Unravelling active sensing in groups with the *Ushichka* dataset

Document updated: 2021-01-02

Collective behaviour and motion has been intensely investigated in passive sensing, visually driven animal groups to date. The behavioural heuristics underlying the impressive animal formations have been revealed through a combination of experimental tracking and modelling. Visually driven collectives can scale well with group size given sufficient light. In contrast to passive sensing animals, active sensing animals like echolocating bats emit probes of energy and sense based on how the probes are modulated by objects around them. Echolocating bats emit loud ultrasonic calls and listen for the returning echoes to detect objects. When many bats echolocate together it is expected that the loud calls of neighbouring bats will mask the detection of returning echoes and thus prevent the detection of objects. However, despite this expectation, bats aggregate and fly together in large groups such as emergences and mating swarms. The study of group echolocationg has also been impeded to date by technological issues in tracking the calls of many bats. To fill the gap in the study of active sensing groups, we present the ‘Ushichka’ dataset. Ushichka is a multi-channel, multi-sensor with audio and video data of bats echolocating as they fly in a cave environment. A LiDAR scan of the recording volume is also part of the dataset.

THINGS TO DO AFTER THE MS FINAL DRAFT IS COMPLETE …..

1. GET THE USHICHKA IN CYRILLIC!!

## Introduction

Many animals across the tree of life move and live in groups, and show impressive coordinated behaviours (Sumpter 2006). Coordinated behaviours shown in these collectives are reproduced by simple heuristics in models with agents responding independently to the movement and positions of their neighbours (Couzin et al. 2002). These coordinated behaviours have been well investigated in visually dominant animal collectives such as birds, mammals and fish (Ballerini et al. 2008; Strandburg-Peshkin et al. 2013; Pita et al. 2016). Vision is a ‘passive’ sensory modality, in that the animal acts solely as a receiver of light (Nelson and MacIver 2006). Visually driven collective behaviour can thus scale well with group size, given sufficient ambient light.

In contrast to passive sensing animals, are ‘active’ sensing animals. Active sensing animals (sensu stricto), such as bats, dolphins and electric fish, emit probes of energy and monitor the modulation of the probes by the environment around them (Nelson and MacIver 2006). Echolocating bats emit loud ultrasonic calls, and listen to the returning echoes to detect objects around them (Griffin 1958). When multiple echolocating bats echolocate together, they can start to negatively affect each other’s ability to detect their own echoes (Ulanovsky and Moss 2008). Every bat emits loud calls, and is attempting to listen to its own echo. However, along with its own returning echoes, each bat hears loud ‘non-target’ sounds (calls and echoes of neighbouring bats) which hamper its own echo detection. A bat that cannot detect its own echoes is metaphorically ‘flying blind’, and risks colliding into its neighbours and of more concern, into hard structures around it. As the number of bats in a group increases, the number of non-target sounds increases rapidly in a non-linear fashion ~. Sensory simulations show that in large groups of bats, a bat may be detecting its neighbours every few calls - and not every time it emits a call (Beleyur and Goerlitz 2019). Despite the expected deterioration of echolocation with increasing group sizes, echolocating bats in nature are social and form very large aggregations in the form of mating swarms, cave roosts and evening emergences (Ortega 2016; Erkert et al. 1982; Gillam et al. 2010). The dynamics of active sensing groups is understudied (but see Theriault et al. 2010). While it is parsimonius to assume that existing heuristic models are able to capture the dynamics of sensorially limited active sensing groups (through models with ‘limited interactions’ sensu (Bode, Franks, and Wood 2011)), it remains to be experimentally seen in what respects the behaviour of active sensing groups differ from the visually driven animal collectives studied so far. Even if active sensing groups do indeed behave much like visually driven passove sensing groups, experimental confirmation is still missing to date.

Here we present the conceptual motivation, biology and technical equipment behind the *Ushichka* dataset. The word Ushichka *(OO-shi-ch-kaa* (*cyrillic here*) in Bulgarian is the dimunitive for something with multiple ears. The name is chosen to highlight the fact that it is a multichannel, multi-sensor dataset which aims to unobtrusively ‘listen’ (observe with technological eyes and ears) to groups of multiple bats echolocating in the wild. The *Ushichka* dataset is, to our knowledge, a unique dataset in that it comprises of multichannel audio and video datasets of bats echolocating in groups in their *natural environment* captured by a LiDAR scane of the space. The multichannel nature of the dataset allows 3D localisation and tracking with high spatio-temporal resolution. Previous experimental on multi-bat echolocation in the lab and field to date has mostly been with upto 2 bats a time (Götze et al. 2016; Giuggioli, McKetterick, and Holderied 2015; Necknig and Zahn 2011; Fawcett and Ratcliffe 2015; Fawcett et al. 2015), or multichannel data of solely video or audio (Theriault et al. 2010; Lin, Abaid, and Müller 2016). The multi-camera, multi-mic nature of *Ushichka* allows us to explore echolocation in groups in much greater detail than previously attempted.

### The experimental study of echolocating bat groups

The study of bat echolocation has a long and grand history dating back over 200 years, first documented in Spallanzi’s experiments in the 18th century. After the formal discovery of echolocation (Griffin 1958; Dijkgraaf 1946), the field has moved leaps and bounds. A detailed understanding of the physiology, behaviour, neurobiology and acoustics behind echolocation has been uncovered in great detail (Popper and Fay 1995; Fenton et al. 2016). While a large body of knowledge is in place about how *individual bats* are able to echolocate through the use of trained animals in laboratory settings and field observations (Fenton et al. 2016), comparitively very little is known on how they echolocate in groups(Ulanovsky and Moss 2008).

Previously, the challenges in echolocation fieldwork included bulky analog recording equipment, and signal analysis; limiting the speed and quantity of data that could be collected, and analysed (Griffin 1958; Habersetzer 1981). However, the onset of the digital revolution has allowed for lightweight devices that can collect more data, which can then be analysed more easily. Current experimental approaches to study echolocation in groups include for example, on-body tags that record audio from the focal animal and its neighbours (along with GPS location and accelerometer data)(Cvikel et al. 2015), microphone-camera pairs to quantify broad echolocation characteristics with group size (Lin, Abaid, and Müller 2016) and thermal camera 3D tracking with sophisticated computer vision to track individuals in the million-strong emergence behaviours of *Tadarida brasiliensis* (Betke et al. 2007; Theriault et al. 2010). Each of these studies have provided exciting and important insights into how echolocating bats manage to echolocate in groups. For instance, (Cvikel et al. 2015) showed that *Rhinopoma hardwickei* do not seem to show the ‘jamming avoidance response’ (sensu Ulanovsky et al. (2004), shifting call frequencies to avoid overlap with neighbours) first seen by (Habersetzer 1981). (Lin, Abaid, and Müller 2016) find that bats echolocating in caves actually increase their pulse rate with group size, in contrast to theoretical expectations and results from another species that decreases its pulse rate (Jarvis, Jackson, and Smotherman 2013). (Theriault et al. 2010) advanced the field by quantifying the geometry of a *T. brasiliensis* emergence, showing that animals in larger groups fly relatively close to each other (0.5m), and that the neighbours are placed uniformly across azimuth and elevation of each individuals, unlike that seen in bird flocks (Ballerini et al. 2008). In starling flocks for instance, neighbours are located non-uniformly in the elevation and azimuth in that they are less likely to be found in front or behind a focal individual.

*Ushichka* represents a natural and quantitative step forward to all the recent work in the study of group echolocation in the field. The one important respect where *Ushichka* differs from previous work is its inherent ‘multi-modality’, in that it consists of simultaneously captured multi-microphone, multi-camera data streams, along with 3D scans of the recording volume for physical context. On-body tags and single microphone recordings lack explicit information of how many or where neighbours were located. Purely video tracking with one or more cameras does not provide insights into the sensory strategies animals use across group sizes. *Ushichka* captures flight and echolocation data of the group, and holds promise to provide access to the auditory scene of each individual in the group. Reconstructing the sensory inputs of individuals in a group allows us to better understand the sensorimotor decisions animals make (Strandburg-Peshkin et al. 2013). While an established method in visually driven animal groups like fish, auditory scene reconstruction is an exciting new frontier for the field of group echolocation, that *Ushichka* promises to provide.

## Experimental methods

### Field site and study species

Recordings of wild echolocating bats were made in the Orlova Chuka cave system (Ruse, Bulgaria). Recordings were made of bat flight and echolocation inside the cave underneath what is locally known as the ‘hanging pyramid’ structure (Figures 1, 2). The primary recording volume (*w x l x h*~5x9x3m) was recorded from 20-5-2018 to 19-08-2018 over the course of multiple nights. The Ushichka dataset also contains a few nights of recordings made outside of the primary recording volume. These recordings were made outside the cave, next to the grilled cave entry/exit door.

Orlova Chuka is known to host at least 9 species of bats (Ivanova 2005; “Natura 2000 Bulgarian Ministry of Environment and Water Report” 2013). During the summer, resident *Myotis myotis* and *Myotis blythii* females form the dominant bat population. These female bats arrive in spring and spend the summer in the cave system raising their young. Though rigorous mark-recapture data is missing , individual mothers are known to be consistent in returning to the cave system over multiple nights (pers comm, Laura Stidsholt, Stefan Greif). Given the dominance of *Myotis myotis* and *Myotis blythii* in the cave system, it can be fairly said that the Ushichka dataset consists mainly of these two species. Moreover, the frequency ranges of *M. myotis* and *M. blythii* are lower than those of the other resident frequency modulating bats in the cave (eg. *M. daubentonii, M. capacinii, Miniopterus schreibersii*) and are very easily distinguishable from the rhinilophid bats (*Rhinolophus ferrumequinum, R.euryale, R. mehelyi*).



Figure 1: The Ushichka setup. The large inverted T shaped array in the centre of the image is the 120cm tristar array with four SANKEN CO-100 microphones. The three blue lights are from the three thermal cameras pointing towards the array. The rest of the microphones are smaller Knowles SPU 0410 units are attached to the wall on the left portion of the image and are not visible.



Figure 2: The Ushichka setup with the smaller Knowles SPU microphones on the wall. In the picture is a *Myotis myotis* bats flying past the array. Photo by Stefan Greif.

### Audio, video and LiDAR systems

#### Audio

Multichannel audio data was recorded by multiple (2-3) Fireface (RME GmbH, Germany) sound cards running at 192kHz (Figure 5) connected to a laptop (DELL Latitude E6540, Windows 7 enterprise, core i5, 16GB RAM). The laptop was verified not to emit ultrasound. All sound cards were Fireface USB type, ASIO protocol based soundcards. The number of channels of audio in the Ushichka dataset varied between 12-22 channels over the course of the field season. The 12-22 channel data consisted of four SANKEN CO-100 microphones with the rest of the 8-18 microphons being Knowles SPU-0410 microphones. All microphones were connected to either the inbuilt pre-amplifiers of the sound cards, or to two RME QuadMic II pre-amplfiers. Later recordings with the Ushichka setup also included a Focusrite OctoPre pre-amplifier.

All recording nights had a 4-channel 120cm tristar mounted in a fairly constant position in front of the memorial plaque in the recording volume. A ‘tristar’ array is a common planar array configuration used in previous publications (Hügel et al. 2017; Goerlitz et al. 2010; Lewanzik and Goerlitz 2018), with three microphones placed at a fixed radius from a central microphone. The three peripheral microphones are separated from each other by a 120 degree angular separation. The entire set of four microphones are accomodated on an inverted-T shaped metal array. The Knowles SPU-0410 microphones were spread around the volume to increase the variety of locations at which bat calls were recorded. The positions of microphone were constant within a recording night, and were often changed between nights. On a few occasions, the microphones were kept in the same position across consecutive recording nights. Microphones were not left in the cave across sessions was due to concerns of humidity affecting the electronic circuits and membranes, especially of moisture build up in the Knowles SPU 0410 recording inlet.

A subset of all possible inter-mic distances were measured every recording session using a GLM 50C laser range finder (Bosch, Germany) (~1mm accuracy). Ultrasonic sweeps were played back from multiple points in the recording volume as part of the automatic mic position calibrations. Most of the microphones were calibrated for their frequency response at least once over the course of their usage during or before the field season. For more information on the microphone automatic position calibrations, frequency response and directionality calibrations please refer to the Supplementary Information. For further details on microphone frequency response calibration and automatic position calculations please see the Supplementary Information.

#### Video

Three thermal cameras (TeAx ThermalCapture 2.0, TeAx Technology GmbH, Germany) running with FLIR Tau 640 cores (640x512 pixel image resolution) at 25Hz formed the camera array. All three cameras were frame synchronised (Figure 3). There were a few occasions where one or two cameras recorded an extra frame, and these extra frames are known from previous tests to be at the end of the recording, and not jitter in recording initiation.

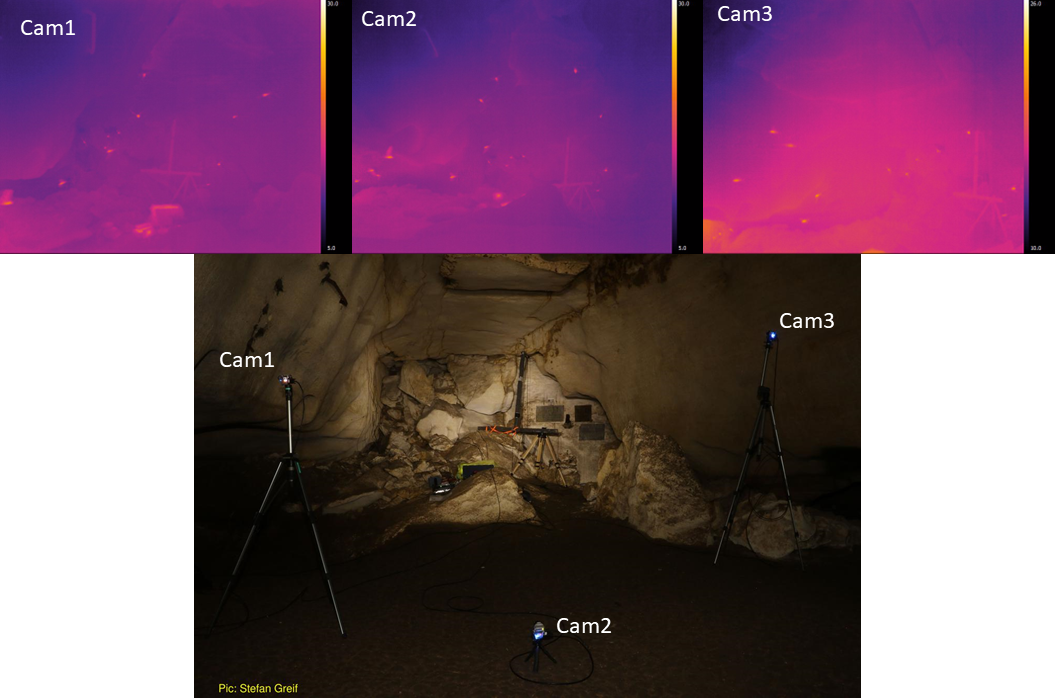


Figure 3: A multi-camera view of bats flying in the recording volume, allowing 3D tracking of flying bats (from the frame-synchronised thermal video recordings).

The microSD cards (SAMSUNG PRO+, 32 GB) were inserted into each camera. Removing the microSD card from the cameras required significant disturbance of their position, and thus the total video recording time was limited by the microSD card memory. The location of the microSD cards *inside* the camera housing meant that the data from the cameras could only be transferred in the morning after the end of the recording session. All cameras were calibrated using the ‘wand’ protocol described in (Theriault et al. 2014). More details of the camera calibration runs are in the Supplementary Information. For more information on the camera intrinsic and extrinsic parameter calibrations, details of the calibration workflows, and synchronisation with audio please refer to the Supplementary Information.

#### LiDAR

A LiDAR scan of the recording volume provides a physical context to place the echolocation and flight behaviour of bats in. With only the audio and video data, we will at most be able to recreate the positions and time of call emissions of only the bats. While bats are likely reacting to each other’s presence, they will also be reacting to the presence or absence of physical obstructions such as cave walls. With the LiDAR scan for example, we will be able to understand if a change in trajectory or call behaviour arose from the bats’ proximity to the cave wall, or to its nearest neighbour.

A LiDAR scan of the main recording volume, along with a large portion of the Orlova Chuka cave was carried out in collaboration with Dr. Asparuh Kamburov (University of Mining and Geology “St. Ivan Rilski”, Sofia). The LiDAR scan provides a high spatial resolution (<1cm) 3D map of the cave’s surface and objects in it (Figure 4). The details of the LiDAR scanning, data processing and description can be found in (Kamburov, Goerlitz, and Beleyur 2018).

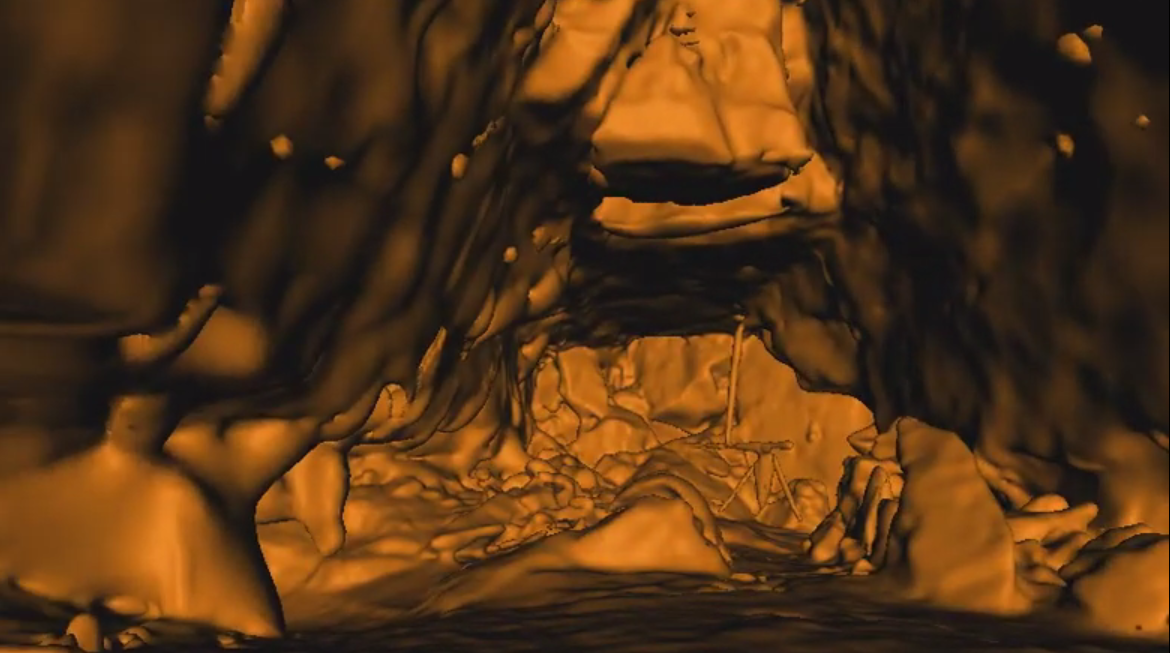


Figure 4: The LiDAR map of the recording volume showing the same view as in Figure 3, with the large tristar array.

#### Weather data

The local weather conditions (humidity, temperature and atmospheric pressure) within the cave were recorded by a Kestrel 4000 weather meter (Nielsen-Kellerman, USA) placed in the recording volume. were logged. The logged weather variables may play a role in the accuracy of position and source level estimates (Goerlitz 2018; Lawrence and Simmons 1982), and were thus measured to later estimate the speed of sound and absorption of ultrasonic sound more accurately across recording sessions.

### Recording schedule

The audio and video arrays were mostly setup before sunset (~21:00 Eastern European Time) and collected data till around sunrise (~5:30 EET) with one break in between. The recording break co-incided with an observed drop in bat activity around midnight. The recording system was switched on again between 3-4am, and left to run till about 1-2 hours after sunset.

Each recording was automatically triggered by bat calls that went above a threshold rms level (between -50 to -40 dB, threshold adjusted across recording sessions) across a subset of monitor channels. The recording was set to stop after a pre-defined duration. There were no other sources of ultrasonic sound in the cave, and it can be said with great certainty that all recordings were triggered only by bat calls. Each recording in the Ushichka dataset consists of 10-15 second long audio-video file set (See SI for details of audio-video synchronisation). The variation in recording duration is due to adjustments made over the course of the field season.

Bats can emit over 10 calls per second, which would lead to an almost continuously triggered recording system. While technically feasible, continuous recording of data would lead to unmanageably large files. Moreover, the TeAX thermal cameras relied on high speed microSD cards that were the limiting factor in the amount of data (32 GB) that could be collected over the course of a recording session. The audio-video recording system was thus set to record at a fixed duty cycle of around 20% across the nights. For example, at a 20% duty cycle and 10 second recording duration meant that recordings could be triggered at most every 50 seconds (10s recordings + 40s inactivity).

There are a few occasions of skipped morning and evening recordings due to various technical and logistical issues that were encountered over the course of the field season.

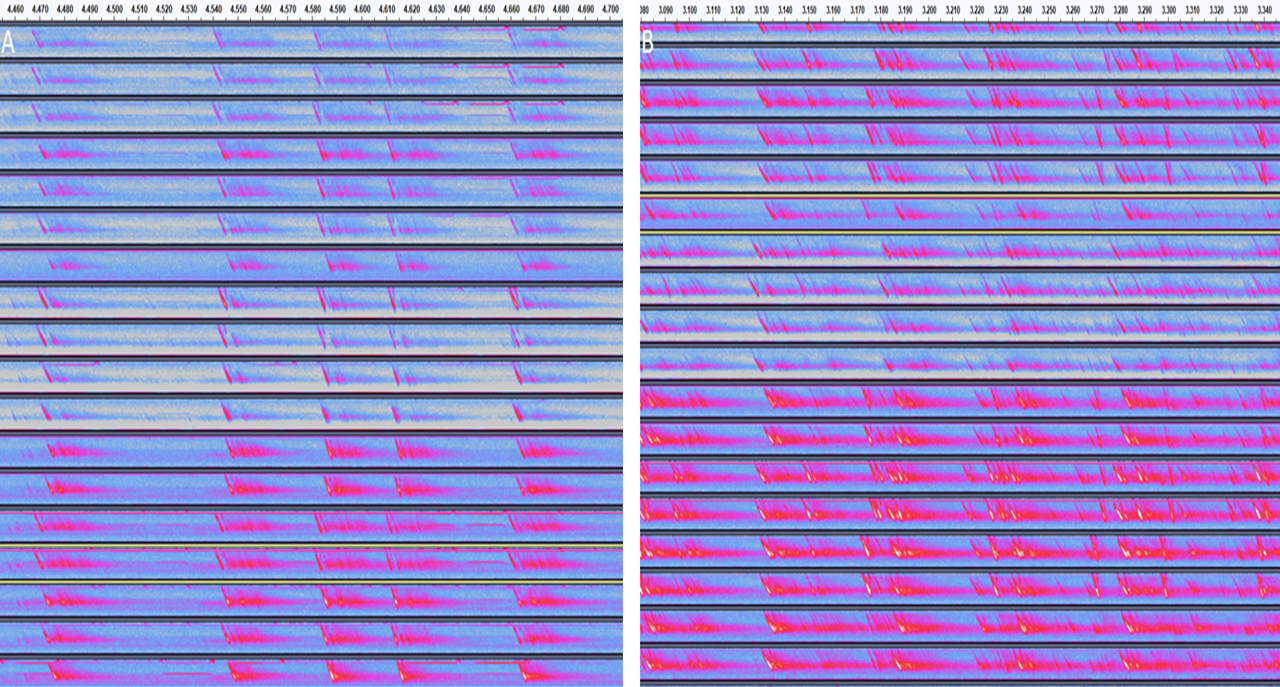


Figure 5: Spectrograms showing a subset channels from the multichannel audio data collected. The spectrograms show highlight some of the challenges of working with the audio data including reverberation, multi-path propagation, call overlaps, and correct call identity matching across channels. A) an example of a recording with a few bats in it B) recording with multiple bats in it.

### Bat activity

Bats began to fly from the interior roosting sites into the recording volume around the time of sunset. The initial activity is typically due to of small groups (2-5) of *R.euryale* and *R. mehelyi* flying very close to the microphones. The activity of the rhinolophid bats however, seemed to reduce with time, and gave way to a slowly rising number of *M. myotis* and *M. blythii* (hereon referred to as *M. myotis/blythii*) bats. The number of bats flying in the recording volume could be as high as thirty bats at a time. Towards end June, it was noticed that the *M. myotis/blythii* bats also formed temporary roosting clusters on the walls of the cave next to the pyramid, and this behaviour seemed to decrease in its frequency as the season progressed. Cloudy and rainy evenings lead to a delay or complete absence of emergence from the cave.

After reaching a peak bat activity in the recording volume about 1-1.5 hours post sun-set the bats then began to exit the cave (Figure 6). The exit seemed relatively rapid, after which only a few bats were seen to still fly in the recording volume. Individual bats returned in the early morning about one to two hours before sunset. It appears that *R. euryale* and *R. mehelyi* activity remained fairly constant through the night, even during the drop in activity around midnight.

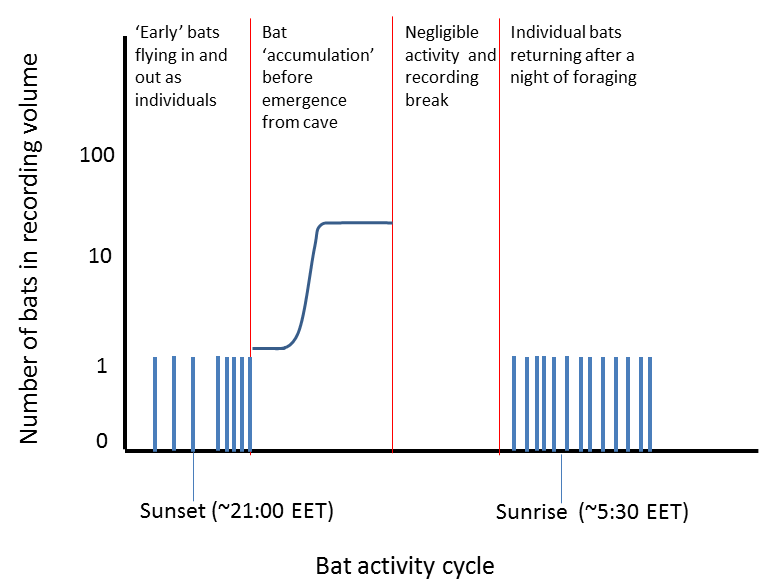


Figure 6: Schematic of typical bat activity pattern of Myotis myotis and Myotis blythii between sunset to sunrise in the Ushichka dataset. The multichannel recording system was typically activated before sunset and was run till after sunrise with a break in between

### Software used in this work

Automatic recording triggering implemented in the fieldrecorder\\_trigger module relies on the open-source Python language (Van Rossum and Drake 2011), and the sounddevice, scipy and numpy (Geier 2015; Virtanen 2019; Oliphant 2006) packages. This manuscript was written using the knitr package (Xie 2015).

## Technical challenges ahead and potential solutions in handling the *Ushichka* dataset

### Handling high call densities in audio data

Analysing the echolocation call data will require surmounting a series of challenges, here discussed in increasing order of difficulty. 1) The recording volume consisted of acoustically reflective surfaces, leading to significant reverberation and multi-path propagation. Dealing with reverberation is likely to be relatively straightforward while performing acoustic tracking, as cross-channel correlations should be able to filter out the noise contributed by reverberation. Multi-path propagation, where a sound arrives multiple times at each microphone due to distinct reflections is likely to be a somewhat more difficult problem to deal with, especially considering that if a single call leads to *N* multi-path reflections, then *M* calls in the volume could then lead to *NxM* reflections. 2) Each bat flying in and out of the volume may be emitting calls at between 10-20 times per second (10-20Hz). With one bat, it is relatively straightforward to acoustically localise the bat as each call is separated by tens of milliseconds from the next one. With an increasing number of bats in the volume, (as happens towards the end of the night, Figure 6), the temporal separation between calls will decrease, and also lead to a higher occurence of call overlaps. Handling such high call densities will be a challenge, and require the latest techniques in call detection in the presence of overlaps (Izadi, Stevenson, and Kloepper 2020), or in general blind signal separation techniques (Brandstein and Ward 2013). 3) Even if calls can be detected reliably in single channels, the problem of cross-channel correspondence still remains to be solved,a common problem faced with any multi-channel acoustic array (Brandstein and Ward 2013). An echolocation call detected in one channel needs to be matched with itself in another channel, despite the possible presence of multiple similar calls in the expected time-window, in every other channel. Individual bats do have unique call characteristics (Masters, Raver, and Kazial 1995, but see @siemersnovoice) (and specifically in *M. myotis*, (Yovel et al. 2009)). The unique echolocation call ‘signatures’ are sometimes apparent from visual inspection of the spectrogram. However, manual correspondence matching is unlikely to scale well in terms of effort required with group sizes beyond 3 bats and will require robust automated algorithms to solve this combinatorial problem.

One important ‘support system’ which may help significantly in solving at least two of the issues mentioned above is the multi-camera video data. The time-synchronised 3D video tracking data contains the microphone positions and bat trajectories. The basic principle of acoustic localisation relies on the fact that with more than 4 microphones spread in space, each position of sound emission will generate a unique set of time-difference-of-arrivals (TDOAs). A TDOA in a microphone array is the difference in arrival time of a sound in comparison to a chosen reference channel. Measuring these TDOAs thus allows inferring the position at which a bat emitted a call from. The 3D video trajectories allow the generation of a series of ‘hypotheses’ of which calls could have been emitted at which point in space. By calculating the time-varying distances between a bat and each microphone across time, it will allow us to generate a set of ‘predicted’ TDOAs that calls could be arriving at. The problem thus boils down to a matching problem, where all TDOAs from all possible call correspondence matches are filtered based on which TDOAs are predictions from a flight trajectory. The video trajectory based filtering approach could in principle elegantly solve the problems of multi-path propagation and cross-channel correspondence at one shot. However, whether the 3D video data will actually be able to deliver sufficiently accurate position estimates, and thus accurate TDOAs remains to be seen.

### Fusing multiple data streams to reconstruct individual auditory scenes

*Ushichka* consists of multiple data streams consisting of acoustic, video and LiDAR data. The acoustic and video streams are tightly linked as they describe the flight and echolocation behaviour of bats in the recording volume. However, the bats in the recording volume are not only trying to detect the presence of other bats, but also trying to avoid collisions with the walls and structures in the cave. This is where the LiDAR dataset provides an extremely important contextual understanding of how and why bats may be choosing to fly or echolocate in the space they occupy. Fusing the acoustic and video data streams is relatively straightforward in that they are time-synchronised and the ‘objects’ being tracked are the same. The 3D positions derived from both data streams can be overlaid to match with a simple rigid rotation of coordinate systems [Figure 7].This apparently simple overlay procedure can also suffer from issues related to localisation and calibration accuracy of both systems. However, while it is clear that all recorded bat flights occured within the volume scanned in the LiDAR data, placing it exactly in space is required to understand which bat was how far away from a wall when it decided to turn. The most natural approach is thus to pair the video and LiDAR data.

One possible approach is to run a structure-from-motion (SfM) analysis (reviewed in Ozyesil et al. 2017) to recreate the 3D structure of the recording volume, and thus find the best fit between the LiDAR surface and the estimated surface. A structure-from-motion analysis typically requires multiple views of the same object from different angles. Given that there were three cameras recording from unique positions on each recording day, it may be possible that there are overall viewpoints available for this analysis. However, the three cameras were not always placed in *completely* unique positions, in that they were kept facing a relatively stable direction (into the recording volume), and each camera had its generalised location fixed (as seen in Figure 3 optimised for the best fields of views for 3D tracking of bats). In addition, SfM has been succesfully applied to a large variety of RGB images which provide three separate color channels to perform feature extraction and matching across views. In contrast to RGB images, thermal images are mono-channel and the cameras we used also have a limited thermal resolution (~1C), along with inter-camera variation in thermal sensitivity. In addition to the technical limitations is the actual thermal environment in the cave. The entire cave system is a closed system which maintains a relatively stable temperature of around 10C through the year. From a thermal perspective, the constrast between one part of the cave and another may not be as large as what is seen in many RGB images. The challenge in overlaying the LiDAR and video streams will be in performing a what may be a data-limited SfM with a low number of views and possibly ‘featureless’ images.

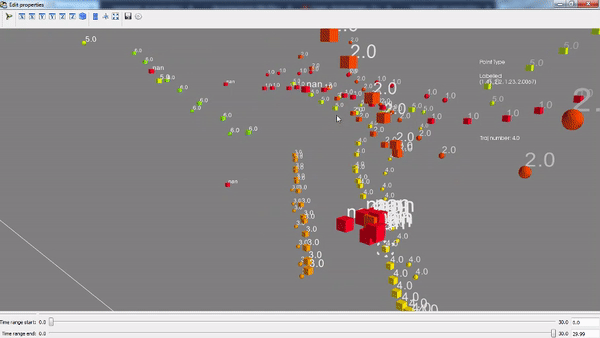


Figure 7: Proof-of-principle data showing audio-video trajectory overlay of bat flights. Data was collected in a flight room using the audio-video tracking system behind *Ushichka*. Each trajectory is represented by a different color. Square and round points represent audio and video positions. As can be seen, the audio and video trajectories are typically close, but do not always overlap exactly likely because of non-uniform localisation accuracies across the volume of interest.

### Towards a field-friendly protocol for microphone array tracking

Working with multi-microphone arrays typically includes carrying many long cables and heavy array frames to hold microphones in known positions while recordings happen. Microphone positions need to be known for successful acoustic localisation, and well-specified mic positions lead to lower localisation errors (Wahlberg 1999). However, carrying large array frames into the field, especially in cluttered environments such as caves or forests is often difficult and time-consuming. Moreover, bats may actually focus their calls specifically on these new objects and the recorded behaviour might actually be an unwanted artifact of the equipment itself. There is thus a pressing need in bioacoustics for the development of acoustic arrays that are frame-less and can be setup universally and quickly.

Once frames have been abandoned, it is still possible to estimate microphone positions indirectly by exhaustively measuring inter-mic distances using a laser range finder or tape-measure (as has been done for *Ushichka*). However, manual inter-mic measurements can be error-prone at larger distances (~4-5m) and do not scale well with increasing microphone numbers in an array (number of measurements to be made scales rapidly with ). Having experienced these problem with array-less microphones, we were looking for solutions that have the ease of camera calibration workflows (eg. Theriault et al. 2014), and found the ‘Structure-from-Sound’ approach of (Zhayida et al. 2016) . (Zhayida et al. 2016) were able to infer microphone positions based on commonly recorded playbacks. We had the fortune of being able to engage in a collaboration and show the success of their workflow with our data using one evening of ground-truthed microphone array positions (Batstone et al. 2019, also see Supplementary Information). Such automatic position inference workflows have the potential to drastically reduce the effort required to perform multi-channel recordings in the field, and open up the realm of exciting experimental setups to observe animals in their natural environments unobtrusively. We are currently continuing the collaboration and looking to formulate calibration workflows that are also field-friendly, and using the Structure-from-Sound to infer positions of the majority of microphones not placed on the array in *Ushichka*.

## Investigative opportunities opened by the Ushichka dataset

Here we detail some of the possible lines of investigations that can be undertaken with *Ushichka* once the necessary analytical workflows are in place.

### Collective behaviour in active sensing groups

The combination of acoustic and video tracking along with a LiDAR scan of the cave provides us sufficient data to reconstruct the sensory inputs of each bat in a group. Given a group of *N* bats, we can thus reconstruct all the sounds each bat would have heard in their inter-pulse intervals (the silent gap between calls). The sounds would include the focal bat’s own echoes, secondary echoes from other bats, and the calls of other bats. Previous studies have attempted to simulate such contexts with biologically parametrised values of the input sounds (Beleyur and Goerlitz 2019; Mazar and Yovel 2020). Our experimental measurements in combination with detailed sound propagation models could allow for *direct* sensory input reconstructions in a whole group of bats simultaneously. Knowing the sensory inputs animals receive over time would allow us to finely parametrise relevant models of collective behaviour (eg. Bode, Franks, and Wood 2011 or a dynamic version of @beleyur2019modeling) and test their fit. Sensory reconstruction approaches in echolocation would also open the field to fine-scale temporal analyses of echolocationa and flight behaviour: do bats turn more often in response to an echo from a wall, or away from the direction of a calling neighbour? Do bats call more often when they may have ‘missed’ echo detections due to call overlaps in the past few interpulse intervals?

The study of biological active sensing collectives is an exciting field waiting to be explored through inter-disciplinary studies. So far, most work on active sensing collectives has gone into technological collectives such as robot or drone swarms (Rubenstein, Ahler, and Nagpal 2012; Vásárhelyi et al. 2018). Exploring how active sensing animal collectives deal with the same problems of collision-avoidance and sensor limitations (directionality,power,overlap) promises to yield the discovery of independent parallels or perhaps even completely new algorithms at work.

### Optimising echolocation and flight strategies under challenging conditions

Bats are known to alter their echolocation behaviour rapidly according to the behavioural context, and ambient soundscape [review, and some noise related reviews]. *Ushichka* offers a direct glimpse into the variety of echolocation strategies in play across a gradient of acoustical challenge as group size increases (Fig 6). Aside from call-level data such as source-level, duration, spectral properties - the presence of multiple microphones spread across the room also provides access to call direction and beam shape modulations (Figure 8A). While other call parameters have been investigated in the presence of other animals, call directionality and beam-shape measurements have so far mainly been done with single animals in the field and lab []. Revealing where bats aim their calls, and how they spread the energy in each call will uncover the sensory priorities bats may have while flying under challening conditions. Do bats choose to focus narrowly on fast moving conspecifics, but broadly onto bigger obstructions like walls. Does their call aim and beam shape change with increasing group size?

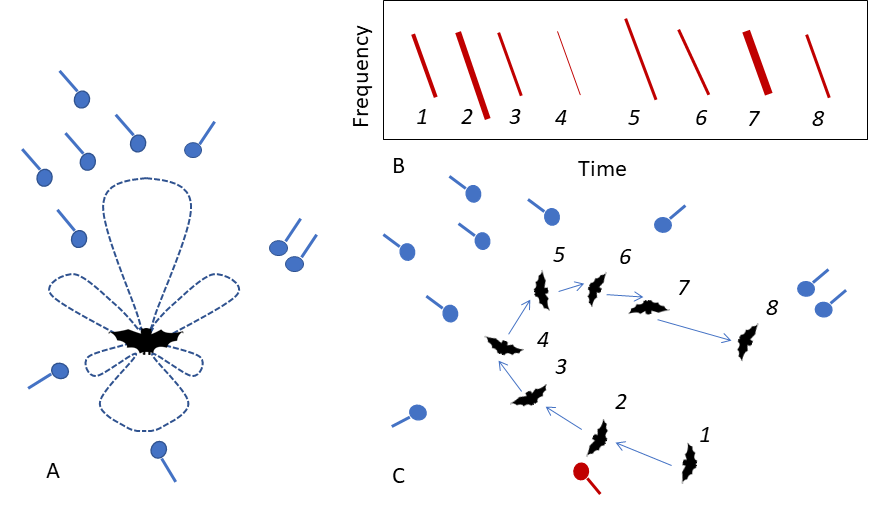


Figure 8: Schematics showing how multi-mic arrays bring many advantages in the study of echolocation and voice recognition (microphone: blue circle with line attached) A) allow beam-shape estimation because of microphones spread all around the animals B,C) a schematic spectrogram in B) showing numbered call-by-call variation recorded at a focal microphone in an array (red microphone in (C)). As the bat moves through the volume in C) the amplitude and spectral properties of the call changes. Having multiple microphones provides a multi-view picture of a call from many directions, allowing a better assessment of intra-individual call variation required for robust voice recognition

The echolocation strategies in use may also inform why certain types of apparently stereotypical flight paths are in use within the recording volume. TB observed that bats sometimes tend to fly a circular loop close to the edges of the recording volume before exiting out or heading back into the roosting site inside (Figure 9). This ‘looping’ flight behaviour was observed in the evening flight activity as bats began to accumulate in the recording volume, but also remained when bats returned in the morning alone. Bats are known to show stereotypic flight behaviours in the lab and in the wild [Barchi, Knowles, and Simmons (2013);????], and in some cases it might be a matter of following the same path the animal first learnt in a cluttered environment. However, in a cave setting, where there are no obvious obstacles, what drives the formation of these seemingly ‘odd’ looping behaviours? Do the loops form a type of stereotypic trajectory that allows the bat to limit its attention only along that route? Why do some bats not take the shortest path and fly from the exit straight into roosting site? TB also had the impression that even when multiple bats were flying together, not all parts of the recording volume were being ‘used’ by the bats. Initial observations of the thermal data suggested that some regions of the recording volume (regions close to walls) seemed to have a higher density of bats. Is this wall-clinging behaviour a behavioural adaptation to maintain a high received level of wall-echoes (in the presence of non-target sounds)? Another interesting parameter to estimate is the dynamics of the loop’s chirality. Alignment is a standard measure of how similar animals are in their direction of movement. When there are multiple animals in the recording volume, a measure of the alignment over time and group size might reveal whether the looping behaviour is an individually driven behaviour or is modulated by the flight directions of group members.

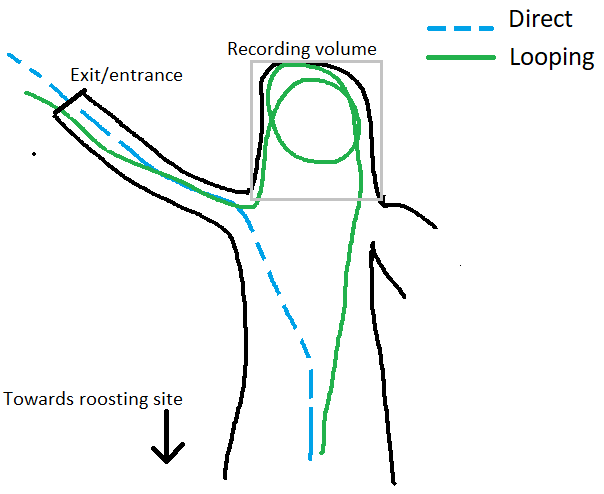


Figure 9: Schematic showing a view of the direct and looped flight paths as seen from above. The direct (broken blue line) and indirect (continuous green line) paths are flown in both directions (exit/entrance-cave interior, cave interior-exit/entrance). The grey quadrilateral outlines the recording volume.

### Voice recognition for short-term and long-term behavioural tracking [NEED TO SHORTEN!!!!]

Vocalising animals across the animal kingdom have uniquely identifiable vocalisations, or voices(Carlson, Kelly, and Couzin 2020). The voice of each animal is a temporally stable cue, unlike experimental markings which may be lost over time. A primary issue with developing a stable ‘voice’ tracker is the stability of the vocalisation itself. Bird call repertoires can change over time, and complicate identification (Stowell et al. 2019). Moreover, variations caused by sound directionality, distance and ambient noise may require larger training datasets, which are often hard to obtain (Stowell et al. 2019). Unlike many types of animal vocalisations (eg. bird song), the spectro-temporal structure of bat echolocation calls are stereotypical and likely to remain stable because they serve a single sensory purpose – to detect objects.

Bat echolocation calls can be assigned to individuals by their acoustic features, though this has typically been done in small lab populations. The methods in (Masters, Raver, and Kazial 1995; Yovel et al. 2009) solve the ‘closed-set’ problem where all animal identities are known. *Ushichka* provides a large dataset over multiple nights and animals that fly by [also see Figure 8B,C]. In contrast to the ‘closed-set’ problems solved in the past, *Ushichka* could be used to generate methods to solve the ‘open-set’ voice recognition task. Open-set classification problems deal with models that uniquely identifying individuals that they have been trained on, but also those they have *not* been explicitly trained on . Even though the identities of bats echolocating in the recordings are unknown, it may be possible to exploit the behavioural patterns of the animals to approximate identity. Bats always return alone in the morning, and these recordings contain echolocation calls from one individual. In addition, it is known that female *M. myotis/blythii* roost in the cave system over the entire summer period. Using these two facts, if successful, a voice recognition model should at least show certain patterns 1) correctly assign identity to an independent subset of calls it was not trained on (validation data) from a morning recording 2) identify at least some individuals across multiple morning returns, and in the best case 3) be able to pick up the individuals as they begin their activity in the recording volume around sunset. Of course, even if a trained model satisfies the three following criteria, it does not directly follow that the detections in scenarios 2) and 3) are not false positives. It may be necessary to systematically collect a larger sample of echolocation calls with hand-release recordings from multiple individuals in the future. The state of signal-processing and machine learning know-how in the field of human voice recognition is relatively mature at this point, with the use of speaker-specific Gaussian Mixture Models compared with Universal Background Models trained against multiple individuals (Hennebert 2009).

While seemingly challenging, if a validated voice recognition model is indeed developed successfully, it will allow revolutionary markerless short-term and long-term tracking of individual behaviours. Short-term behaviours such as flight and echolocation decisions over the course of a few seconds can provide insights into the individual sensorimotor strategies and the tracking of social interactions. Long-term decisions such as the change exit-time from cave, and morning return times can reveal changing physiological status of the resident females as their pups mature. Voice based individual recognition mimics the advantages of tracking animals (zebras, tigers) with unique markings on their bodies in that it allows unobtrusive observations to be made at multiple time-scales.

## Supplementary code and website

The fieldrecorder\_trigger.py module used to trigger automatic recording is uploaded at the following repository: <https://github.com/thejasvibr/fieldrecorder>. Those interested in viewing more snippets of the the *Ushichka* dataset and the progress of research related to it, may wish to visit <https://thejasvibr.github.io/ushichka/>.

## Field work permits

All field work at the Orlova Chuka cave was done with permission from the relevant authorities. *Ushichka* is a purely observational dataset, no animals were handled or subjected to experimental treatments of any sort over the course of data collection.

## Author Contributions

TB: conception, experiment design, field data collection. HRG: conception, experiment and recording-system design.

## Acknowledgements

We would like to thank Antoniya Hubancheva, for her constant help and support in the collection of the Ushichka dataset (also for christening the dataset with its Bulgarian name) and the Tabachka field crew of 2018. The electronics (Markus Abels, Hannes Sagunsky, Reinhard Biller) and Feinmechanik (Erich Koch, Felix Hartl, Klaus Pichler) teams were of invaluable help and support all through the process of experimental setup design and troubleshooting. Joanna Furmankiewicz for supporting initial recordings at the Jaskinia Niedźwiedzia and Diane Theriault for providing advice on thermal cameras. Special thanks and acknowledgement to the field assistance of Aditya Krishna and Neetash MR, without whom the data collection would not have been possible. T.B. was funded by a doctoral fellowship from the German Academic Exchange Service (DAAD) and the International Max Planck Research School for Organismal Biology. H.R.G. was funded by the Emmy Noether program of the German Research Foundation (DFG, GO 2091/2-1, GO 2091/2-2). We would like to thank the members of the Acoustic and Functional Ecology groups and the Max Planck Institute for Ornithology, Seewiesen for its support and infrastructure.

## References

Ballerini, Michele, Nicola Cabibbo, Raphael Candelier, Andrea Cavagna, Evaristo Cisbani, Irene Giardina, Vivien Lecomte, et al. 2008. “Interaction Ruling Animal Collective Behavior Depends on Topological Rather Than Metric Distance: Evidence from a Field Study.” *Proceedings of the National Academy of Sciences* 105 (4): 1232–7.

Barchi, Jonathan R, Jeffrey M Knowles, and James A Simmons. 2013. “Spatial Memory and Stereotypy of Flight Paths by Big Brown Bats in Cluttered Surroundings.” *Journal of Experimental Biology* 216 (6): 1053–63.

Batstone, K., G. Flood, T. Beleyur, V. Larsson, H. R. Goerlitz, M. Oskarsson, and K. Åström. 2019. “Robust Self-Calibration of Constant Offset Time-Difference-of-Arrival.” In *ICASSP 2019 - 2019 Ieee International Conference on Acoustics, Speech and Signal Processing (Icassp)*, 4410–4.

Beleyur, Thejasvi, and Holger R Goerlitz. 2019. “Modeling Active Sensing Reveals Echo Detection Even in Large Groups of Bats.” *Proceedings of the National Academy of Sciences* 116 (52): 26662–8.

Betke, Margrit, Diane E Hirsh, Angshuman Bagchi, Nickolay I Hristov, Nicholas C Makris, and Thomas H Kunz. 2007. “Tracking Large Variable Numbers of Objects in Clutter.” In *2007 Ieee Conference on Computer Vision and Pattern Recognition*, 1–8. IEEE.

Bode, N. W. F. F., D. W. Franks, and A. J. Wood. 2011. “Limited Interactions in Flocks: Relating Model Simulations to Empirical Data.” *J. R. Soc. Interface* 8: 301–4.

Brandstein, Michael, and Darren (Eds.) Ward. 2013. *Microphone Arrays: Signal Processing Techniques and Applications*. Springer Science & Business Media.

Carlson, Nora V, E McKenna Kelly, and Iain Couzin. 2020. “Individual Vocal Recognition Across Taxa: A Review of the Literature and a Look into the Future.” *Philosophical Transactions of the Royal Society B* 375 (1802): 20190479.

Couzin, I. D., J. Krause, R. James, G. D. Ruxton, and N. R. Franks. 2002. “Collective Memory and Spatial Sorting in Animal Groups.” *J. Theor. Biol* 218: 1–11.

Cvikel, Noam, Eran Levin, Edward Hurme, Ivailo Borissov, Arjan Boonman, Eran Amichai, and Yossi Yovel. 2015. “On-Board Recordings Reveal No Jamming Avoidance in Wild Bats.” *Proceedings of the Royal Society B: Biological Sciences* 282 (1798): 20142274.

Dijkgraaf, Sven. 1946. “Die Sinneswelt der Fledermäuse.” *Experientia* 2 (11): 438–48. <https://doi.org/10.1007/BF02153592>.

Erkert, Hans G., M Brock Fenton, James S. Findley, Theodore H. Fleming, Raymond E. Heithaus, Thomas H. Kunz, Adrian G. Marshall, et al. 1982. *Ecology of Bats*. <https://doi.org/10.1007/978-1-4613-3421-7>.

Fawcett, Kayleigh, David S Jacobs, Annemarie Surlykke, and John M Ratcliffe. 2015. “Echolocation in the Bat, Rhinolophus Capensis: The Influence of Clutter, Conspecifics and Prey on Call Design and Intensity.” *Biology Open* 4 (6): 693–701.

Fawcett, Kayleigh, and John M Ratcliffe. 2015. “Clutter and Conspecifics: A Comparison of Their Influence on Echolocation and Flight Behaviour in Daubenton’s Bat, Myotis Daubentonii.” *Journal of Comparative Physiology A* 201 (3): 295–304.

Fenton, M Brock, Alan D Grinnell, Arthur N Popper, and Richard R Fay Editors. 2016. *Bat Bioacoustics*. Edited by Brock M Fenton, Alan D Grinnell, Arthur N Popper, and Richard R Fay. Vol. 54. New York: Springer Science+Business Media k. <https://doi.org/10.1007/978-1-4939-3527-7>.

Geier, M. 2015. <https://python-sounddevice.readthedocs.io/en/0.3.5/_modules/sounddevice.html.>

Gillam, Erin H, Nickolay I Hristov, Thomas H Kunz, and Gary F McCracken. 2010. “Echolocation Behavior of Brazilian Free-Tailed Bats During Dense Emergence Flights.” *Journal of Mammalogy* 91 (4): 967–75.

Giuggioli, L., T. J. McKetterick, and M. Holderied. 2015. “Delayed Response and Biosonar Perception Explain Movement Coordination in Trawling Bats.” *PLoS Comput Biol* 11 (3).

Goerlitz, Holger R. 2018. “Weather Conditions Determine Attenuation and Speed of Sound: Environmental Limitations for Monitoring and Analyzing Bat Echolocation.” *Ecology and Evolution* 8 (10): 5090–5100.

Goerlitz, Holger R., Hannah M. Ter Hofstede, Matt R. K. Zeale, Gareth Jones, and Marc W. Holderied. 2010. “An aerial-hawking bat uses stealth echolocation to counter moth hearing.” *Current Biology* 20 (17): 1568–72. <https://doi.org/10.1016/j.cub.2010.07.046>.

Götze, S., J. C. Koblitz, A. Denzinger, and H. U. Schnitzler. 2016. “No Evidence for Spectral Jamming Avoidance in Echolocation Behavior of Foraging Pipistrelle Bats.” *Sci Rep* 6 (July).

Griffin, Donald R. 1958. “Listening in the dark: the acoustic orientation of bats and men.”

Habersetzer, J. 1981. “Adaptive Echolocation Sounds in the Bat Rhinopoma Hardwickei.” *J. Comp. Physiol* 144: 559–66.

Hennebert, Jean. 2009. “Speaker Recognition, Overview.” In *Encyclopedia of Biometrics*, edited by Stan Z. Li and Anil Jain, 1262–70. Boston, MA: Springer US. <https://doi.org/10.1007/978-0-387-73003-5_198>.

Hügel, Theresa, Vincent van Meir, Amanda Muñoz-Meneses, B. Markus Clarin, Björn M. Siemers, and Holger R. Goerlitz. 2017. “Does similarity in call structure or foraging ecology explain interspecific information transfer in wild Myotis bats?” *Behavioral Ecology and Sociobiology* 71 (11). <https://doi.org/10.1007/s00265-017-2398-x>.

Ivanova, Teodora. 2005. “Important Bat Underground Habitats (IBUH) in Bulgaria.” *Acta Zoologica Bulgarica* 57 (2): 197–206.

Izadi, Mohammad Rasool, Robert Stevenson, and Laura N Kloepper. 2020. “Separation of Overlapping Sources in Bioacoustic Mixtures.” *The Journal of the Acoustical Society of America* 147 (3): 1688–96.

Jarvis, Jenna, William Jackson, and Michael Smotherman. 2013. “Groups of Bats Improve Sonar Efficiency Through Mutual Suppression of Pulse Emissions.” *Frontiers in Physiology* 4: 140.

Kamburov, Asparuh, Holger R Goerlitz, and Thejasvi Beleyur. 2018. “Geospatial Modelling Inside the "Orlova Chuka" Cave in Bulgaria.” *XXVIII International Symposium on Modern Technologies, Education and Professional Practice in Geodesy and Related Fields*. <http://symp2018.geodesy-union.org/wp-content/uploads/2018/11/20.pdf>.

Lawrence, Beatrice D, and James A Simmons. 1982. “Measurements of Atmospheric Attenuation at Ultrasonic Frequencies and the Significance for Echolocation by Bats.” *The Journal of the Acoustical Society of America* 71 (3): 585–90.

Lewanzik, Daniel, and Holger R. Goerlitz. 2018. “Continued source level reduction during attack in the low-amplitude bat Barbastella barbastellus prevents moth evasive flight.” *Functional Ecology* 32 (5): 1251–61. <https://doi.org/10.1111/1365-2435.13073>.

Lin, Yuan, Nicole Abaid, and Rolf Müller. 2016. “Bats Adjust Their Pulse Emission Rates with Swarm Size in the Field.” *The Journal of the Acoustical Society of America* 140 (6): 4318–25.

Masters, W Mitch, Kelley AS Raver, and Karry A Kazial. 1995. “Sonar Signals of Big Brown Bats, Eptesicus Fuscus, Contain Information About Individual Identity, Age and Family Affiliation.” *Animal Behaviour* 50 (5): 1243–60.

Mazar, Omer, and Yossi Yovel. 2020. “A Sensorimotor Model Shows Why a Spectral Jamming Avoidance Response Does Not Help Bats Deal with Jamming.” *Elife* 9: e55539.

“Natura 2000 Bulgarian Ministry of Environment and Water Report.” 2013. <http://natura2000.moew.government.bg/Home/ProtectedSite?code=BG0000608&siteType=HabitatDirective>.

Necknig, Veronika, and Andreas Zahn. 2011. “Between-Species Jamming Avoidance in Pipistrelles?” *Journal of Comparative Physiology A* 197 (5): 469–73.

Nelson, M. E., and M. A. MacIver. 2006. “Sensory Acquisition in Active Sensing Systems.” *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol* 192: 573–86.

Oliphant, T. E. 2006. *A Guide to Numpy*. Trelgol Publishing USA.

Ortega, Jorge. 2016. *Sociality in bats*. Edited by Jorge Ortega. Springer International Publishing Switzerland. <https://doi.org/10.1007/978-3-319-38953-0>.

Ozyesil, Onur, Vladislav Voroninski, Ronen Basri, and Amit Singer. 2017. “A Survey of Structure from Motion.” *arXiv Preprint arXiv:1701.08493*.

Pita, Diana, Bertrand Collignon, José Halloy, and Esteban Fernández-Juricic. 2016. “Collective Behaviour in Vertebrates: A Sensory Perspective.” *Royal Society Open Science* 3 (11): 160377.

Popper, Arthur N, and Richard R Fay. 1995. *Hearing by Bats*. Springer.

Rubenstein, Michael, Christian Ahler, and Radhika Nagpal. 2012. “Kilobot: A Low Cost Scalable Robot System for Collective Behaviors.” In *2012 Ieee International Conference on Robotics and Automation*, 3293–8. IEEE.

Siemers, Björn M, and Gerald Kerth. 2006. “Do Echolocation Calls of Wild Colony-Living Bechstein’s Bats (Myotis Bechsteinii) Provide Individual-Specific Signatures?” *Behavioral Ecology and Sociobiology* 59 (3): 443–54.

Stowell, Dan, Tereza Petrusková, Martin Šálek, and Pavel Linhart. 2019. “Automatic Acoustic Identification of Individuals in Multiple Species: Improving Identification Across Recording Conditions.” *Journal of the Royal Society Interface* 16 (153): 20180940.

Strandburg-Peshkin, Ariana, Colin R Twomey, Nikolai WF Bode, Albert B Kao, Yael Katz, Christos C Ioannou, Sara B Rosenthal, et al. 2013. “Visual Sensory Networks and Effective Information Transfer in Animal Groups.” *Current Biology* 23 (17): R709–R711.

Sumpter, David JT. 2006. “The Principles of Collective Animal Behaviour.” *Philosophical Transactions of the Royal Society B: Biological Sciences* 361 (1465): 5–22.

Theriault, D. H., Z Wu, N. I. Hristov, S. M. Swartz, K. S. Breuer, T. H. Kunz, and M. Betke. 2010. *Reconstruction and Analysis of 3D Trajectories of Brazilian Free-Tailed Bats in Flight*. Bost Univ Comput Sci Technical Rep.

Theriault, Diane H., Nathan W. Fuller, Brandon E. Jackson, Evan Bluhm, Dennis Evangelista, Zheng Wu, Margrit Betke, and Tyson L. Hedrick. 2014. “A protocol and calibration method for accurate multi-camera field videography.” *Journal of Experimental Biology* 217 (11): 1843–8. <https://doi.org/10.1242/jeb.100529>.

Ulanovsky, Nachum, M Brock Fenton, Asaf Tsoar, and Carmi Korine. 2004. “Dynamics of Jamming Avoidance in Echolocating Bats.” *Proceedings of the Royal Society of London. Series B: Biological Sciences* 271 (1547): 1467–75.

Ulanovsky, Nachum, and Cynthia F Moss. 2008. “What the Bat’s Voice Tells the Bat’s Brain.” *Proceedings of the National Academy of Sciences* 105 (25): 8491–8.

Van Rossum, Guido, and Fred L Drake. 2011. *The Python Language Reference Manual*. Network Theory Ltd.

Vásárhelyi, Gábor, Csaba Virágh, Gergő Somorjai, Tamás Nepusz, Agoston E Eiben, and Tamás Vicsek. 2018. “Optimized Flocking of Autonomous Drones in Confined Environments.” *Science Robotics* 3 (20).

Virtanen, P. 2019. “SciPy 1.0–Fundamental Algorithms for Scientific Computing in Python.”

Wahlberg, Magnus. 1999. “Positioning accuracy of a large‐aperture hydrophone array for sperm whale research.” *The Journal of the Acoustical Society of America* 105 (2): 1318–8. <https://doi.org/10.1121/1.424536>.

Xie, Yihui. 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. <https://yihui.org/knitr/>.

Yovel, Yossi, Mariana Laura Melcon, Matthias O Franz, Annette Denzinger, and Hans-Ulrich Schnitzler. 2009. “The Voice of Bats: How Greater Mouse-Eared Bats Recognize Individuals Based on Their Echolocation Calls.” *PLoS Comput Biol* 5 (6): e1000400.

Zhayida, Simayijiang, Simon Segerblom Rex, Yubin Kuang, Fredrik Andersson, and Kalle Åström. 2016. “An Automatic System for Acoustic Microphone Geometry Calibration Based on Minimal Solvers.” *CoRR* abs/1610.02392. <http://arxiv.org/abs/1610.02392>.