingdom, **5** Bellevue, Washington, United States of America, **6** Wickford, Essex, United Kingdom, **7** British Trust for Ornithology, The Nunnery, Thetford, Norfolk, United Kingdom, **8** Institute of Biodiversity and Ecosystem Research, Bulgaria Academy of Sciences, Sofia, Bulgaria, **9** School of Earth, Environmental and Biological Sciences, Queensland University of Technology (QUT), Brisbane, QLD, Australia, **10** Ridgeway Ecology, Warwick, United Kingdom, **11** Romanian Bat Protection Association, Satu Mare, Romania, **12** Green Balkans—Stara Zagora, Stara Zagora, Bulgaria, **13** Department of Mathematics, Imperial College London, London, United Kingdom

* o.macaodha@cs.ucl.ac.uk (OMA); kate.e.jones@ucl.ac.uk (KEJ)

OPEN ACCESS

Citation: Mac Aodha O, Gibb R, Barlow KE, Browning E, Firman M, Freeman R, et al. (2018) Bat detective—Deep learning tools for bat acoustic signal detection. PLoS Comput Biol 14(3): e1005995. https://doi.org/10.1371/journal.pcbi.1005995

Editor: Brock Fenton, University of Western Ontario, CANADA

Received: August 9, 2017

Accepted: January 21, 2018

Published: March 8, 2018

Data Availability Statement: All training and test data, including user and expert annotations, along with the code to train and evaluate our detection algorithms are available on our GitHub page (https://github.com/macaodha/batdetect).

Funding: This work was supported financially through the Darwin Initiative (Awards 15003, 161333, EIDPR075), the Zooniverse, the People's Trust for Endangered Species, Mammals Trust UK, the Leverhulme Trust (Philip Leverhulme Prize for KEJ), NERC (NE/P016677/1), and EPSRC (EP/

Abstract

Passive acoustic sensing has emerged as a powerful tool for quantifying anthropogenic impacts on biodiversity, especially for echolocating bat species. To better assess bat population trends there is a critical need for accurate, reliable, and open source tools that allow the detection and classification of bat calls in large collections of audio recordings. The majority of existing tools are commercial or have focused on the species classification task, neglecting the important problem of first localizing echolocation calls in audio which is particularly problematic in noisy recordings. We developed a convolutional neural network based open-source pipeline for detecting ultrasonic, full-spectrum, search-phase calls produced by echolocating bats. Our deep learning algorithms were trained on full-spectrum ultrasonic audio collected along road-transects across Europe and labelled by citizen scientists from www.batdetective.org. When compared to other existing algorithms and commercial systems, we show significantly higher detection performance of search-phase echolocation calls with our test sets. As an example application, we ran our detection pipeline on bat monitoring data collected over five years from Jersey (UK), and compared results to a widely-used commercial system. Our detection pipeline can be used for the automatic detection and monitoring of bat populations, and further facilitates their use as indicator species on a large scale. Our proposed pipeline makes only a small number of bat specific design decisions, and with appropriate training data it could be applied to detecting other species in audio. A crucial novelty of our work is showing that with careful, non-trivial, design and implementation considerations, state-of-the-art deep learning methods can be used for accurate and efficient monitoring in audio.

K015664/1 and EP/K503745/1). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Author summary

There is a critical need for robust and accurate tools to scale up biodiversity monitoring and to manage the impact of anthropogenic change. For example, the monitoring of bat species and their population dynamics can act as an important indicator of ecosystem health as they are particularly sensitive to habitat conversion and climate change. In this work we propose a fully automatic and efficient method for detecting bat echolocation calls in noisy audio recordings. We show that our approach is more accurate compared to existing algorithms and other commercial tools. Our method enables us to automatically estimate bat activity from multi-year, large-scale, audio monitoring programmes.

Introduction

There is a critical need for robust and accurate tools to scale up biodiversity monitoring and to manage the impact of anthropogenic change [1, 2]. Modern hardware for passive biodiversity sensing such as camera trapping and audio recording now enables the collection of vast quantities of data relatively inexpensively. In recent years, passive acoustic sensing has emerged as a powerful tool for understanding trends in biodiversity [3–6]. Monitoring of bat species and their population dynamics can act as an important indicator of ecosystem health as they are particularly sensitive to habitat conversion and climate change [7]. Close to 80% of bat species emit ultrasonic pulses, or echolocation calls, to search for prey, avoid obstacles, and to communicate [8]. Acoustic monitoring offers a passive, non-invasive, way to collect data about echolocating bat population dynamics and the occurrence of species, and it is increasingly being used to survey and monitor bat populations [7, 9, 10].

Despite the obvious advantages of passive acoustics for monitoring echolocating bat populations, its widespread use has been hampered by the challenges of robust identification of acoustic signals, generation of meaningful statistical population trends from acoustic activity, and engaging a wide audience to take part in monitoring programmes [11]. Recent developments in statistical methodologies for estimating abundance from acoustic activity [4, 12, 13], and the growth of citizen science networks for bats [9, 10] mean that efficient and robust audio signal processing tools are now a key priority. However, tool development is hampered by a lack of large scale species reference audio datasets, intraspecific variability of bat echolocation signals, and radically different recording devices being used to collect data [11].

To date, most full-spectrum acoustic identification tools for bats have focused on the problem of species classification from search-phase echolocation calls [11]. Existing methods typically extract a set of audio features (such as call duration, mean frequency, and mean amplitude) from high quality search-phase echolocation call reference libraries to train machine learning algorithms to classify unknown calls to species [11, 14–19]. Instead of using manually defined features, another set of approaches attempt to learn representation directly from spectrograms [20, 21]. Localising audio events in time (defined here as 'detection'), is an important challenge in itself, and is often a necessary pre-processing step for species classification [22]. Additionally, understanding how calls are detected is critical to quantifying any biases which may impact estimates of species abundance or occupancy [12, 23]. For example, high levels of background noise, often found in highly disturbed anthropogenic habitats such as cities, may have a significant impact on the ability to detect signals in recordings and lead to a bias in population estimates.

Detecting search-phase calls by manual inspection of spectrograms tends to be subjective, highly dependent on individual experience, and its uncertainties are difficult to quantify [24]. There are a number of automatic detection tools now available which use a variety of methods, including amplitude threshold filtering, locating areas of smooth frequency change, detection of set search criteria, or based on a cross-correlation of signal spectrograms with a reference spectrogram [see review in 11]. While there are some studies that analyse the biases of automated detection (and classification) tools [25–30], this is generally poorly quantified, and in particular, there is very little published data available on the accuracy of many existing closed source commercial systems. Despite this, commercial systems are commonly used in bat acoustic survey and monitoring studies, albeit often with additional manual inspection [9, 10]. This reliance on poorly documented algorithms is scientifically undesirable, and manual detection of signals is clearly not scalable for national or regional survey and monitoring. In addition, there is the danger that manual detection and classification introduces a bias towards the less noisy and therefore more easily identifiable calls. To address these limitations, a freely available, transparent, fast, and accurate detection algorithm that can also be used alongside other classification algorithms is highly desirable.

Here, we develop an open source system for automatic bat search-phase echolocation call detection (i.e. localisation in time) in noisy, real world, recordings. We use the latest developments in machine learning to directly learn features from the input audio data using supervised deep convolutional neural networks (CNNs) [31]. CNNs have been shown to be very successful for classification and detection of objects in images [32, 33]. They have also been applied to various audio classification tasks [34–36], along with human speech recognition [37, 38]. Although CNNs are now starting to be used for bioacoustic signal detection and classification tasks in theoretical or small-scale contexts (e.g. bird call detection) [39], to date there have been no application of CNN-based tools for bat monitoring. This is mainly due to a lack of sufficiently large labelled bat audio datasets for use as training data. To overcome this, we use data collected and annotated by thousands of citizen scientists as part of our Indicator Bats Programme [7] and Bat Detective (www.batdetective.org). We validate our system on three different challenging test datasets from Europe which represent realistic use cases for bat surveys and monitoring programmes, and we present an example real-world application of our system on five years of monitoring data collected in Jersey (UK).

Materials and methods

Acoustic detection pipeline

We created a detection system to determine the temporal location of any search-phase bat echolocation calls present in ultrasonic audio recordings. Our detection pipeline consisted of four main steps (Fig 1) as follows: (1) *Fast Fourier Transform Analysis*—Raw audio (Fig 1A) was converted into a log magnitude spectrogram (FFT window size 2.3 milliseconds, overlap of 75%, with Hanning window), retaining the frequency bands between 5kHz and 135kHz (Fig 1B). Recordings with a sampling rate of 44.1kHz, time expansion factor of 10, and 2.3ms FFT window, resulted in a window size of 1,024 samples. We used spectrograms rather than raw audio for analysis, as it provides an efficient means of dealing with audio that has been recorded at different sampling rates. Provided the frequency and time bins of the spectrogram are of the same resolution, audio with different sampling rates can be input into the same network. (2) *De-noising*–We used the de-noising method of [40] to filter out background noise by removing the mean amplitude in each frequency band (Fig 1C), as this significantly improved performance. (3) *Convolutional Neural Network Detection*–We created a convolutional neural network (CNN) that poses search-phase bat echolocation call detection as a

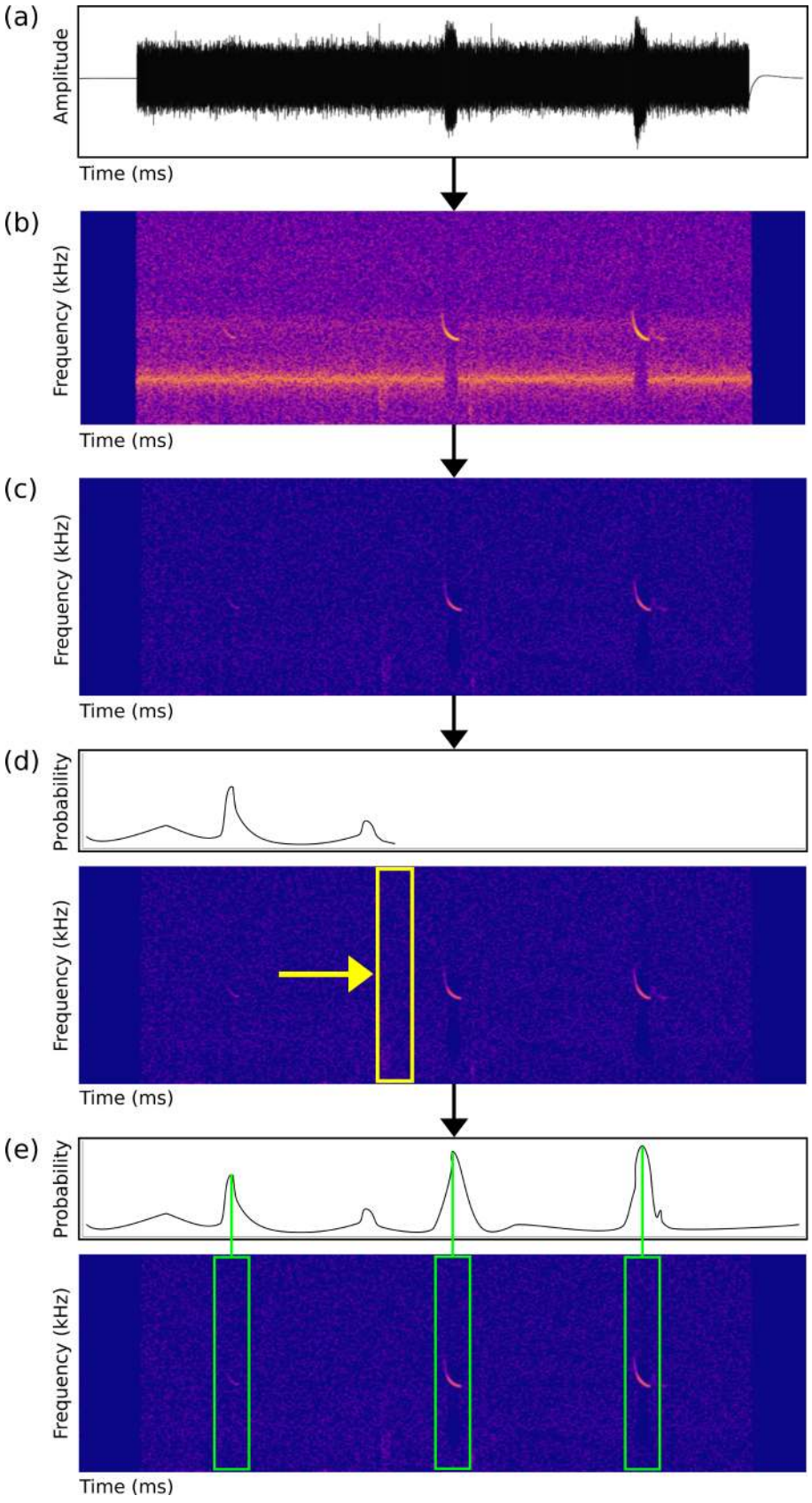
(a)
Amplitude
Time (ms)
(b)
Frequency (kHz)
Time (ms)
(c)
Frequency (kHz)
Time (ms)
(d)
Probability
Frequency (kHz)
Time (ms)
(e)
Probability
Frequency (kHz)
Time (ms)

Fig 1. Detection pipeline for search-phase bat echolocation calls. (a) Raw audio files are converted into a spectrogram using a Fast Fourier Transform (b). Files are de-noised (c), and a sliding window Convolutional Neural Network (CNN) classifier (d, yellow box) produces a probability for each time step. Individual call detection probabilities using non-maximum suppression are produced (e, green boxes), and the time in file of each prediction along with the classifier probability are exported as text files.

https://doi.org/10.1371/journal.pcbi.1005995.g001

binary classification problem. Our CNN_{FULL} consisted of three convolution and max pooling layers, followed by one fully connected layer (see Supplementary Information Methods for further details). We halved the size of the input spectrogram to reduce the input dimensionality to the CNN which resulted in an input array of size of 130 frequency bins by 20 time steps, corresponding to a fixed length, detection window size of 23ms. We applied the CNN in a sliding window fashion, to predict the presence of a search-phase bat call at every instance of time in the spectrogram (Fig 1D). As passive acoustic monitoring can generate large quantities of data, we required a detection algorithm that would run faster than real time. While CNNs produce state of the art results for many tasks, naïve application of them for detection problems at test time can be extremely computationally inefficient [33]. So, to increase the speed of our system we also created a second, smaller CNN which included fewer model weights that can be run in a fully convolutional manner (CNN_{FAST}) (Supplementary Information Methods, Supplementary Information S1 Fig). (4) *Call Detection Probabilities*–The probabilistic predictions produced by the sliding window detector tended to be overly smooth in time (Fig 1D). To localise the calls precisely, we converted the probabilistic predictions into individual detections using a non-maximum suppression to return the local maximum for each peak in the output prediction (Fig 1E). These local maxima corresponded to the predicted locations of the start of each search-phase bat echolocation call, with associated probabilities, and were exported as text files.

Acoustic training datasets

We trained our BatDetect CNNs using a subset of full-spectrum time-expanded (TE) ultrasonic acoustic data recorded between 2005–2011 along road-transects by citizen scientists as part of the Indicator Bats Programme (iBats) [7] (see Supplementary Information Methods for detailed data collection protocols). During surveys, acoustic devices (Tranquility Transect, Courtplan Design Ltd, UK) were set to record using a TE factor of 10, a sampling time of 320ms, and sensitivity set on maximum, giving a continuous sequence of 'snapshots', consisting of 320ms of silence (sensor listening) and 3.2s of TE audio (sensor playing back x 10). As sensitivity was set at maximum, and no minimum amplitude trigger mechanism was used on the recording devices, our recorded audio data contained many instances of low amplitude and faint bat calls, as well as other night-time 'background' noises such as other biotic, abiotic, and anthropogenic sounds.

We generated annotations of the start time of search-phase bat echolocation calls in the acoustic recordings by uploading the acoustic data to the Zooniverse citizen science platform (www.zooniverse.org) as part of the Bat Detective project (www.batdetective.org), to enable public users to view and annotate them. The audio data were first split up into 3.84s long sound clips to include the 3.2s of TE audio and buffered by sensor-listening silence on either side. We then uploaded each sound clip as both a wav file and a magnitude spectrogram image (represented as a 512x720 resolution image) onto the Bat Detective project website. As the original recordings were time-expanded, therefore reducing the frequency, sounds in the files were in the audible spectrum and could be easily heard by users. Users were presented with a spectrogram and its corresponding audio file, and asked to annotate the presence of bat calls in each 3.84s clip (corresponding to 320ms of real-time recordings) (Supplementary

Information S2 Fig). After an initial tutorial (S1 Video), users were instructed to draw bounding boxes around the locations of bat calls within call sequences and to annotate them as being either: (1) search-phase echolocation calls; (2) terminal feeding buzzes; or (3) social calls. Users were also encouraged to annotate the presence of insect vocalisations and non-biotic mechanical noises.

Between Oct 2012 and Sept 2016, 2,786 users (including only the number of users which had registered with the site and performed more than five annotations) listened to 127,451 unique clips and made 605,907 annotations. 14,339 of these clips were labelled as containing a bat call, with 10,272 identified as containing search-phase echolocation calls. Due to the inherent difficulty of identifying bat calls and the inexperience of some of our users, we observed a large number of errors in the annotations provided. How to best merge different annotations for multiple users is an open research question. Instead, we visually inspected a subset of the annotations from our most active user and found that they produced high quality annotations. This top user had viewed 46,508 unique sound clips and had labelled 3,364 clips as containing bat search-phase echolocation calls (a representative sample is shown in Supplementary Information S3 Fig). From this we randomly selected a training set of 2,812 clips, consisting of 4,782 individual search-phase echolocation call annotations from Romania and Bulgaria, with which to train the CNNs (corresponding to data from 347 road-transect sampling events of 137 different transects collected between 2006 and 2011) (Fig 2A). Data were chosen from these countries as they contain the majority of the most commonly occurring bat species in Europe [41]. This training set was used for all experiments. The remaining annotated clips from the same user were used to create one of our test sets, iBats Romania and Bulgaria (Fig 2A and see below). Occasionally, call harmonics and the associated main call were sometimes labelled with different start times in the same audio clip. To address this problem, we automatically merged annotations that occurred within 6 milliseconds of each other, making the assumption that they belonged to the same call. We measured the top user's annotation accuracy on the test set from Romania and Bulgaria compared to the expert curated ground truth. This resulted in an average precision of 0.845 (computed from 455 out of 500 test files this user had labelled). This is in contrast with the second most prolific annotator who had an average precision of 0.67 (based on 311 out of 500 files).

Acoustic testing datasets and evaluation

To evaluate the performance of the detection algorithms, we created three different test datasets of approximately the same size (number and length of clips) (Fig 2A and 2B, Supplementary Information S1 Table). These datasets were chosen to represent three different realistic use cases commonly used for bat surveys and monitoring programmes and included data collected both along road-transects (resulting in noisier audio), and using static ultrasonic detectors. The test sets were as follows: (1) *iBats Romania and Bulgaria*—audio recorded from the same region, by the same individuals, with the same equipment, and sampling protocols as the training set, corresponding to 161 sampling events of 81 different transect routes; (2) *iBats UK*—audio recorded from a different region (corresponding to data from 176 sampling events of 111 different transects recorded between 2005–2011 in the United Kingdom, chosen randomly), by different individuals, using the same equipment type, and identical sampling protocols as part of the iBats programme [7] as the training set; and (3) *Norfolk Bat Survey*—audio recorded from a different region (Norfolk, UK), by different individuals, using different equipment types (SM2BAT+ Song Meter, Wildlife Acoustics) and different protocols (static devices from random sampling locations) as part of the Norfolk Bat Survey [9] in 2015. These data corresponded to 381 sampling events from 246 static recording locations (1km^2 grid cells),

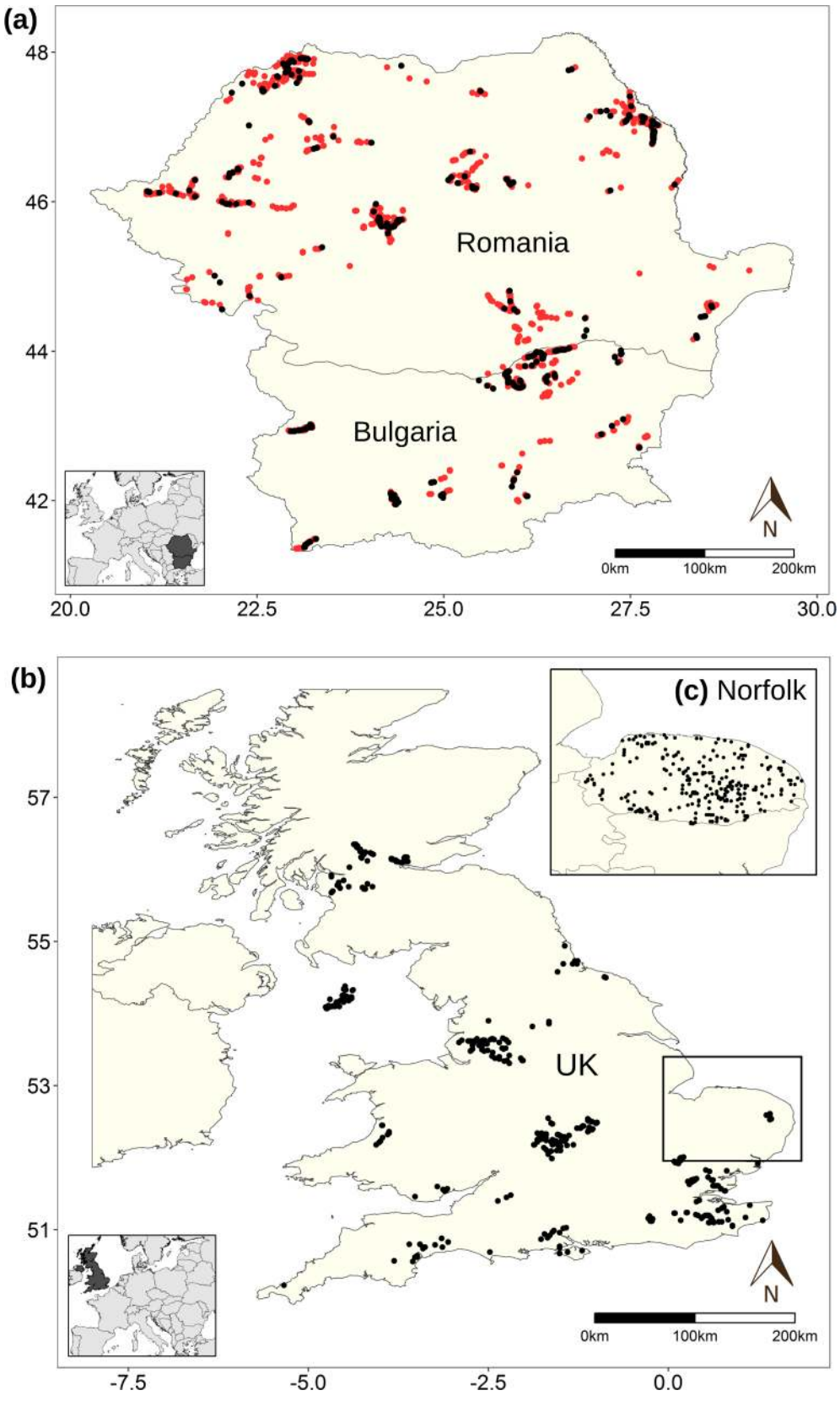
(a)
48
46
44
42
Romania
Bulgaria
N
0km
100km
200km
20.0
22.5
25.0
27.5
30.0
(b)
(c) Norfolk
57
55
53
51
UK
N
0km
100km
200km
-7.5
-5.0
-2.5
0.0

Fig 2. Spatial distribution of the BatDetect CNNs training and testing datasets. (a) Location of training data for all experiments and one test dataset in Romania and Bulgaria (2006–2011) from time-expanded (TE) data recorded along road transects by the Indicator Bats Programme (iBats) [7], where red and black points represent training and test data, respectively. (b) Locations of additional test datasets from TE data recorded as part of iBats car transects in the UK (2005–2011), and from real-time recordings from static recorders from the Norfolk Bat Survey from 2015 (inset). Points represent the start location of each snapshot recording for each iBats transect or locations of static detectors for the Norfolk Bat Survey.

https://doi.org/10.1371/journal.pcbi.1005995.g002

randomly chosen. The start times of the search-phase echolocation calls in these three test sets were manually extracted. For ambiguous calls, we consulted two experts, each with over 10 years of experience with bat acoustics.

As these data contained a significantly greater proportion of negative (non-bat calls) as compared to positive examples (bat calls), standard error metrics used for classification such as overall accuracy were not suitable for evaluating detection. Instead, we report the interpolated average precision and recall of each method displayed as a precision-recall curve [42]. Precision was calculated as the number of true positives divided by the sum of both true and false positives. We consider a detection to be a true positive if it occurred within 10ms of the expert annotated start time of the search-phase echolocation call. Recall was measured as the overall fraction of calls that were present in the audio that were correctly detected. Curves were obtained by thresholding the detection probabilities from zero to one and recording the precision and recall at each threshold. Algorithms that did not produce a continuous output were represented as a single point on the precision-recall curves. We also report recall at 0.95 precision, a metric that measures the fraction of calls that were detected while accepting a false positive rate of 5%. Thus a detection algorithm gets a score of zero if it was not capable of retrieving any calls with a precision greater than 0.95.

We compared the performance of both BatDetect CNNs to three existing closed-source commercial detection systems: (1) SonoBat (version 3.1.7p) [43]; (2) SCAN'R version 1.7.7. [44]; and (3) Kaleidoscope (version 4.2.0 alpha4) [45]. For SonoBat, calls were extracted in batch mode. We set a maximum of 100 calls per file (there are never more than 20 search-phase calls in a test file), and set 'acceptable call quality' and 'skip calls below this quality' parameters both to zero, and used an auto filter of 5KHz. For SCAN'R, we used standard settings as follows: setting minimum and maximum frequency cut off at 10 kHz and 125 kHz, respectively; minimum call duration at 0.5 ms; and minimum trigger level of 10 dB. We used Kaleidoscope in batch mode, setting 'frequency range' to 15-120kHz, 'duration range' to 0-500ms, 'maximum inter-syllable' to 0ms, and 'minimum number of pulses' to 0. We also compared two other detection algorithms that we implemented ourselves, which are representative of typical approaches used for detection in audio files and in other bat acoustic classification studies: (4) Segmentation—an amplitude thresholding segmentation method [46], this is related to the approach of [47]; and (5) Random Forest–a random forest-based classifier [48]. Where relevant, the algorithms for (4) and (5) used the same processing steps as the BatDetect CNNs. For the Segmentation method, we thresholded the amplitude of the input spectrogram resulting in a binary segmentation. Regions that were greater than the threshold S_t, and bigger than size S_r, were considered as positive instances. We chose the values of S_t and S_r on the iBats (Romania and Bulgaria) test dataset that gave the best test results to quantify its best case performance. For the *Random Forest* algorithm, as opposed to extracting low dimensional audio features we instead we used the raw amplitude values from the gradient magnitude of the log magnitude spectrogram as a higher dimensional candidate feature set. This enabled the Random Forest to learn features that it deemed useful for detecting calls. We compared the total processing time for each of our own algorithms, and timings were calculated on a desktop

with an Intel i7 processor, 32Gb of RAM, and a Nvidia GTX 1080 GPU. With the exception of the BatDetect CNN_{FULL}, which used a GPU at test time, all the other algorithms were run on the CPU.

Ecological monitoring application

To demonstrate the performance of our method in a large-scale ecological monitoring application, we compared the number of bat search-phase echolocation calls found using our BatDetect CNN_{FAST} algorithm to those produced from a commonly used commercial package using SonoBat (version 3.1.7p) [43] as a baseline, using monitoring data collected in iBats programme in Jersey, UK from 2011–2015. Audio data was collected twice yearly (July and August) from 11 road-transect routes of approximately 40km by volunteers using the iBats protocols (Supplementary Information, Supplementary Methods), corresponding to 5.7 days of continuous TE audio over five years (or 13.75 hours of real-time data). For the BatDetect CNN_{FAST} analysis, we ran the pipeline as described above, using a conservative probabilistic threshold of 0.90 (so as to only include high precision predictions). Computational analysis timings for the CNN_{FAST} for this dataset were calculated as before. For the comparison to SonoBat, we used the results from an existing real-world analysis in a recent monitoring report [49], where the audio files were first split into 1 min recordings, and then SonoBat was used to detect search-phase calls and to fit a frequency-time trend line to the shape of the call [49]. All fitted lines were visually inspected and calls where the fitted line included background noise or echoes, were rejected. Typically, monitoring analyses group individual calls into sequences (a bat pass) before analysis. To replicate that here in both analyses, individual calls were assumed to be part of the same call sequence (bat pass) if they occurred within the same 3.84s sound clip and if the sequence continued into subsequent sound clips. We compared number of bat calls and passes detected per transect sampling event across the two analyses methods using generalized linear mixed models (GLMM) using lme4 [50] in R v. 3.3.3 [51] in order to control for the spatial and temporal non-independence of our survey data (Poisson GLMM including analysis method as a fixed effect and sampling event, transect route and date as random effects).

Results

Acoustic detection performance

Both versions of our BatDetect CNN algorithm outperformed all other algorithms and commercial systems tested, with consistently higher average precision scores and recall rates across the three different test datasets (Table 1, Fig 3A–3C). In particular, the CNNs detected a substantially higher proportion of search-phase calls at 0.95 precision (maximum 5% false positives) (Table 1). All the other algorithms underestimated the number of search-phase echolocation calls in each dataset, except Segmentation, which produced high recall rates but with low precision (a high number of false positives). The CNNs relative improvement compared to other methods was higher on the road transect datasets (iBats Romania & Bulgaria; iBats UK; Table 1, Fig 3A and 3B). Overall the performance of CNN_{FAST} was slightly worse than the larger CNN_{FULL} across all test datasets, with the exception of improved recall at 0.95 precision in the static Norfolk Bat Survey dataset (Fig 3C, Table 1). Precision scores for all commercial systems (SonoBat, SCAN'R and Kaleidoscope) were reasonably good across all test datasets (>0.7) (Fig 3A–3C). However, this was at the expense of recall rates, which were consistently lower than for the CNNs and Random Forest, where the maximum recall rates were 44–60% of known calls detected (Fig 3C). The recall rates fell to a maximum of 25% of known calls for the road transect datasets (Fig 3A and 3B).

Table 1. Average precision and recall results for bat search-phase call detection algorithms across three different test sets iBats Romania and Bulgaria; iBats UK; and Norfolk Bat Survey.

Detection Algorithms							
				BatDetect			
Average Precision	**SonoBat**	**SCAN'R**	**Kaleidoscope**	**Segment**	**Random Forest**	**CNN_{FAST}**	**CNN_{FULL}**
iBats (R&B)	0.265	0.239	0.189	0.299	0.674	0.863	<u>0.895</u>
iBats (UK)	0.200	0.142	0.144	0.324	0.648	0.781	<u>0.866</u>
NBP (Norfolk)	0.473	0.456	0.553	0.506	0.630	0.861	<u>0.882</u>
Recall at 0.95							
iBats (R&B)	0	0.251	0	0	0.568	0.777	<u>0.818</u>
iBats (UK)	0	0	0	0	0.324	0.570	<u>0.670</u>
NBP (Norfolk)	0.184	0.470	0	0	0.049	<u>0.781</u>	0.754

Large numbers indicate better performance. Recall results are reported at 0.95 precision, where zero indicates that the detector algorithm was unable to achieve a precision greater than 0.95 at any recall level. The results for the best performing algorithm are underlined. Details of the test datasets and detection algorithms are given in the text.

https://doi.org/10.1371/journal.pcbi.1005995.t001

The Random Forest baseline performed significantly better than the commercial systems on the two challenging roadside recorded datasets (Fig 3A and 3B). This is a result of the training data and the underlying power of the model. However, unlike our CNNs, the simple tree based model is limited in the complexity of the representations it can learn, which results in worse performance. For the static Norfolk Bat Survey its performance more closely matches that of SonoBat, but with improved recall.

CNN_{FULL}, CNN_{FAST}, Random Forest, and the Segmentation algorithms took 53, 9.5, 11, and 17 seconds respectively, to run the full detection pipeline on the 3.2 minutes of full spectrum iBats Romania and Bulgaria test dataset. Compared to CNN_{FULL} there was therefore a significant decrease in the time required to perform detection using CNN_{FAST}, which was also the fastest of our methods overall. Notably, close to 50% of the CNN runtime was spent generating the spectrograms for detection, making this the most computationally expensive stage in the pipeline.

Ecological monitoring application results

Our BatDetect CNN_{FAST} algorithm detected a significantly higher number of bat echolocation search-phase calls per transect sampling event, across 5 years of road transect data from iBats Jersey, compared to using SonoBat (CNN_{FAST} mean = 107.69, sd = 48.01; SonoBat mean = 64.95, sd = 28.53, Poisson GLMM including sampling event, transect route and date as random effects $p<2e^{-16}$, n = 216) (Fig 4, Supplementary Information S2 Table). The differences between the two methods for bat passes was much smaller per sampling event, although CNN_{FAST} still detected significantly more passes per transect recording (CNN_{FAST} mean = 29.57, sd = 11.26; SonoBat mean = 27.27, sd = 10.85; Poisson GLMM including sampling event, transect route and date as random effects p = 0.00143, n = 216) (Fig 4, Supplementary Information S2 Table). Running only on the CPU, the CNN_{FAST} algorithm took 24 seconds to process one hour of time-expanded audio.

Discussion

The BatDetect deep learning algorithms show a higher detection performance (average precision and recall) for search-phase echolocation calls with the test sets, when compared to other existing algorithms and commercial systems. In particular, our algorithms were better at

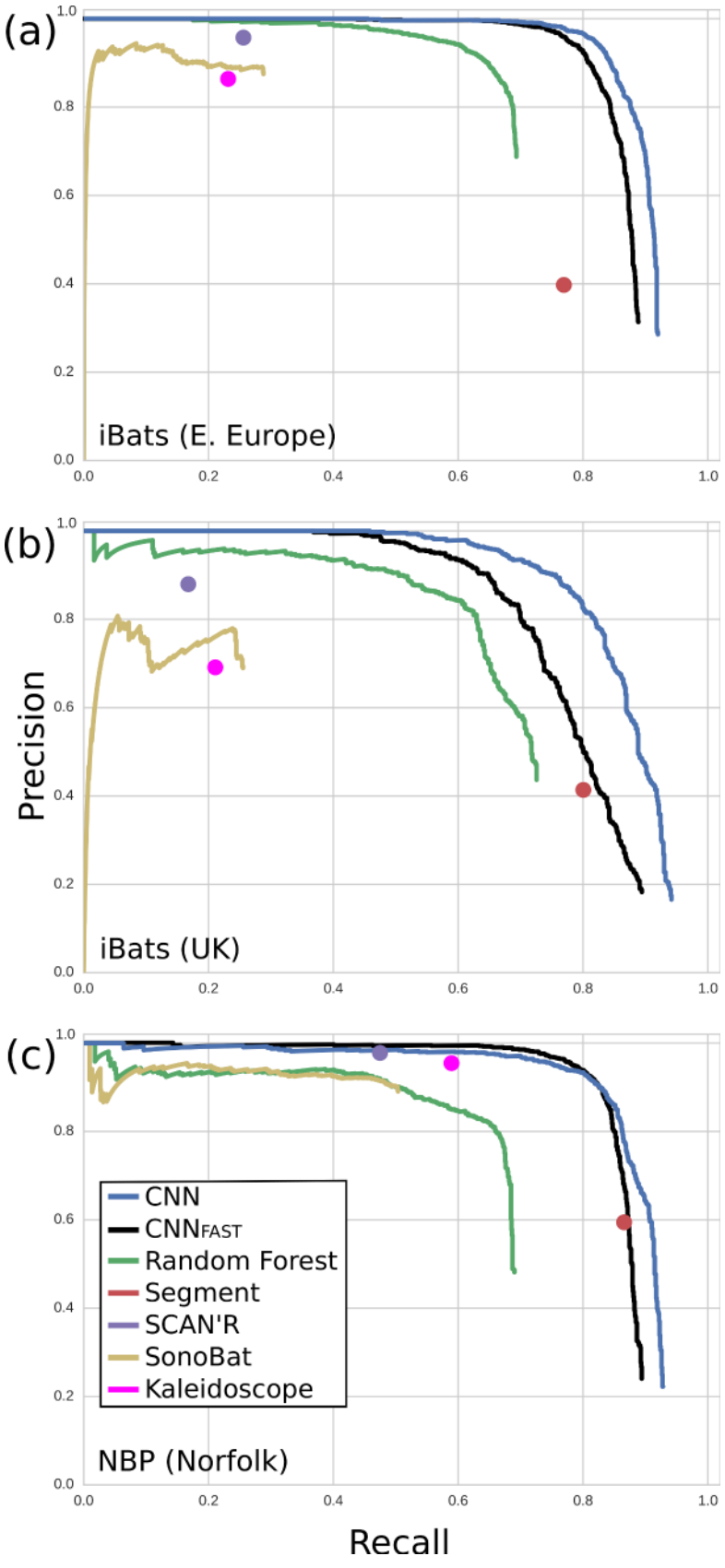
(a)
iBats (E. Europe)
(b)
iBats (UK)
Precision
(c)
CNN
CNNFAST
Random Forest
Segment
SCAN'R
SonoBat
Kaleidoscope
NBP (Norfolk)
Recall

Fig 3. Precision-recall curves for bat search-phase call detection algorithms across three testing datasets; (a) iBats Romania and Bulgaria; (b) iBats UK; and (c) Norfolk Bat Survey. Curves were obtained by sweeping the output probability for a given detector algorithm and computing the precision and recall at each threshold. The commercial systems or algorithms that did not return a continuous output or probability (SCAN'R, Segment, and Kaleidoscope) were depicted as a single point.

https://doi.org/10.1371/journal.pcbi.1005995.g003

detecting calls in road-transect data, which tend to contain noisier recordings, suggesting that these are extremely useful tools for measuring bat abundance and occurrence in such datasets. Road-transect acoustic monitoring is a useful technique to assess bat populations over large areas and programmes have now been established by government and non-government agencies in many different countries [e.g., 7, 52, 53–55]. Noisy sound environments are also likely to be a problem for other acoustic bat monitoring programmes. For example, with the falling cost and wider availability of full-spectrum audio equipment, the range of environments being acoustically monitored is increasing, including noisy urban situations [56, 57]. Individual bats further from the microphone are less likely to be detected as their calls are fainter, and high ambient noise levels increase call masking and decrease call detectability. Additionally, a growth in open-source sensor equipment for bat acoustics using very cheap MEMs microphones [58] may also require algorithms able to detect bats in lower quality recordings, which may have a lower signal to noise ratio or a reduced call band-width due to frequency-dependent loss. Our open-source, well documented algorithms enable biases and errors to be directly incorporated into any acoustic analysis of bat populations and dynamics (e.g. occupancy models [e.g., 23]. The detections with BatDetect can be directly used as input for population monitoring programmes when species identification is difficult such as the tropics, or to other CNN systems to determine bat species identity when sound libraries are available.

Our result that deep learning networks consistently outperformed other baselines, is consistent with the suggestion that CNNs offer substantially improved performance over other supervised learning methods for acoustic signal classification [39]. The major improvement of both CNNs over Random Forest and the three commercial systems was in terms of recall, i.e. increasing the proportion of detected bat calls in the test datasets. Although the precision of the commercial systems was often relatively high, the CNNs were able to detect much fainter and partially noise-masked bat calls that were missed by the other methods, with fewer false positives, and very quickly, particularly with CNN_{FAST}. Previous applications of deep learning networks to bioacoustic and environmental sound recognition have used small and high-quality datasets [e.g., 35, 39]. However, our results show that, provided they are trained with suitably large and varied training data, deep learning networks have good potential for applied use in real-world heterogeneous datasets that are characteristic of acoustic wildlife monitoring (involving considerable variability in both environmental noise and distance of animal from sensor). Our comparison of CNN_{FULL} and CNN_{FAST} detectors was favourable, although CNN_{FAST} had a slightly poorer performance showing a trade-off between speed and accuracy. This suggests that CNN_{FAST} could potentially be adapted to work well with on-board low power devices (e.g. Intel's Edison device) to deliver real-time detections. Avoiding the spectrogram generation stage entirely and using the raw audio samples as input [59], could also speed up performance of the system in the future, as currently over 50% of the CNN test time is taken up by computing spectrograms.

While our results have been validated on European bats, no species or region-specific knowledge, or particular acoustic sensor system is directly encoded into our system, making it possible to easily generalise to other systems (e.g. frequency division recordings), regions and species with additional training data. Despite this flexibility, this version of our deep network may be currently biased towards common species found along roads, although the algorithms

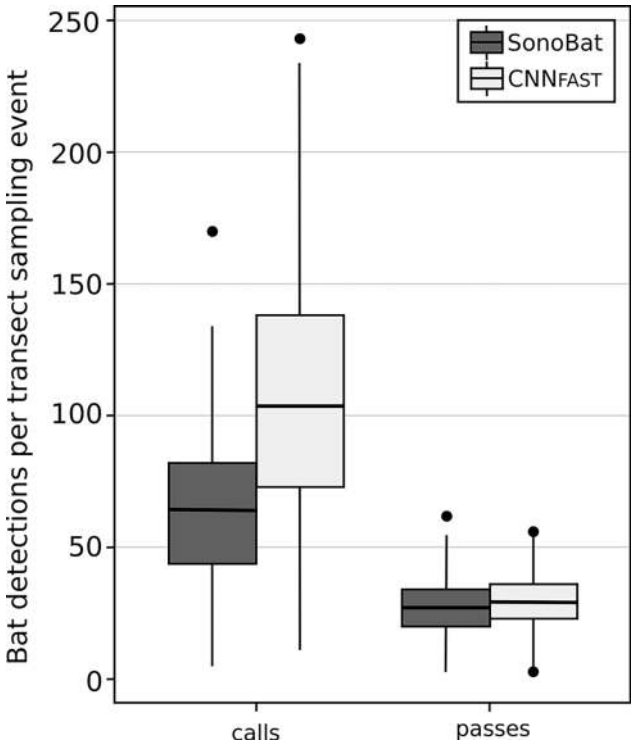

Fig 4. Comparison of the predicted bat detections (calls and passes) for two different acoustic systems using monitoring data collected from Jersey, UK. Acoustic systems used were SonoBat (version 3.1.7p) [43] using analysis in [49], and BatDetect CNN_{FAST} using a probability threshold of 0.90. Detections are shown within each box plot, where the black line represents the mean across all transect sampling events from 2011–2015, boxes represent the middle 50% of the data, whiskers represent variability outside the upper and lower quartiles, with outliers plotted as individual points. See text for definition of a bat pass.

https://doi.org/10.1371/journal.pcbi.1005995.g004

did perform well on static recordings on a range of common and rare species in a range of habitats in the Norfolk Bat Survey [9]. Nevertheless, in future, extending the training dataset to include annotated bat calls from verified species-call databases to increase geographic and taxonomic coverage, will further improve the generality of our detection tool. Other improvements to the CNN detectors could also be made to lessen taxonomic bias. For example, some bat species have search phase calls longer than the fixed input time window of 23ms of both CNNs (e.g. horseshoe bats). This may limit our ability currently to detect species with these types of calls. One future approach would be to resize the input window [33], thus discarding some temporal information, or to use some form of recurrent neural network such as a Long Short-Term Memory (LSTM) [60] that can take a variable length sequence as input. There are many more unused annotations in the Bat Detective dataset that could potentially increase our training set size. However, we found some variability in the quality of the citizen science user annotations, as in other studies [61]. To make best use of these annotations, we need user models for understanding which annotations and users are reliable [62, 63]. The Bat Detective dataset also includes annotations of particular acoustic behaviours (feeding buzzes and social calls), which in future can be used to train detection algorithms for different acoustic behaviours [e.g., 64].

Our evaluation on large-scale ecological monitoring data from Jersey [49], demonstrated that our open-source BatDetect CNN_{FAST} pipeline performs as well or better (controlling for spatial and temporal non-independence) compared with an existing widely-used commercial system (SonoBat) that had been manually filtered (false positives were removed). Here we

assume that the manually filtered data represents the ground truth, although it may slightly underrepresent the true number of calls due to missing detections on the part of SonoBat. Interestingly, although the CNN_{FAST} consistently detected more of the faint and partially-masked calls, most bat passes are likely to still contain at least one call that is clearly-recorded enough to be detected by SonoBat, meaning that the total number of detected bat passes is similar across the two methods. No manual filtering is performed for $CNN_{FAST,}$ but the increase in detected calls mirrors the results observed in Fig 3 at high thresholds (i.e. the left of the curves), where both CNNs detected 10–20% more calls than SonoBat on the related driving transect based test sets. Our system results in a large reduction processing time—several minutes for our automatic approach compared to several days split between automatic processing and manual filtering as reported by the authors of [49]. Specifically, it takes CNN_{FAST} under 10 seconds to process the 500 files in the iBats Romania and Bulgaria test set compared to 30 minutes for SonoBat in batch mode. This increase in performance both in terms of speed and accuracy is crucial for future large scale monitoring programmes.

The results in our monitoring application raises an interesting question—what is the value of the additional detected calls? Fig 4 shows a large increase in the number of detected calls and a slight increase in the number of detected bat passes. It may be the case that our current heuristic for merging calls into passes is too aggressive and as a result under reports the true number of bats when there were multiple calling at the same time. Further improvements to our system may come from a better understanding of the patterns of search-phase calls within sequences [65]. Instead of the existing heuristic we would ideally also be able to learn the relationship between individual calls and passes from labelled training data.

The current generation of algorithms for bat species classification that are based on extracting simple audio features may perhaps not be best suited to make use of the extra calls we detect. However, when large collections of diverse species data become available only relatively minor architectural changes will be required to our detection pipeline to adapt it for species classification (e.g. changing the final layer of our CNNs). As we have already observed for detection, with enough data, representation learning based approaches can also be applied to the problem of species classification with the promise of large increases in accuracy. These extra calls will be invaluable to create more powerful models, enabling them to perform accurately in diverse and challenging real world situations. For some noisy and faint bat calls it may always be difficult to identify them to the species level, and as a result a coarser taxonomic prediction may have to suffice.

Our BatDetect search-phase bat call detector significantly outperforms existing methods for localising the position of bat search-phase calls, particularly in noisy audio data. It could be combined with automatic bat species classification tools to scale up the monitoring of bat populations over large geographic regions. In addition to making our system available open source, we also provide three expertly annotated test sets that can be used to benchmark future detection algorithms.

Data reporting

All training and test data, including user and expert annotations, along with the code to train and evaluate our detection algorithms are available on our GitHub page (https://github.com/macaodha/batdetect).

Supporting information

S1 Text. Supplementary methods. Description of the CNN architectures, training details, and information about how the training data was collected.
(PDF)

S1 Fig. CNN_{FAST} network architecture description. The CNN_{FAST} network consists of two convolution layers (Conv1 and Conv2), with 16 filters each (shown in yellow, with the filter size shown inset). Both convolution layers are followed by a max pooling layer (Max Pool1 and Max Pool2), and the network ends with a fully connected layer with 64 units (Fully Connect). CNN_{FAST} computes feature maps (shown as white boxes) across the entire input spectrogram, resulting in less computation and a much faster run time. The fully connected layer is also evaluated as a convolution. The output of the detector is a probability vector (shown in green) whose length is one quarter times the width of the input spectrogram. The numbers below each layer indicate the height, weight, and depth of the corresponding layer.
(TIF)

S2 Fig. Spectrogram annotation interface from Bat Detective. Boxes represent example user annotations of sounds in a spectrogram of a 3840ms sound clip, showing annotations of two sequences of search-phase echolocation bat calls (blue boxes), and an annotation of an insect call (yellow box).
(TIF)

S3 Fig. Example search-phase bat echolocation calls from iBats Romania & Bulgaria training dataset. Each example is represented as a spectrogram of duration 23 milliseconds and frequency range from 5–115 kHz using the same FFT parameters as the main paper, and contains examples of different search-phase echolocation call type, but also a wide variety of background non-bat biotic, abiotic and anthropogenic sounds.
(TIF)

S1 Table. Description of BatDetect CNNs test datasets. TE represents time-expansion recordings (x10); RT real-time recordings. Note that the length of the clips is approximately comparable for both the iBats and the Norfolk Bat Survey data as the total iBats clip length of 3.84s corresponds to 320ms of ultrasonic sound slowed down ten times (3.2s) and buffered by silence on either side.
(PDF)

S2 Table. Full details of the Poisson Generalised Linear Mixed Model (GLMM) used to model bat detections (calls and passes) for two acoustic analytical systems. β represents slope, Std standard deviation, Z Z-value, p probability. Analytical systems compared were SonoBat (version 3.1.7p) [14] and BatDetect CNN_{FAST}, using a 0.9 probability threshold. Data from using acoustic monitoring data collected from Jersey, UK between 2011–2015. See main text for definition of a bat pass. GLMMs were fitted using lme4 [15] with model formula: *detections ~ analytical_method + (1|sampling_event) + (1|transect) + (1|date)*.
(PDF)

S1 Video. Overview of our system, Bat Detective annotation steps, and sample results.
(MP4)

Acknowledgments

We are enormously grateful for the efforts and enthusiasm of the amazing iBats and Bat Detective volunteers, for the many hours spent collecting data and providing valuable annotations. We would also like to thank Ian Agranat and Joe Szewczak for useful discussions and access to their systems. Finally, we would like to thank Zooniverse for setting up and hosting the Bat Detective project.

Author Contributions

Conceptualization: Oisin Mac Aodha, Kate E. Jones.

Data curation: Oisin Mac Aodha, Rory Gibb, Kate E. Barlow, Ella Browning, Robin Freeman, Briana Harder, Libby Kinsey, Gary R. Mead, Stuart E. Newson, Ivan Pandourski, Jon Russ, Abigel Szodoray-Paradi, Farkas Szodoray-Paradi, Elena Tilova, Kate E. Jones.

Funding acquisition: Mark Girolami, Gabriel Brostow, Kate E. Jones.

Investigation: Kate E. Barlow, Kate E. Jones.

Methodology: Oisin Mac Aodha, Rory Gibb, Ella Browning, Michael Firman, Robin Freeman, Libby Kinsey, Stuart E. Newson, Stuart Parsons, Kate E. Jones.

Project administration: Kate E. Barlow, Abigel Szodoray-Paradi, Farkas Szodoray-Paradi, Elena Tilova, Mark Girolami, Kate E. Jones.

Resources: Mark Girolami, Gabriel Brostow, Kate E. Jones.

Software: Oisin Mac Aodha, Ella Browning, Robin Freeman, Libby Kinsey.

Supervision: Stuart E. Newson, Gabriel Brostow, Kate E. Jones.

Validation: Oisin Mac Aodha, Rory Gibb, Ella Browning, Stuart E. Newson, Jon Russ, Kate E. Jones.

Visualization: Oisin Mac Aodha, Rory Gibb, Kate E. Jones.

Writing – original draft: Oisin Mac Aodha, Kate E. Jones.

Writing – review & editing: Oisin Mac Aodha, Rory Gibb, Michael Firman, Libby Kinsey, Stuart E. Newson, Stuart Parsons, Gabriel Brostow, Kate E. Jones.

References

1. Turner W. Sensing biodiversity. Science. 2014; 346(6207):301. https://doi.org/10.1126/science.1256014 PMID: 25324372
2. Cardinale BJ, Duffy JE, Gonzalez A, Hooper DU, Perrings C, Venail P, et al. Biodiversity loss and its impact on humanity. Nature. 2012; 486(7401):59–67. http://www.nature.com/nature/journal/v486/n7401/abs/nature11148.html#supplementary-information. https://doi.org/10.1038/nature11148 PMID: 22678280
3. Blumstein DT, Mennill DJ, Clemins P, Girod L, Yao K, Patricelli G, et al. Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. Journal of Applied Ecology. 2011; 48(3):758–67. https://doi.org/10.1111/j.1365-2664.2011.01993.x
4. Marques TA, Thomas L, Martin SW, Mellinger DK, Ward JA, Moretti DJ, et al. Estimating animal population density using passive acoustics. Biological Reviews. 2013; 88(2):287–309. https://doi.org/10.1111/brv.12001 PMID: 23190144
5. Penone C, Le Viol I, Pellissier V, Julien J-F, Bas Y, Kerbiriou C. Use of Large-Scale Acoustic Monitoring to Assess Anthropogenic Pressures on Orthoptera Communities. Conservation Biology. 2013; 27(5):979–87. https://doi.org/10.1111/cobi.12083 PMID: 23692213
6. Sueur J, Pavoine S, Hamerlynck O, Duvail S. Rapid Acoustic Survey for Biodiversity Appraisal. PLOS ONE. 2009; 3(12):e4065. https://doi.org/10.1371/journal.pone.0004065 PMID: 19115006
7. Jones KE, Russ JA, Bashta A-T, Bilhari Z, Catto C, Csősz I, et al. Indicator Bats Program: A System for the Global Acoustic Monitoring of Bats. Biodiversity Monitoring and Conservation: Wiley-Blackwell; 2013. p. 211–47.
8. Schnitzler H-U, Moss CF, Denzinger A. From spatial orientation to food acquisition in echolocating bats. Trends in Ecology & Evolution. 2003; 18(8):386–94. http://dx.doi.org/10.1016/S0169-5347(03)00185-X.
9. Newson SE, Evans HE, Gillings S. A novel citizen science approach for large-scale standardised monitoring of bat activity and distribution, evaluated in eastern England. Biological Conservation. 2015; 191:38–49. http://dx.doi.org/10.1016/j.biocon.2015.06.009.

10. Barlow KE, Briggs PA, Haysom KA, Hutson AM, Lechiara NL, Racey PA, et al. Citizen science reveals trends in bat populations: The National Bat Monitoring Programme in Great Britain. Biological Conservation. 2015; 182:14–26. http://dx.doi.org/10.1016/j.biocon.2014.11.022.
11. Walters CL, Collen A, Lucas T, Mroz K, Sayer CA, Jones KE. Challenges of Using Bioacoustics to Globally Monitor Bats. In: Adams RA, Pedersen SC, editors. Bat Evolution, Ecology, and Conservation. New York, NY: Springer New York; 2013. p. 479–99.
12. Lucas TCD, Moorcroft EA, Freeman R, Rowcliffe JM, Jones KE. A generalised random encounter model for estimating animal density with remote sensor data. Methods in Ecology and Evolution. 2015; 6(5):500–9. https://doi.org/10.1111/2041-210X.12346 PMID: 27547297
13. Stevenson BC, Borchers DL, Altwegg R, Swift RJ, Gillespie DM, Measey GJ. A general framework for animal density estimation from acoustic detections across a fixed microphone array. Methods in Ecology and Evolution. 2015; 6(1):38–48. https://doi.org/10.1111/2041-210X.12291
14. Skowronski MD, Harris JG. Acoustic detection and classification of microchiroptera using machine learning: lessons learned from automatic speech recognition. The Journal of the Acoustical Society of America. 2006; 119:1817–33. PMID: 16583922
15. Armitage DW, Ober HK. A comparison of supervised learning techniques in the classification of bat echolocation calls. Ecological Informatics. 2010; 5:465–73.
16. Parsons S, Jones G. Acoustic identification of twelve species of echolocating bat by discriminant function analysis and artificial neural networks. The Journal of Experimental Biology. 2000; 203:2641–56. PMID: 10934005
17. Russo D, Jones G. Identification of twenty-two bat species (Mammalia: Chiroptera) from Italy by analysis of time-expanded recordings of echolocation calls. Journal of Zoology. 2002; 258(01):91–103.
18. Walters CL, Freeman R, Collen A, Dietz C, Brock Fenton M, Jones G, et al. A continental-scale tool for acoustic identification of European bats. Journal of Applied Ecology. 2012; 49(5):1064–74. https://doi.org/10.1111/j.1365-2664.2012.02182.x
19. Zamora-Gutierrez V, Lopez-Gonzalez C, MacSwiney Gonzalez MC, Fenton B, Jones G, Kalko EKV, et al. Acoustic identification of Mexican bats based on taxonomic and ecological constraints on call design. Methods in Ecology and Evolution. 2016; 7(9):1082–91. https://doi.org/10.1111/2041-210X.12556
20. Stathopoulos V, Zamora-Gutierrez V, Jones KE, Girolami M. Bat echolocation call identification for biodiversity monitoring: A probabilistic approach. Journal of the Royal Statistical Society Series C: Applied Statistics. 2017.
21. Stowell D, Plumbley MD. Automatic large-scale classification of bird sounds is strongly improved by unsupervised feature learning. PeerJ. 2014; 2:e488. https://doi.org/10.7717/peerj.488 PMID: 25083350
22. Stowell D, Wood M, Stylianou Y, Glotin H, editors. Bird detection in audio: a survey and a challenge. Machine Learning for Signal Processing (MLSP), 2016 IEEE 26th International Workshop on; 2016: IEEE.
23. Clement MJ, Rodhouse TJ, Ormsbee PC, Szewczak JM, Nichols JD. Accounting for false-positive acoustic detections of bats using occupancy models. Journal of Applied Ecology. 2014; 51(5):1460–7. https://doi.org/10.1111/1365-2664.12303
24. Skowronski MD, Fenton MB. Model-based detection of synthetic bat echolocation calls using an energy threshold detector for initialization. The Journal of the Acoustical Society of America. 2008; 123:2643–50. https://doi.org/10.1121/1.2896752 PMID: 18529184
25. Adams AM, Jantzen MK, Hamilton RM, Fenton MB. Do you hear what I hear? Implications of detector selection for acoustic monitoring of bats. Methods in Ecology and Evolution. 2012; 3(6):992–8.
26. Jennings N, Parsons S, Pocock M. Human vs. machine: identification of bat species from their echolocation calls by humans and by artificial neural networks. Canadian Journal of Zoology. 2008; 86 (5):371–7.
27. Clement MJ, Murray KL, Solick DI, Gruver JC. The effect of call libraries and acoustic filters on the identification of bat echolocation. Ecology and evolution. 2014; 4(17):3482–93. https://doi.org/10.1002/ece3.1201 PMID: 25535563
28. Fritsch G, Bruckner A. Operator bias in software-aided bat call identification. Ecology and evolution. 2014; 4(13):2703–13. https://doi.org/10.1002/ece3.1122 PMID: 25077021
29. Russo D, Voigt CC. The use of automated identification of bat echolocation calls in acoustic monitoring: A cautionary note for a sound analysis. Ecological Indicators. 2016; 66:598–602.
30. Rydell J, Nyman S, Eklöf J, Jones G, Russo D. Testing the performances of automated identification of bat echolocation calls: A request for prudence. Ecological Indicators. 2017; 78:416–20.
31. LeCun Y, Bottou L, Bengio Y, Haffner P. Gradient-based learning applied to document recognition. Proceedings of the IEEE. 1998; 86(11):2278–324.

32. Krizhevsky A, Sutskever I, Hinton GE, editors. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems; 2012.
33. Girshick R, Donahue J, Darrell T, Malik J, editors. Rich feature hierarchies for accurate object detection and semantic segmentation. Proceedings of the IEEE conference on computer vision and pattern recognition; 2014.
34. Piczak KJ, Environmental sound classification with convolutional neural networks. 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP); 2015: IEEE.
35. Salamon J, Bello JP. Deep convolutional neural networks and data augmentation for environmental sound classification. arXiv preprint arXiv:160804363. 2016.
36. Hershey S, Chaudhuri S, Ellis DP, Gemmeke JF, Jansen A, Moore RC, et al. CNN Architectures for Large-Scale Audio Classification. arXiv preprint arXiv:160909430. 2016.
37. Hinton G, Deng L, Yu D, Dahl GE, Mohamed A-r, Jaitly N, et al. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. IEEE Signal Processing Magazine. 2012; 29(6):82–97.
38. Hannun A, Case C, Casper J, Catanzaro B, Diamos G, Elsen E, et al. Deep speech: Scaling up end-to-end speech recognition. arXiv preprint arXiv:14125567. 2014.
39. Goeau H, Glotin H, Vellinga W-P, Planque R, Joly A, editors. LifeCLEF Bird Identification Task 2016. The Arrival of Deep Learning. Working Notes of CLEF 2016-Conference and Labs of the Evaluation forum; 2016; Évora, Portugal.
40. Aide TM, Corrada-Bravo C, Campos-Cerqueira M, Milan C, Vega G, Alvarez R. Real-time bioacoustics monitoring and automated species identification. PeerJ. 2013; 1:e103. https://doi.org/10.7717/peerj.103 PMID: 23882441; PubMed Central PMCID: PMCPMC3719130.
41. The IUCN Red List of Threatened Species. Version 2017–1 [Internet]. 2017 [cited Downloaded on 12 May 2017.]. Available from: http://www.iucnredlist.org.
42. Everingham M, Van Gool L, Williams CK, Winn J, Zisserman A. The pascal visual object classes (voc) challenge. International journal of computer vision. 2010; 88(2):303–38.
43. Szewczak JM. Sonobat 2010.
44. Binary Acoustic Technology. SCAN'R. 2014.
45. Wildlife Acoustics. Kaleidoscope. 2012.
46. Lasseck M, editor Large-scale Identification of Birds in Audio Recordings. CLEF (Working Notes); 2014.
47. Bas Y, Bas D, Julien J-F. Tadarida: A Toolbox for Animal Detection on Acoustic Recordings. Journal of Open Research Software. 2017; 5:6. http://doi.org/10.5334/jors.154
48. Breiman L. Random forests. Machine learning. 2001; 45(1):5–32.
49. Walters CL, Browning E, Jones KE. iBats Jersey Review. London, UK: 2016.
50. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models Using lme4. 2015. 2015; 67(1):48. Epub 2015-10-07. https://doi.org/10.18637/jss.v067.i01
51. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2009.
52. Roche N, Langton S, Aughney T, Russ JM, Marnell F, Lynn D, et al. A car-based monitoring method reveals new information on bat populations and distributions in Ireland. Animal Conservation. 2011; 14:642–51.
53. Whitby MD, Carter TC, Britzke ER, Bergeson SM. Evaluation of Mobile Acoustic Techniques for Bat Population Monitoring. Acta Chiropterologica. 2014; 16:223–30.
54. Loeb SC, Rodhouse TJ, Ellison LE, Lausen CL, Reichard JD, Irvine KM, et al. A plan for the North American Bat Monitoring Program (NABat). General Technical Report SRS-208. Asheville, NC: U.S.: Department of Agriculture Forest Service, Southern Research Station., 2015.
55. Azam C, Le Viol I, Julien J-F, Bas Y, Kerbiriou C. Disentangling the relative effect of light pollution, impervious surfaces and intensive agriculture on bat activity with a national-scale monitoring program. Landscape Ecology. 2016; 31(10):2471–83. https://doi.org/10.1007/s10980-016-0417-3
56. Merchant ND, Fristrup KM, Johnson MP, Tyack PL, Witt MJ, Blondel P, et al. Measuring acoustic habitats. Methods in Ecology and Evolution. 2015; 6(3):257–65. https://doi.org/10.1111/2041-210X.12330 PMID: 25954500
57. Lintott PR, Bunnefeld N, Minderman J, Fuentes-Montemayor E, Mayhew RJ, Olley L, et al. Differential Responses to Woodland Character and Landscape Context by Cryptic Bats in Urban Environments. PLOS ONE. 2015; 10(5):e0126850. https://doi.org/10.1371/journal.pone.0126850 PMID: 25978034

58. Whytock RC, Christie J. Solo: an open source, customizable and inexpensive audio recorder for bioacoustic research. Methods in Ecology and Evolution. 2017; 8(3):308–12. https://doi.org/10.1111/2041-210X.12678

59. van den Oord A, Dieleman S, Zen H, Simonyan K, Vinyals O, Graves A, et al. Wavenet: A generative model for raw audio. arXiv preprint arXiv:160903499. 2016.

60. Hochreiter S, Schmidhuber J. Long short-term memory. Neural computation. 1997; 9(8):1735–80. PMID: 9377276

61. Kosmala M, Wiggins A, Swanson A, Simmons B. Assessing data quality in citizen science. Frontiers in Ecology and the Environment. 2016; 14(10):551–60. https://doi.org/10.1002/fee.1436

62. Welinder P, Branson S, Perona P, Belongie SJ, editors. The multidimensional wisdom of crowds. Advances in neural information processing systems; 2010.

63. Swanson A, Kosmala M, Lintott C, Packer C. A generalized approach for producing, quantifying, and validating citizen science data from wildlife images. Conservation Biology. 2016; 30(3):520–31. https://doi.org/10.1111/cobi.12695 PMID: 27111678

64. Prat Y, Taub M, Yovel Y. Everyday bat vocalizations contain information about emitter, addressee, context, and behavior. Scientific Reports. 2016; 6:39419. https://doi.org/10.1038/srep39419 PMID: 28005079

65. Kershenbaum A, Blumstein DT, Roch MA, Akçay Ç, Backus G, Bee MA, et al. Acoustic sequences in non-human animals: a tutorial review and prospectus. Biological Reviews. 2016; 91(1):13–52. https://doi.org/10.1111/brv.12160 PMID: 25428267