

Theoretical and empirical investigation of echolocation in bat groups

Thejasvi Beleyur

2021 – 01 – 12

Summary	7
Zusammenfassung	9
1 General Introduction	11
1.1 Active sensing in groups : a tricky proposition	13
1.2 The nascent field of group echolocation	14
1.3 Science-technology loops in echolocation	14
1.4 Thesis outline	15
2 Modeling active sensing in groups of bats reveals echo detection even in large groups	19
Abstract	20
2.1 Introduction	21
2.2 Materials and methods	22
2.3 Results	26
2.4 Discussion	29
2.5 Conclusion	33
2.6 Supplementary Information	34
3 Horseshoe bat craziness paper	53
4 ushichka	55
4.1 Example one	55
4.2 Example two	55

5 Robust Self-Calibration of Constant Offset Time-Difference-of-Arrival	57
Abstract	58
5.1 Introduction	59
5.2 Time-difference-of-arrival self calibration	59
5.3 Local optimization and the low rank relaxation	60
5.4 Minimal problems and solvers	61
5.5 Using RANSAC for five rows	62
5.6 Robust estimation of parameters	63
5.7 Experimental Validation	63
5.8 Conclusions	65
6 tacost: Testing and simulating the performance of acoustic tracking systems	67
Abstract	68
6.1 Introduction	69
6.2 Statement of need	69
6.3 Design	69
6.4 Examples	70
6.5 Future directions	72
6.6 Acknowledgements	73
7 itsfm, an open-source package to reliably segment and measure sounds by frequency modulation	75
Abstract	76
7.1 Introduction	77
7.2 Package description	78
7.3 Methods evaluation	84
7.4 Discussion	87
7.5 Open-source software and packages used	87
7.6 Supporting information	87
7.7 Acknowledgements	88

8 General Discussion	89
8.1 Active sensing : a tricky and yet manageable situation.	89
8.2 <i>Ushichka</i> as a model dataset for group active sensing	89
8.3 Lowering the technical barriers to study group echolocation . . .	90
8.4 Sensory systems	90
8.5 Inverse problems all the way	90
9 Acknowledgements	91
10 Author contributions	95

Summary

English English English English English English English English English English

Zusammenfassung

*Bulla! I know not who I am,
 Bulla! I know not who I am
 ...
 Nor am I the believer in the mosque,
 Nor am I in the rituals of the infidel,
 Nor am I the pure in the impure,
 Nor am I inherent in the Vedas
 ...
 Nor am I of the water nor of the land,
 Nor am I fire nor air,
 Bulla! I know not who I am,
 Nor am I Arabic nor from Lahore,
 ...
 Nor did I name myself,
 Beginning or end I know just the self,
 Do not acknowledge duality,
 There's none wiser than I*

Who is this Bulla Shah?

*Bulla! I know not who I am,
 Bulla! I know not who I am*

- Bulla Shah, Sufi mystic and poet ¹. Translation by Sonyia Jamal (2017).

Echolocation is still a topic that resists neat classification into any conventional category of science. It ramifies into widely different scientific disciplines including ethology, mammalogy, physiology, psychology, acoustics, and the mathematical theory of signal detectionl. The results of new investigations have reemphasized the necessity of a broad biological perspective for any serious appreciation of echolocation.

- Donald Griffin, Listening in the Dark

¹I interpret this poem as an apt expression of the simultaneous pleasure, confusion and contentment gained from inter-disciplinary research

1

General Introduction

Animal sensory systems and behaviours evolve in response to a host of selection pressures such as predators, food and the environment (?). Moths with ultrasonic hearing are able to evade echolocating bats, that have in turn developed echolocation strategies that evade moth hearing (?). The physical environment (eg. light, heat, humidity) also plays a strong role in selecting sensory systems. Cave fish inhabiting dark cave systems repeatedly evolve a loss of vision (Jeffery, 2001), while ABCDEFGHIJKLMNOP. The social environment, consisting of other conspecifics and interactions with them, is also of great importance in the life-cycle of most sexually reproducing animals. Even classically solitary animals such as tigers and spiders will (or seek to) encounter mates over the course of their life time. Sexual selection moulds behaviours and sensory systems as seen in territorial displays and female sensory drives in a wide variety of taxa (?) as epitomised by the dance and colourful patterns of the male peacock spider (?).

The social environment of animals in long-term social groups (eusocial insects and mammals) or short-term groups (eg. lekks, maternal colonies) is particularly filled with a continuous stream of multi-modal stimuli and presents a host of challenges that individual sensory systems and behaviours have evolved with. To illustrate this point, we only need to look at ourselves as a species. The sensory challenge of conducting a conversation with a person in the middle of a gathering lies in trying to listen to what the person opposite is saying, and in trying to make ourselves heard. This problem has been termed the ‘cocktail party problem’ (Cherry, 1953) and our own sensory systems and behaviours show a variety of responses to such situations. As ‘emitters’ of speech we talk louder in the presence of background noise (the Lombard effect @). As ‘receivers’ there are a host of auditory and attentional processes that come into play. Our pinnae provide directional cues to allowing us to hone into sounds coming from the relevant direction (Yost, 2007) while the auditory centers in the brain are able to selectively choose the best speech signals (with the higher signal-to-noise ratio)

from either cochlea (Brungart and Iyer, 2012). Many other taxa such as birds, frogs and crickets face very similar cocktail-party like issues (Bee and Micheyl, 2008; Brumm and Slabbekoorn, 2005), and show a diversity of strategies to solve the challenges of dealing with emitting and detecting signals that overlap in time and are very similar spectrally. As emitters, animals may attempt to avoid signal overlap wherever possible, for instance in certain species of frogs and crickets, pairs of males may alternate their calls to reduce temporal overlap (Bee and Micheyl, 2008; Hartbauer and Römer, 2016), while in other species males synchronise their calls to increase the conspicuousness of the emitted signal and attract females over a longer distance range (Greenfield, 2015). As receivers, animals in groups need to be able to identify and separate signal sources and move in response to them. In certain vertebrates such as penguins, bats and seals, young are raised in large maternal colonies. Mothers leave their young ones in these colonies as they go to forage and return afterwards. A common strategy seen for mother-young reunion is the presence of individual-specific contact calls (and the ability to detect the inter-individual variation) in either the mother or the pup (Bradbury and Vehrencamp, 1998). The task of detecting a pup's or mother's call in the midst of a few hundreds to thousand other individuals' calls is no trivial task, and mistakes do occur: 17% of observed mother-pup pairs in a Mexican free-tailed bat (*Tadarida brasiliensis*) colony (McCracken, 1984) were found to be unrelated genetically.

Animal sensory systems are more than static 'receivers' or 'emitters', but exist in a dynamic interaction with the outside world through sensorimotor loops that form the basis of behaviours such as walking, running, flight (@@@ ?). Individuals in groups must also deal with the consequences of having conspecifics around even in simple tasks such as movement and path-planning. In contrast to moving in surroundings with stationary objects and predictable trajectories, moving among conspecifics implies anticipating and updating one's own direction of movement constantly. The apparent ease with which individuals in a murmuration of starlings or school of fish move belies the complex underlying sensorimotor loops that drive the movement of each individual in the group. These collective behaviours can be replicated in computational agent-based models with abstract and simple behavioural heuristics based on the relative proximity and position of each agent's neighbours (Reynolds, 1987; Couzin et al., 2002). The behavioural heuristics in these models mimics what animals in a group are likely to be doing (Herbert-Read, 2016), and broadly consists of the following cycle: 1) the individual assessing the relative position and movement direction of its neighbours, 2) using a decision criterion to integrate information on the neighbour positions and movements and 3) moving in a direction using the integrated information.

A key step in the initiation and maintenance of any collective behaviour is neighbour detection through the sensory modalities available to the animal. Many of the collective movements studied experimentally to date have been in visually dominant animals (Pita et al., 2016) such as birds, fish, mammals and flies. The fact that each animal acts as an independent receiver of light,

and does not affect the amount of light received by others majorly (aside from light obstruction due to body opacity) means that visually driven collective behaviour can scale well with group size. Given sufficient ambient light, individuals in all sizes of groups (small to large) are expected to independently be able to detect their neighbour's and thus show co-ordinated movement. This expectation is supported by experimental observations of thousand strong starling flocks (Ballerini et al., 2008), schools with ten-hundreds of fish (Tunstrøm et al., 2013), or even humans as we aggregate in small to very large groups (Moussaïd et al., 2011; Dyer et al., 2009). Concrete support for the use of visual cues, and their role in modulating individual decisions in group movement is provided by studies that reconstruct individual sensory *umwelts* (Turner and Penn, 2002; Strandburg-Peshkin et al., 2013; Belden et al., 2019). In contrast to simple models that assume all neighbours in a fixed sensory volume can be detected (Reynolds, 1987; Couzin et al., 2002), sensory reconstruction approaches include the details of neighbour positions and visual acuity (Strandburg-Peshkin et al., 2013; Belden et al., 2019), and thus allow a detailed analysis of individual sensory inputs and motor outputs.

1.1 Active sensing in groups : a tricky proposition

Broadly classified, sensory systems are of two types: passive and active. Passive sensing systems such as vision and olfaction are receivers that ‘take in’ stimuli from the environment. Active sensing animals such as bats, cetaceans, cave swiftlets and weakly electric fish(*sensu strictu* Nelson and MacIver (2006)) emit pulses of energy and perceive the environment through modulations of this pulse. In this respect, active sensing differs from passive sensing as animals generate their own stimuli to sense the world around them. The fact that animals must constantly emit pulses of energy to detect their surroundings raises cocktail-party like problems for individuals in groups. Electric fish detect objects through fluctuations they cause in the electric field along their body (?). When multiple electric fish are in close proximity, they will mutually disrupt each other’s ability to detect objects due to the similarity in the emitted electric field properties of individuals (Nelson and MacIver, 2006). This mutual disruption has been termed ‘jamming’, inspired by the technical term used in RADAR. Jamming is a common phenomenon expected in all active sensing animals irrespective of the type of energy that forms the probe (sound, electric field, etc.).

In the face of jamming from other active sensing conspecifics or experimental treatments, individuals show ‘jamming-avoidance responses’ (Bullock et al., 1972). Bullock et al. (1972), while working in electric fish that emit ‘wave-type’ electric signals. Wave-type electric fish continuously emit electric discharges at an individual-specific frequency (Heiligenberg, 2012). When faced with an experimental signal that is close to their own emitted frequency, wave-type fish shift their discharge frequency to prevent spectral overlap and deterioration in electrolocation. Another group of electric fish, the ‘pulse’ type fish, emit

short electric discharges at intervals. Pulse type electric fish resort to increasing their discharge rate, or adjusting discharges to avoid overlap when faced with interfering experimental pulses (Heiligenberg, 2012). Studies of jamming avoidance in electric fish have helped form the basis of group active sensing, but have remained to my knowledge primarily focused on individual-level responses to experimental treatments, studies with multiple electric fish in the field are recently on the rise (Henninger et al., 2020). Another important aspect that may minimise the overall extent of jamming in weakly electric fish is the limited spatial range of their electrolocation (Nelson and MacIver, 2006).

Echolocating bats emit loud ultrasonic calls and detect objects by listening for the echoes reflected by them.

1.1.1 Low and high duty-cycle echolocation

1.2 The nascent field of group echolocation

- grand history. starting with spallanzani, moving onto the first ‘proof’ by Griffin and etc - the focus has always been on how individual bats managed to do stuff.
- The question of how bats detect echoes in the midst of noise is however not very new. Griffin and the crazy noise room experiment with 50/80? speakers spread all around the room.
-
- much like electric fish - two types of bats, the low-duty-cycle FM bats like the pulse type, and the high-duty cycle CF-FM bats
- Mostly studies on low duty-cycle FM bats, Both of these species show different responses to
- Questions of why they do what they do are open
- CF-FM bats remain understudied due to the difficulty in analysing long overlapping sounds
- Given that they have lower sensory inputs - how do they manage to show coordinated behaviours in huge groups???

1.3 Science-technology loops in echolocation

The very discovery of echolocation could only happen because a biologist and a physicist used the then newly developed ultrasonic sensor and pointed it at

bats. Without stressing too much on the disciplinary labels of the researchers themselves, it remains that studying echolocation has meant using technological tools to make this alien sensory modality studiable to us humans.

1.3.1 Active sensing in bats and in noise

1. Echolocating bats emit loud calls and listen for echoes –
2. The returning echoes serve as packets of information that the bat uses to gain insights about its surroundings. The time delay, echo level, interference and spectral properties and other blah blah tell the bat more about the objects.
3. Echolocation might work based on a type of efference copy based mechanism.
- 4.

1.3.2 Focus on open and reproducible science

Uploading all my data, documenting code and publishing code as packages with publicly available tutorials are part of my attempts to share the science I do as much as possible.

1.4 Thesis outline

In this thesis I will investigate the sensory challenge faced by active sensing animals in groups using a combination of experimental and theoretical methods. Through modelling I will reveal and quantify the sensory challenge FM echolocating bats face in groups of various sizes. I will then present the results of an observational study detailing the echolocation strategies of CF-FM bats in the presence of conspecifics, followed by a novel experimental dataset and workflow to study groups of echolocating bats in the wild. I will end with a series of methodologically driven reports that will significantly reduce the effort needed to perform acoustic tracking at scale, help users simulate the expected performance of their arrays and ease the computational reproduciblity of echolocation call analysis.

Echolocation is expected to become increasingly infeasible with growing group sizes due to masking. In response to masking in the form of artificial playbacks and conspecific calls, bats show a variety of changes to their echolocation behaviour. While many experimental studies have reported the changes bats show as ‘emitters’ of sound, very little focus had been given to date on bats as ‘receivers’. Computational work modelling group echolocation has been

restricted to models that are qualitatively based on echolocation, with minimal biological details. In chapter 2 I present an experimentally parameterised computational model that estimates the sensory inputs a bat may receive as it echolocates in a group. The sensory inputs are quantified in terms of the number of echoes detected. Each detected echo corresponds to the detection of one neighbour in the group. I formulate a conceptual framework to estimate echo/neighbour detection in the presence of masking sounds. By quantifying echo detection from small to large groups, I find that even in group sizes of 200 bats, individuals may still be detecting one neighbour every few calls. When alone, a bat detects echoes reflecting off all objects in its surroundings. In contrast to echolocating when alone, bats in groups thus face a much lower ‘update rate’ in groups. Despite this lowered sensory update rate, I argue that the detection of occasional echoes may provide the sensory basis for the impressive feats of collective behaviour bats show in roosting sites, mating swarms and emergences.

In Chapter 3 takes us into our first experimental study looking at echolocation in groups of horseshoe bats.

In Chapter 4 (Uschichka), I do kat kat kat kat. From the time of Griffin & Galambos, who were the first to use the newly invented ultrasonic microphone to record ultrasound calls - the field of echolocation has always been a field that has ‘resisted’

Chapter 5 describes a methodology to automatically estimate microphone positions in an acoustic array. Multi-microphone arrays have been an important and indispensable tool in the study of echolocation in the laboratory ever since the work of Aubauer (1994). Working with microphone arrays in the field however comes with a series of challenges. Handling multi-microphone arrays in the field can be logically and physically straining: the long cables involved must be brought to order each time, the array frames themselves can be bulky and must be assembled on site each time. The frames themselves are often conspicuous and invite artifactual inspection behaviours from the animals themselves. In certain configurations the positions of each microphone may need to be measured each time the array is setup. While rewarding in terms of the data it generates, the process of setting up a microphone array can be strenuous especially when it comes to taking positional measurements. The use of TotalStation surveying systems that provide direct XYZ measurements are not common in the field (they are themselves another piece of bulky equipment). In contrast to acoustic arrays, camera arrays are much easier to setup and calibrate. Field-friendly protocols like ? mean the field biologist may freely place the cameras, record animal activity and then perform calibrations by moving a common object in front of the cameras. While looking for camera-type calibration workflows, I found the work of ? with their ‘Structure-From-Sound’ method. ? were able to automatically infer microphone positions using only the common sounds recorded across channels as input. I was able to initiate a collaboration with Prof. Åström (Uni. Lund) and convey my enthusiasm

for their method when applied to our arrays in the field. This chapter with Prof. Åström and colleagues shows the viability of automatic microphone position estimation in field conditions, using recordings I had made in the Orlova Chuka cave system. This chapter is but a beginning to the development of field-friendly workflows in acoustic tracking. The ‘Structure-From-Sound’ method used in this chapter and ? promise to significantly reduce the barrier to multi-microphone acoustic tracking for bioacousticians. For one it may completely eliminate the need to place microphones in specific geometries on conspicuous bulky frames and secondly, there will be no more need to perform tedious measurements before or after recordings. Being able to handle a larger number of microphones, the bioacoustician is now free to generate even higher resolution measurements of animal sounds. The ease and anticipatory joy of such an effortless acoustic tracking workflow can only be described by the words of a senior colleague¹ who said knowing there were such methods in development was a ‘Christmas gift’ in itself.

In Chapter 6 I present the **tacost** software package to aid the design and analysis of acoustic tracking systems. The accuracy of acoustic tracking is affected by multiple parameters: array geometry, source sound position, signal-to-noise ratios and the analytical approaches used to process and calculate source sound position itself. Experimental verifications of tracking systems is time and labour intensive, while mathematical analysis of array configurations are by their nature restricted to specific array geometries. Simulated audio data presents a quick and simple method to estimating the real-world performance of acoustic tracking systems either before or after experiments. **tacost** generates multi-channel audio data according to user-specified array geometry, source sound and source positions. **tacost** can be used to plan array geometry to optimise tracking accuracy before experiments, or post-hoc to estimate the maximal accuracy of the tracking system used. I detail the use of **tacost** to estimate patterns in tracking accuracy across space with two types of experimentally used array systems, the planar ‘tristar’ array with microphones placed regularly on a frame, and an array configuration used in Chapter 4 & 5 with freely-placed microphones spread around a volume. While **tacost** does not itself perform any acoustic tracking, it serves as a useful tool to estimate baseline performance for bioacousticians working with acoustic tracking. The package is written in the non-proprietary Python language, and released with an open-source license. To ease its use, the package also has detailed online and offline documentation

In Chapter 7, I present the **itsfm** software package originally written to perform the systematic and reproducible analysis of echolocation calls detailed in Chapter 3. CF-FM calls appear like staple-pins when viewed on a spectrogram. The ‘rising’ and ‘declining’ portions form the frequency-modulated (FM) parts, while the flat portion, forms the constant-frequency (CF) portion. The two call portions serve different sensory functions, and

¹here’s to the encouragement of Lasse Jakobsen

can be altered independently. To date, studies have adopted a variety of automated and manual methods to segment and then quantify the alterations in the CF and FM call portions. Manual methods of call segmentation are not reproducible and suspect to experimenter bias, while the automated software based methods remain custom scripts whose implementation is briefly detailed in the publications themselves. Publicly available implementations are important for the scrutiny and comparison of published methods. Purely description based implementations can show important differences from original implementations. Additionally, none of the software based methods include assessments of segmentation accuracy. `itsfm` implements a commonly described method (that I call the *peak-percentage* method) to segment CF-FM calls, and introduces a new and more reliable method (the *pwvd* method) to segment CF-FM calls. I also create a synthetic dataset of CF-FM calls, and compare the performance of the *peak-percentage* and *pwvd* algorithms in their segmentation accuracy. Results show that the *pwvd* method is overall superior to the *peak-percentage* method, and is thus the recommended algorithm to use for segmenting CF-FM calls. The *pwvd* method, unlike the *peak-percentage* method, is also potentially applicable to vocalisations of any kind as it does not rely on a specific spectro-temporal shape assumption. The `itsfm` package is written in the non-proprietary Python language, and released with an open-source license. To ease its use and encourage the further development of the package, detailed online and offline documentation has been made available.

2

Modeling active sensing in groups of bats reveals echo detection even in large groups

This chapter was published as a peer-reviewed paper in the Proceedings of the National Academy of Sciences of the United States of America:

Beleyur, T., & Goerlitz, H. R. (2019). Modeling active sensing reveals echo detection even in large groups of bats. Proceedings of the National Academy of Sciences, 116(52), 26662-26668.

TODO STUFF:

- Some references are still in NUMBERS!!!
- Check if all figure references are correct
- Fix multi paragraph caption problem in Figure S5
- Table S1 (masking data publications) - needs to fit into the page
- and Table S2 (logistic regression results) needs to fit into the page
- Check that all 'Section' headers are correctly referenced

Abstract

Active sensing animals perceive their surroundings by emitting probes of energy and analyzing how the environment modulates these probes. However, the probes of conspecifics can jam active sensing, which should cause problems for groups of active sensing animals. This problem was termed the cocktail party nightmare for echolocating bats: as bats listen for the faint returning echoes of their loud calls, these echoes will be masked by the loud calls of other close-by bats. Despite this problem, many bats echolocate in groups and roost socially. Here, we present a biologically parametrized framework to quantify echo detection in groups. Incorporating properties of echolocation, psychoacoustics, acoustics, and group flight, we quantify how well bats flying in groups can detect each other despite jamming. A focal bat in the center of a group can detect neighbors in group sizes of up to 100 bats. With increasing group size, fewer and only the closest and frontal neighbors are detected. Neighbor detection is improved by longer call intervals, shorter call durations, denser groups, and more variable flight and sonar beam directions. Our results provide a quantification of the sensory input of echolocating bats in collective group flight, such as mating swarms or emergences. Our results further generate predictions on the sensory strategies bats may use to reduce jamming in the cocktail party nightmare. Lastly, we suggest that the spatially limited sensory field of echolocators leads to limited interactions within a group, so that collective behavior is achieved by following only nearest neighbors.

2.1 Introduction

Active sensing animals use self-generated energy to sense their surroundings by analyzing how objects around them change the emitted energy (Nelson and MacIver, 2006). Bats emit loud ultrasonic calls and detect objects around them by listening to the echoes (Fenton, 2013a, Griffin (1958a))reflected off these objects. Active sensing is an effective sensory modality when the animal is solitary. However, when multiple active sensing animals emit pulses of energy in close proximity, they may “jam” each other and mutually interfere with their ability to detect objects in their environment (Nelson and MacIver, 2006, Matsubara and Heiligenberg (1978)). If groups of echolocating bats mutually jam or mask each other, they would not be able to detect each other. Due to the intense jamming, individuals would have a progressively difficult time detecting the echoes reflecting off their neighbors, and thus not detect their neighbors at all. Without detecting each other, groups of individuals cannot show collision-free flight. However, many bat species are very gregarious, and fly and echolocate together in groups of tens to millions of bats. Bat groups also show coordinated behaviors in cave flights, evening emergences, and mating swarms (Ortega, 2016, Kunz (1982)). How is their ability to detect each other impaired by increasing group size? How many of its neighbors does a bat actually detect in the presence of intense jamming? What strategies may improve echo detection and thus neighbor detection when many active sensing animals are together? We present biologically parametrized simulations to answer how bats manage to echolocate in the face of intense jamming.

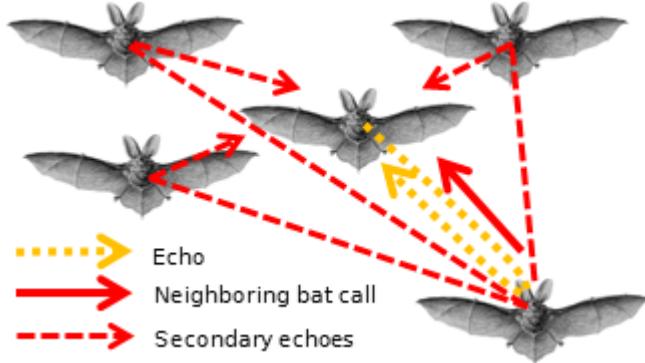
In human psychophysics, the sensory challenge of perceiving an auditory cue among other similar sounds has been called the “cocktail party problem”(Cherry, 1953, Bee and Micheyl (2008)). When applied to bat echolocation, the cocktail party problem has been elevated to the “cocktail party nightmare”, given the high repetition rate, similarity, and amplitude of echolocation calls. On top of these factors is the nonlinear increase in the number of masking sounds with increasing group size (Ulanovsky and Moss, 2008). Empirical studies to date have investigated the cocktail party problem from a sender’s perspective (Bee and Micheyl, 2008, Ulanovsky and Moss (2008), Brumm and Slabbekoorn (2005)). Through field observations, playback studies, and on-body tags (Amichai et al., 2015, Cvikel et al. (2015a), Gillam et al. (2010),Gillam and Montero (2016),Lin et al. (2016),Ulanovsky et al. (2004),Habersetzer (1981), Jones et al. (1993),Jarvis et al. (2013),Adams et al. (2017),Falk et al. (2014),Surlykke (1993)), we now know a range of echolocation strategies that bats show under challenging acoustic conditions. Bats can increase their call intensity, alter their call duration and frequency range, or suppress calling in the presence of conspecifics and noise playbacks (Amichai et al., 2015, Adams et al. (2017),Tressler and Smotherman (2009),Møhl and Surlykke (1989)). In contrast to the many reports of bats’ responses to noisy conditions, very little work has been done in conceptually understanding how receiver strategies might contribute to dealing with the cocktail party nightmare (Lin and Abaid,

2015, Perkins et al. (2017)). To our knowledge, biological modeling of the cocktail party nightmare from a receiver’s perspective that includes the details of bat echolocation and auditory processing is lacking. We fill this gap in conceptual understanding by presenting a biologically parametrized model based on the known properties of bat audition and the acoustics of a multi bat echolocation scenario. We quantified how well a bat flying with conspecifics can perceive its neighbors in terms of the returning echoes it detects. Through our simulations, we arrive at a sensory estimate of what a bat in the cocktail party nightmare may be detecting, if anything at all.

2.2 Materials and methods

We model the echolocation of frequency-modulating (FM) bats. The calls of FM bats are typically downward frequency-modulated and of short duration (≤ 5 ms). Each call is followed by a longer silence (80–150 ms) called the interpulse interval (Jones et al. (1999)). FM bats thus sense their world “stroboscopically” by emitting a call and listening for the echoes returning during the interpulse interval (Griffin and Galambos (1941)). In the absence of any loud conspecific calls, a bat is able to hear all returning echoes and thus to detect all objects around it. However, in the presence of other loud bat calls, some of its own returning echoes may be masked. In that case, the bat will hear a few or none of the returning echoes. This corresponds to the bat detecting a few or none of the surrounding objects. In the cocktail party nightmare the “objects” each bat is trying to detect are its neighbors

Our model of the cocktail party nightmare is designed to describe the auditory scene (Ulanovsky and Moss (2008)) of a bat emerging from a cave in a group as it echolocates on the wing. A focal bat flying in a group of N bats may detect up to $N-1$ of its neighbors (excluding itself), which is equivalent to hearing $N-1$ returning echoes. The focal bat receives 2 kinds of loud masking sounds that interfere with the detection of its neighbors: 1) the $N-1$ loud calls emitted by other bats in the group, and 2) the secondary echoes created by the call of a neighboring bat, reflecting once off another bat, and arriving at the focal bat. Every neighboring bat call generates $N-2$ secondary echoes, meaning that the focal bat can receive up to $N-1 \times N-2$ secondary echoes (Figure 2.1). We implemented a spatially explicit 2-dimensional (2D) simulation of bat echolocation, sound propagation, and sound reception and include mammalian auditory phenomena to quantify how many and which neighbors a bat can detect in the sonar cocktail party nightmare. We then explored how changes in group size and in sender strategies affect neighbor detection in a group.



2.1: Schematic of the cocktail party nightmare. Arrows indicate the different types of sounds received by a focal bat: it needs to hear the echoes returning from its own calls (orange) to detect its neighbors, despite the masking by the calls of neighboring bats (solid red) and their secondary echoes (dashed red). Here, only 1 target echo off a single neighbor, only 1 representative neighboring bat call, and its set of secondary echoes are shown. In total, for a group of N bats, the focal bat will receive $N - 1$ echoes, $N - 1$ neighboring bat calls, and $N - 1 \times N - 2$ secondary echoes. Bat image courtesy of Wikimedia Commons/Ernst Haeckel.

2.2.1 Model Scenarios

We ran 2 model scenarios to test the effect of 1) increasing group size and of 2) variation in call parameters, group geometry, and acoustic parameters on neighbor detection. In all models, we used the central-most bat in the group as the focal bat.

- Scenario 1: *Effect of group size on neighbor detection:* We simulated groups of 5, 10, 30, 50, 75, 100, and 200 well-aligned bats with identical echolocation and hearing properties flying at a minimum inter-bat distance of 0.5 m (Table 2.1 for full model parameters). The number and location of neighbors detected by the focal bat were recorded in every simulation run.
- Scenario 2: *Effect of call parameters, group geometry, and acoustic parameters on neighbor detection:* Here, we varied other parameters relevant to the cocktail party nightmare (Table 2.1) while keeping group size constant ($N_{bats} = 100$, i.e. the largest group size from Scenario 1 with a biologically relevant neighbor detection rate). We varied call parameters (interpulse interval, call duration, source level), group parameters (heading variation, minimum inter-bat spacing), and acoustic parameters (atmospheric absorption, acoustic shadowing).

2.1: Model parameters for both model scenarios: Scenario 1 modeled the effect of group size, while other parameters were fixed, resulting in 7 parameter combinations (1 per group size). Scenario 2 modeled the effect of other relevant parameters, while group size was kept constant at 100 bats, resulting in a combined set of 1,200 parameter combinations.

Parameter	Scenario 1: Effect of group size	Scenario 2: Effect of call parameters, group geometry, and acoustics
Group size	5, 10, 30, 50, 75, 100, 200	100
Interpulse interval (ms)	100	25, 50, 100, 200 ,300
Call duration (ms)	2.5	1, 2.5
Source level (dB SPL re 20 μ Pa at 1m)	100	94, 100, 106, 112, 120
Minimum interneighbour distance (m)	0.5	0.5,1.0
Group heading variation ($^{\circ}$)	10	10, 90
Atmospheric attenuation (dB/m)	-1	0, -1, -2
Acoustic shadowing	Yes	No, Yes

2.2.2 Model Implementation

Each model run simulated 1 interpulse interval of the focal bat, and we calculated the timing and received level of all sounds (target echoes, masking calls, and secondary echoes) that arrived at the focal bat during that interpulse interval. Each model run simulated a series of sounds that arrived during an interpulse interval following the focal bats' call, based on a spatially explicit distribution of a group of bats (SI Appendix, Schematic 2.6.5.1). At the beginning of every model run, N bats were placed in a 2D space with randomly assigned heading directions (SI Appendix, 2.6.1.6 and 2.6.1.7). For each neighboring bat, we calculated its angle and distance to the focal bat. The received level was calculated based on a common source level for all bats, spherical and atmospheric spreading over each call's and echo's travel distance, and acoustic shadowing. Acoustic shadowing is the reduction in received level of a sound due to obstructions in its path. A sound in the cocktail party nightmare may pass around obstacles (other bats) as it propagates from source to receiver. The reduction in received level was measured and calculated as a linear function of the number of bats obstructing the path between source and receiver (SI Appendix 2.6.1.9). For target and secondary echoes, we also considered monostatic and bistatic target strengths measured in this paper (SI Appendix, 2.6.1.8).

The arrival time of target echoes within the interpulse interval was determined according to the 2-way travel time to the echo-reflecting neighboring bat. The

arrival time of masking calls and secondary echoes was assigned randomly with uniform probability across the interpulse interval. The random arrival time assignment of calls and secondary echoes recreates the uncoordinated echolocation of all bats in the group. It is unlikely that multiple bats in large groups can coordinate their calls effectively, and independent calling has been reported even in small groups of 4 bats (Hase (2018)).

All bats in a group were identical in their calling properties, and we treated all sounds as constant tones of equal duration, i.e., we did not explicitly model spectral emission, propagation, and reception properties. The only difference between each of the sounds was their path and source of sound production. The omission of spectral properties is a conservative choice that assumes maximal masking of the primary echoes, thus allowing us to study the role of intensity differences and temporal separation between target echoes and masking sounds.

Once we calculated the timing and received level of all sounds at the focal bat, we accounted for directional hearing sensitivity (SI Appendix, Figure 2.7) and spatial unmasking. Spatial unmasking describes the reduction in experienced masking as the arrival angle between masker and target sound increases (Ebata (2003), Sümer et al. (2009)). We simulated spatial unmasking by the reduction of a masker's effective received level based on its angular separation to an echo. For each echo, the same masker will have a different effective masking level as its relative angle of arrival will be unique for each echo. We thus calculated the effective masking level of each masker for each echo. The effective masking levels of all maskers were then combined to form a time-variant and echo-specific “masker SPL profile” (SI Appendix, Figure 2.9D). This is essentially the joint sound pressure level (SPL) of all maskers over time. We then expressed this echo-specific masker SPL profile in relation to the echo's SPL, thus obtaining a relative “echo-to-masker ratio profile” (SI Appendix, Figure 2.9E). This is equivalent to a signal-to-noise ratio profile, where the echo is the signal and the masker profile is the noise.

In addition to angular separation, signal detection is also determined by the temporal separation between signal (echo) and masker (Møhl and Surlykke (1989), Yost (2007), Siewert et al. (2004)). Masking increases as the masker arrives closer in time to the echo. Masking occurs over longer durations when maskers arrive before the signal (forward masking) than afterward (backward masking). We recreated the asymmetric masking by a “temporal masking envelope” temporally centered at the echo (SI Appendix, Figure 2.5). The echo was considered heard if the echo-to-masker ratio profile was above the temporal masking envelope. We allowed short drops of the echo-to-masker ratio profile below the temporal masking envelope, for a combined maximum duration of less than 25% of an echo's duration. Alternatively, we defined an echo to be masked (= not heard), if the echo-to-masker ratio profile was below the temporal masking envelope for more than 25% of the echo duration. The 25% threshold was an arbitrarily chosen conservative value to prevent masking by rare and short bursts of high sound pressure level that are unlikely to affect

echo detection biologically (SI Appendix, 2.6.2.7).

2.2.3 Model Parametrization

We implemented a detailed set of echolocation, group and sound properties in our model, including call and hearing directionality, spatial unmasking, temporal masking, group geometry, and details of sound propagation. These properties were parameterized based on published results wherever available. Acoustic shadowing and target strengths (monostatic and bistatic) of bats were specifically measured for this work. All details of the model parameters including our respective measurements and on model implementation are presented in the Supporting Information.

2.3 Results

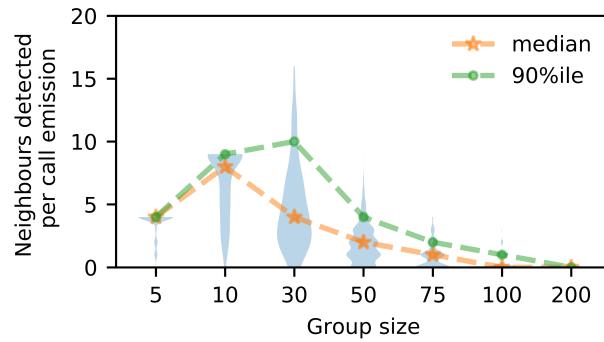
2.3.1 Effect of Group Size on Neighbor Detection

At group sizes of 5 and 10, the focal bat hears the echoes of most or all of its neighbors per call (median: 4 and 8 echoes per call at $N=5$ and 10, respectively; Figure 2.2). At progressively larger group sizes, the median number of detected neighbors drops from 4 to 0 at group sizes of 30 to 200. Yet even in a group of 100 bats, while the median number of detected neighbors is zero, the 90th percentile is 1, showing that a neighbor is not detected with each call, but occasionally. Beyond a group of 100 bats, the focal bat typically detects no neighbors at all. The initial rise in detected neighbors in groups of 5 to 30 bats is primarily caused by the increased number of neighbors that could be detected, which is soon counter-acted by the intense masking that rises non linearly with group size.

We next derived the probability of detecting at least 1 neighbor per call, which describes the average rate of neighbor detection (Figure 2.3A, blue). At smaller groups of 5 to 30 bats, the focal bat detects at least 1 neighbor per call at above 0.95 probability. At larger group sizes (50 to 100), the probability of detecting at least 1 neighbor drops rapidly to 0.3 per call in a group of 100 bats, and is basically zero for a group of 200 bats (0.004 probability). A bat (with 10 Hz calling rate) flying in a group of 100 bats will thus detect at least 1 neighbor around 3 times per second (~ 3 Hz detection rate), while a bat flying in a group of 30 bats will detect at least 1 neighbor almost every time it calls (9.5 Hz detection rate). The probability of detecting multiple bats per call is lower than just detecting at least 1 bat (Figure 2.3A). Yet, even in a group of 50 bats, the focal bat has a probability of detecting at least 2 and 4 neighbors per call of about 0.5 and 0.1, respectively.

We next quantified which neighbors the focal bat detects. Detection is generally limited to nearby neighbors (Figure 2.3B) and, with increasing group size, to

neighbors in front of the focal bat (Figure 2.3C). At a group size of 30 bats, the focal bat occasionally detects neighbors that are up to 2 m away in radial distance, which is the furthest neighbor distance. With increasing group sizes, despite the group being more spread out, the focal bat can only detect its nearest neighbors (e.g., neighbors at ~0.5 m in a group of 200 bats; Figure 2.3B). In the azimuthal plane, at small group sizes, the focal bat initially detects neighbors all around it (95%-neighbor detection angle range 237° for up to 50 bats; Figure 2.3C). With increasing group size, a frontal bias in neighbor detection appears (95%-neighbor detection angle range: 191 to 35° for 100 and 200 bats; Figure 2.3C).

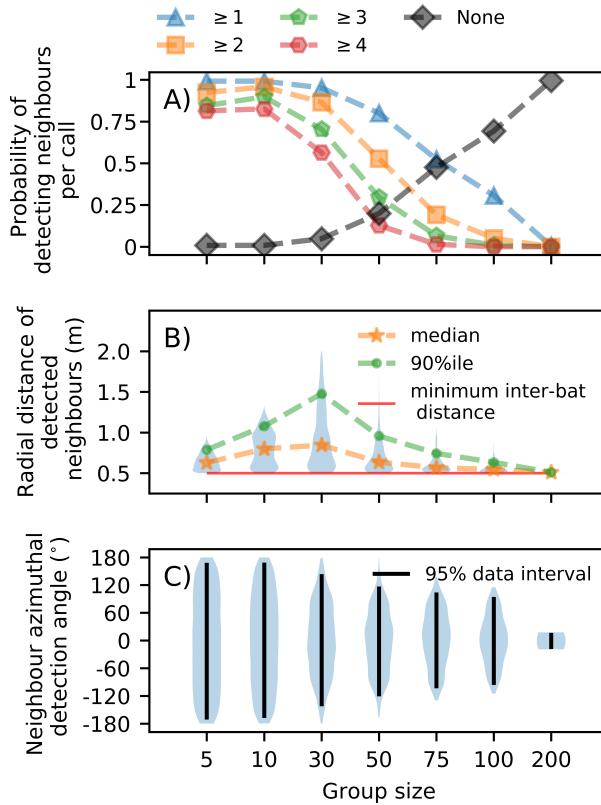


2.2: Number of detected neighbors per call by a focal bat in the center of a group. The initial rise in the number of detected neighbors is because there are indeed more neighbors and the degree of masking is low. However, with increasing group size, most of the neighbors cannot be detected anymore, and progressively fewer neighbors are detected per call. Violin plots show the distribution of the number of neighbors detected per call, and their median (stars, orange) and 90th percentile (dots, green).

2.3.2 Effect of Call Parameters, Group Geometry, and Acoustic Parameters on Neighbor Detection

We next analyzed how variation in call parameters, group structure, and acoustic parameters affected neighbor detection. We fixed the group size to 100, as at this size, the focal bat could typically detect at most 1 neighbor (90th percentile, Figure 2.2) at 0.3 probability (Figure 2.3A) per call. We thus reduced the output of each simulation run to a binary neighbor detection score of 1 (detection) or 0 (no detection). We analyzed the effect of each parameter on neighbor detection with a logistic regression, treating all parameters as categorical and using their value in Scenario 1 as reference (parameter range in Table 2.1).

The call parameters interpulse interval and call duration showed the strongest effect (Figure 2.4A,B and SI Appendix, Table ??). Increasing the interpulse interval from 100 ms to 200 and 300 ms increases neighbor detection probability by about 15 and 75 times, while reducing it to 50ms lowers neighbor detection

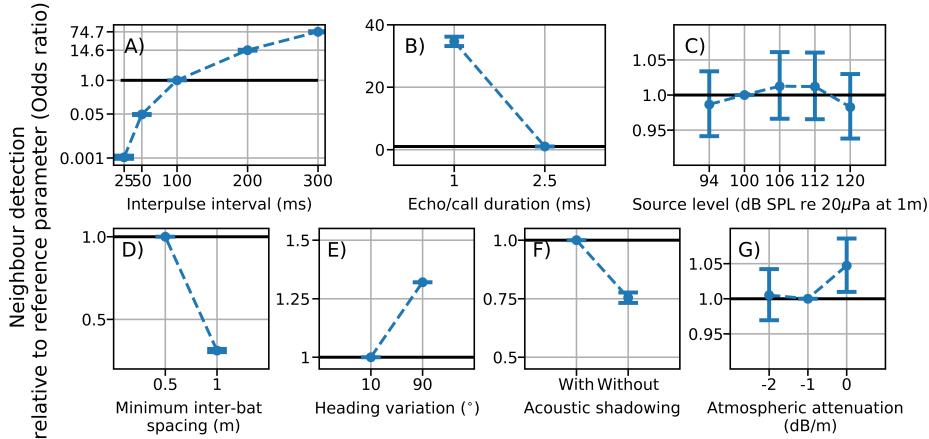


2.3: Characterization of the focal bat's perception. (A) The probability of detecting $\leq X$ neighbors per call ($X = 1, 2, 3, 4$, or none). Even in groups of up to 100 bats, the focal bat has a ~ 0.3 probability of detecting at least 1 neighbor per call. In even larger groups (200 bats), no neighbors are detected anymore. (B) With increasing group size, a focal bat only detects its closest neighbors. Initially, the radial distance of detected neighbors increases because the spatial extent of a group increases with group size (at 5, 10, 30 bats: radius = 0.75, 1.12, 1.97 m), but it then drops down to the nearest neighbors beyond 30 bats. (C) The azimuthal location of detected neighbors, showing an increasing frontal bias with increasing group size. Although neighbors were uniformly distributed in azimuth, the frontal bias of call and hearing directionality means that frontal returning echoes are louder than peripheral ones.

to 0.05 times Scenario 1 (Figure 2.4A). Shortening call duration from 2.5 ms to 1 ms led to 35 times higher neighbor detection (Figure 2.4B). Call source level had no effect (Figure 2.4C).

Group geometry also influenced neighbor detection probability, but less than changing call parameters. Flying at larger interbat distances of 1.0 m leads to 0.31 times lower neighbor detection compared to denser groups with 0.5 m interbat distance (Figure 2.4D). Groups where individuals head in a more variable direction have 1.32 times better neighbor detection than groups with a generally common heading (or echolocation beam) direction (Figure 2.4E).

Among the physical parameters, acoustic shadowing increased neighbor detection (without acoustic shadowing, neighbor detection is 0.75 times lower than with acoustic shadowing), while atmospheric attenuation had a negligible effect (Figure 2.4F and G)



2.4: Effect of call parameters (A–C), group geometry (D and E), and acoustic parameters (F and G) on neighbor detection. Each plot shows the probability of neighbor detection (model estimate and 95% confidence interval of odds ratio) when changing model parameters relative to the reference parameter used in the simulations of Scenario 1 (Table reftab:modelsenarios). Odds ratios above and below 1 indicate a higher and lower neighbor detection probability, respectively, indicated by the horizontal reference line. (A–C) Call parameters: Longer interpulse intervals (A) and shorter call durations (B) increase neighbor detection probability, while call source level (C) has no effect. (D and E) Group geometry: Neighbor detection is better in groups that are tightly packed (D) and with higher heading variation (E). (F and G) Effect of acoustic parameters: Acoustic shadowing by bats in groups improves neighbor detection probability (F), while atmospheric attenuation has a negligible effect (G)

2.4 Discussion

We present a conceptual framework to quantify what a focal bat experiences in the sonar cocktail party nightmare. We quantified the probability of detecting

neighbors across a range of group sizes, which allows calculating the rate at which a focal bat detects its neighbors. When flying alone, a focal bat will detect objects around it at a rate equal to its call rate, while in a group, its object detection rate is reduced due to masking. We show that even in a group of 100 bats, bats still detect at least 1 neighbor per call about 3 times per second (for a 10 Hz call rate), while in smaller group sizes, neighbor detection rate is larger at 5 to 10 Hz. Bat echolocation is generally “stroboscopic,” meaning that information is received intermittently with time gaps (Griffin (1958a)). We suggest that bats in smaller group sizes still experience a sufficiently high information update rate for performing collision avoidance and neighbor following. With increasing group size, perception might become “hyper-stroboscopic,” i.e., so scarce that different sensorimotor heuristics might be required to maintain group coordination.

The low level of masking at smaller group sizes allows the focal bat to detect all its neighbors per call. With increasing group size, however, the focal bat detects maximally 1 neighbor per call in a group of 100 bats. This neighbor detection rate of at least 1 neighbor per call even in large group sizes provides a formal sensory basis for group movement in active sensing animals. While a bat in a large group cannot track the position of all its neighbors, it still can track the movement of a few neighbors, specifically those close to and in front of it. This reduction in rate, range, and direction of detected neighbors has predictive consequences for the kind of collective behavior bat groups may show in nature. Many models of collective movement assume that each individual in a group detects the position and orientation of neighbors in the whole of its sensory volume, and then performs an averaging across all neighbors to decide its next movement (Couzin et al. (2002), Grégoire and Chaté (2004), Vicsek et al. (1995), Reynolds (1987)), leading to the impressive coordinated behaviors of fish schools and insect swarms (Sumpter (2006), Vicsek and Zafeiris (2012)). As the number of neighbors that an individual detects decreases, more “limited interactions” begin to dominate, causing anisotropy in the group structure (Bode et al. (2011), Ballerini et al. (2008)). For bats in the cocktail party nightmare, we predict that large groups may show higher anisotropy than smaller groups due to the limited number of neighbors that they can detect and react to. All things being equal, we predict that in large groups (>50 bats), the neighbors in the frontal field of a bat will have a disproportionate influence on its movement decisions. Bats in larger groups may thus maintain higher alignment with their frontal neighbors compared to bats in smaller groups.

Our simulations allow for a direct quantitative comparison of the effects of echolocation, group geometry, and acoustic parameters in group echolocation. Among the call parameters tested, reducing call rate (increasing interpulse interval) was most effective in increasing neighbor detection in jamming conditions, matching experimental evidence for reduced calling rate in *Tadarida brasiliensis* (Jarvis et al. (2013), Adams et al. (2017)). In contrast, other FM bat species increase their call rates in groups and background noise (Amichai et al. (2015), Lin et al. (2016), Cvikel et al. (2015b), Luo et al. (2015)). Likewise, our result that shorter call duration should improve neighbor detection is

opposite to experiments showing that most bat species increased call duration in the presence of maskers (11, 23, 24, 43, 44), except (Cvikel et al. (2015b)). Lastly, our result of no effect of changing source level on neighbor detection might also seem to differ from experimental data showing that bats in laboratory conditions do increase source level in the presence of maskers (Amichai et al. (2015), Tressler and Smotherman (2009), Luo et al. (2015), Hase et al. (2016)). While there might be species-specific variation, we suggest that these differences are mostly due to the experimental situation. Bats in these experiments experienced constant maskers. Calling more often, for longer, and louder thus improved the bats' signal redundancy, echo-to-masker ratio, and overall echo detection. In contrast, our model simulates group flight of many bats with simultaneous and uniform changes in their call parameters. When all bats in a group shorten call duration, this reduces the overall duration of masking sounds, thus improving echo detection. Likewise, when all bats in a group increase their call amplitudes to optimize their own echo-to-masker ratios, all bats will eventually call at their maximum, with no overall effect on neighbor detection. Analyzing bat calls in mass emergences is technically challenging and it remains unknown whether *T. brasiliensis* and other gregarious bat species reduce their call rate in the field.

Bat aggregations show a variety of structures across behavioral contexts, from well-aligned almost parallel flight during roost emergences, to more variable and less-aligned flight in mating swarms and when circling in limited cave volumes. We show that this group structure itself affects how well bats can detect each other. Bats detect their neighbors better in less-aligned groups compared to more aligned groups. During aligned emergence flight, the focal bat always receives loud forward-directed masking calls from bats behind it, in addition to the relatively loud side-calls emitted by neighbors to its left and right. In contrast, during less-aligned swarming flight, the relative orientation of the bats is more distributed and changing, with the focal bat experiencing a wider dynamic range of masker levels (i.e., louder and fainter masking calls originating from a wider range of angular directions). This increased dynamic masker range allows for better echo detection, as there will be drops in echo-to-masker ratios due to changing received masker level. This effect is beneficial for enabling swarming flight, as the collision risk in less-aligned flight is likely higher compared to the more aligned emergence flights. Inter-individual distance is another parameter of group structure, and we show that neighbor detection is better in dense groups. This might seem unexpected given that the received SPL of the maskers is higher the closer the bats are to each other. However, received echo levels are also higher when bats are closely spaced. Since echo SPL drops by 12 dB per doubling of distance, but masker call SPL only by 6 dB per doubling of distance, the echo-to-masker ratio is higher at shorter compared to longer inter-bat distances. It would be interesting to examine if perhaps large groups in the field actually fly closer to each other than smaller groups.

While we only modeled neighbor detection for the central-most bat in a group, its position in the group (e.g., central, frontal, or at the back) is likely to also

have an effect on the number and received level of maskers, and thus on the number of detected echoes. However, we expect the obtained trends to remain qualitatively the same regardless of focal bat position. Particularly, we predict that masking will increase with group size, and only the exact group size at which a given level of masking (e.g., X% neighbor detection probability) is obtained will change depending on the focal bat’s position in the group.

We furthermore show that it is important to consider bats not only as sources of reflected echoes and masking sounds, but also as obstructions to sound that actually alleviate the cocktail party nightmare. Typically, the detected echoes originate from nearby bats and are not shadowed. In contrast, the masking calls and secondary echoes can arrive from distant neighbors, thus passing by multiple other bats. Shadowing thus consists of the overall reduction in masker levels, which increases echo-to-masker ratios for the comparatively loud echoes returning from nearby neighbors.

Our results show that the cocktail party may not be as much of a “nightmare” as previously thought (Ulanovsky and Moss (2008)). We show that the modeled psychoacoustic, spatial, and acoustic properties act together to alleviate the nightmare into a challenge. When bats are flying in a multi echo environment, our results show that a bat will always hear some echoes after a call emission, and very rarely no echoes at all. This parallels the phenomenon of auditory “glimpsing” reported in the human auditory cocktail party where individuals may follow conversations by perceiving parts of detected speech rather than whole sounds (Miller and Licklider (1950)).

2.4.1 Improved Echo Detection in Real-World Situations

We present a first approximation to the sonar cocktail party nightmare, including many relevant biological, physical, and auditory mechanisms. Bats are expert echolocators and can detect echoes and fly under challenging conditions (Møhl and Surlykke (1989), Surlykke (1992), Petrites et al. (2009), Bates et al. (2008)). Bats rapidly adjust their call behavior in terms of their call duration, source level, and interpulse intervals (Luo and Moss (2017), Corcoran and Moss (2017)), integrate echoic information over multiple call emissions (Simmons and Gaudette (2012)), and actively track objects by aiming their calls at them (Ghose et al. (2006), Ghose and Moss (2006)). While we tested a range of different echolocation call parameters, our model implemented these parameters as fixed values that do not vary over time, thus lacking the dynamic nature of a real bat in the wild.

Furthermore, we did not model the spectral content of echo or masker sounds, and analyzed echo detection based on a fixed threshold of echo-to-masker-ratio. In contrast, real echolocation calls possess a time-variant spectral pattern that is species and even individual specific (Gillam et al. (2010), Yovel et al. (2009)), which can reduce echo masking. Masking is strongest when target

and masker overlap both in time and in frequency (i.e., fall within the same critical band of the auditory system) (Ebata (2003), Fletcher (1940)). The frequency-modulation of bat calls means that even when maskers and echoes partially overlap in time, they will not necessarily overlap in frequency, thus reducing the likelihood of masking. The individuality of bat calls may help a bat reject the secondary echoes from other bats' calls by forming separate auditory streams (Fay (2008)) for its own echoes and others' echoes. Given the scarcity of empirical data to parametrize the effect of spectral differences on echo detection in masking conditions, we did not include it in our model, thus simulating a conservative worst case scenario where all sounds lie in the same frequency band. Additionally, attentional processes strongly improve target detection by improving the signal-to-noise ratio in the presence of maskers with similar time-frequency structure (Hafter et al. (2008)). Under real-world conditions, it is likely that masking in groups is even less than simulated here.

Due to the scarcity of published data, the inter-individual and interspecific variation in the temporal and spatial masking functions used in our model is unknown. The temporal masking envelope will arguably be similar in many bat species, showing the typical mammalian pattern of worse target detection threshold with shorter temporal separation between target and masker. Spatial unmasking occurs through the nonlinear interaction of pinnae shape, cochlear and higher auditory processing (Ebata (2003), Culling and Stone (2017)). As pinna shape and associated acoustic receiver characteristics strongly vary in echolocating bats (Obrist et al. (1993)), this will lead to species-specific spatial unmasking and echo detection rates in the sonar cocktail party nightmare.

2.5 Conclusion

We provide a conceptual framework to explain how active sensing animals such as echolocating bats successfully navigate in groups despite mutually jamming each other. The intense jamming in groups might lead to individuals only detecting their nearest frontal neighbors, which might drive limited interactions within a group. We also show that call parameters and group geometry determine the challenge in the sonar cocktail party nightmare. Recent advances in on-body acoustic tags (Cvikel et al. (2015b), Stidsholt et al. (2019)), signal analysis (Aodha (2018)), and acoustic tracking (Seibert et al. (2015)) of echolocating animals in the field might facilitate future experimental validation of our model predictions. As our model formulation is not constrained to echolocation in bats, it can be parametrized to other echolocators such as oilbirds, swiftlets, and odontocetes (Brinkløv et al. (2013), Surlykke et al. (2014)) that also echolocate in groups and suffer from cocktail-party-like conditions.

2.5.1 Acknowledgements

We thank the members of the Acoustic and Functional Ecology Group for insightful comments and support, Claire Guérin for contributing to code development, the Max Planck Institute for Ornithology for excellent research infrastructure, and 2 anonymous reviewers for helpful comments. 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). The simulation runs in this paper were funded by a Google (Research Credits) grant to T.B.

2.6 Supplementary Information

2.6.1 Model parametrization

Here, we present the details of how we parametrized our model of the sonar cocktail party nightmare, based on empirical data of behavioral studies in bats and our own measurements.

2.6.1.1 Temporal masking envelope

We derived the echo-to-masker sound pressure level (SPL) ratios for forward, backward and simultaneous masking from two empirical target-detection studies in echolocating bats (Table 2.2). We only chose studies where the target and masker were co-located along the same direction. Both studies presented the ratio between the echo and masker SPL at target detection for various delays between echo and masker arrival time. We linearly interpolated (in a piecewise fashion) the echo-to-masker SPL ratios between each of the time delays measured in the studies to obtain the full temporal masking envelope ranging from -0.65 and +24 ms delay of the target echo relative to the masker edge (Figure 2.5).

2.6.1.2 Spatial unmasking function

Sümer et al. (2009) performed a backward masking study to address spatial unmasking in the bat *Eptesicus fuscus*. In a two-alternative forced-choice paradigm, they increased the angular separation between a target object and a masker object while also varying the level of the target object's echo (by varying the size of the target object and thus its target strength).

We define the spatial unmasking function as the echo-to-masker SPL ratio at just-noticeable echo detection as a function of angular separation. To obtain the echo-to-masker level ratios, we subtracted the target object's target strengths from the masker object's target strength. We normalized the spatial unmasking function to the co-localised echo-masker case (i.e., when both echo and masker arrive from the same direction). This describes the reduction in echo-to-masker ratio required for echo detection as a function of angular separation between target and masker, compared to the co-localized case. We digitized the data points from Figure 2.4B of Sümer et al. (2009) by hand with WebPlotDigitiser (Rohatgi, 2015) to obtain the target's target strength as a function of angular separation. Masker target strength was given by Sümer et al. (2009) as -14.5 dB. We then calculated the target-masker SPL ratios and interpolated them with a quadratic polynomial fit (Figure 2.6). The interpolated data was then further upsampled to 0.47° intervals. As Sümer et al. (2009) only measured angular separations up to 23°. We conservatively used the echo-masker SPL ratio at 23° also for all larger angular separations.

2.6.1.3 Call and hearing directionality

Echolocation calls have a directional beam shape, meaning that the emitted SPL generally decreases with increasing angular distance from the main call direction, which has the highest call SPL (Figure 2.7). The highest call SPL is typically towards the front, and reduces towards the back of the bat. Despite this directionality and additional variation with call frequency and behavioral context (Jakobsen and Surlykke, 2010, Jakobsen et al. (2013), Surlykke et al. (2012), Giuggioli et al. (2015)), bats still emit a significant amount of sound pressure into the backward direction. The average call SPL behind a bat is about 14 dB lower than in the forward direction (Giuggioli et al., 2015, Stidsholt et al. (2019)). Call directionality leads to a drop in the effective number of masking calls from neighbors as only those calls arriving from a limited range of directions will have sufficiently high SPL (Figure 2.7). For example, in an emergence situation with approximately parallel flight directions, the focal bat will receive the loudest calls from those bats flying behind it. Similarly, the lowest received call levels in an emergence will be from those bats flying in front of the focal bats.

Like call production, hearing is also directional. The pinna structure of a bat attenuates or amplifies the same sound depending on its direction of arrival (Firzlaff and Schuller, 2003, F et al. (2008)). In *Myotis daubentonii*, hearing directionality between 35-45 kHz leads to an average amplification of frontally arriving sounds by 4 dB in comparison to those arriving from behind. We used data of Giuggioli et al (2015) to describe the average call and hearing directionality of our modelled bats (Figure 2.7).

2.6.1.4 Atmospheric attenuation

Ultrasound in air is heavily attenuated by atmospheric attenuation, even over short distances of a few meters (Goerlitz, 2018, Lawrence and Simmons (1982)). Atmospheric attenuation will reduce the received SPL of a masker or echo at the focal bat. We chose a range of values for the atmospheric attenuation coefficient between 0 to -2 dB/m. These values approximate the atmospheric attenuation experienced by a bat calling at very low (≤ 20 kHz, ~ 0 dB/m) to high (60 kHz, ~ -2 dB/m) peak frequencies.

2.6.1.5 Geometric attenuation

Sound pressure level reduces with increasing distance from the source, called geometric attenuation. For all sounds in our model (target echoes, masking calls and secondary echoes), we implemented spherical geometric spreading, i.e., uniform spreading of sound in all directions (Speaks, 1996).

2.6.1.6 Group geometry

A group of bats might organize themselves tightly or well spread in the field. The spacing between bats will decide how loud the returning target echoes, masking calls and secondary echoes are. We simulated a group of bats by placing individual bats on a 2D plane using the Poisson disk algorithm (Bridson, 2007). The advantage of using the Poisson disk algorithm is that points are spaced relatively uniformly in space compared to a random placement of points from two independent distributions. The other advantage of the Poisson disk algorithm is that it allows the specification of a minimum distance between two points. For the first simulations varying group size only, we chose 0.5 m as inter-bat distance (see Table 2.1, main text), matching the average interbat-distance in dense swarms of *T. brasiliensis* in the field (Theriault et al., 2010). In addition to 0.5 m minimum inter-bat distance, we also studied how a sparser 1.0 m minimum inter-bat distance affects neighbor detection (see Table 2.1, main text). The Poisson disk arranged bats showed average inter-bat distances of between 1-1.5 times the specified minimum distance between points.

2.6.1.7 Heading variation

Active sensing animals are known to ‘scan’ their environments by emitting energy in varying directions of interest according to the behavioral context (Wisniewska, 2015, Bullock et al. (2005)). Bats alter the shape and direction of their sonar beam while they fly (Jakobsen and Surlykke, 2010, Ghose et al. (2006), Lee et al. (2017)). The directions into which each bat in a group aims its calls could affect how well each bat in the group can detect echoes. A group of bats calling into the same direction may experience high masking, as a focal individual will receive many loud calls from the bats behind it. In contrast, a group of bats calling into a larger range of directions may experience less masking. The focal bat may receive a mix of fainter off-axis calls and loud on-axis calls from neighbors.

We simulated the scanning behavior of individual bats in the group by setting a heading angle for each individual. Each individual called into the direction of its heading angle, and we chose two levels of variation of heading angles in the group. Groups with a low heading variation were all pointing their beams in more or less the same direction. Groups with high heading variation were pointing their beams in a wider range of directions. A low heading variation simulates an emergence situation where each bat is calling approximately into the overall flight direction of the group. A high heading variation simulates a swarming situation where each bat is calling at a unique direction. Given the lack of empirical data to guide our estimates, we chose $\pm 10^\circ$ for the low heading variation, and $\pm 90^\circ$ for the high heading variation. The heading angle for each individual was randomly drawn from a uniform distribution covering the respective range.

2.6.1.8 Monostatic and bistatic target strength of a flying bat

Quantifying the received levels of echoes and secondary echoes requires knowledge of the target strength of a bat when emitter and receiver are at the same and at different locations. Here, we measured monostatic and bistatic target strengths (Richards et al., 2010) of a flying stuffed *Myotis myotis* bat. Monostatic target strengths refer to the situation where the emitter and receiver are at the same location, i.e., they are the same bat (this is the ‘classical’ target strength usually considered in echolocation research). Bistatic target strength refers to a situation where the emitter and receiver are at different locations, i.e., the receiving bat hears the echo of a call that was emitted by another bat, i.e. a secondary echo.

In the simulations, all incoming and outgoing sounds at the bat are between $\text{abs}(0\text{-}180)^\circ$. Sounds with 0° angle are along the heading direction of the focal bat. Sounds arriving/reflecting on the left have negative angles ($0^\circ \geq \theta \geq -180^\circ$), and those on the right have positive angles ($0^\circ > \theta > 180^\circ$).

Methods: We ensonified a stuffed *Myotis myotis* with outstretched wings, which was suspended from the ceiling at ~ 1 m height and placed on a rotating base, which could be rotated in 45° steps. A speaker (electrostatic Polaroid, custom built) and microphone (CM16/CMPA, Avisoft Bioacoustics, Glienicker, Germany) were placed at a 1 m radial distance to the center of the bat (Figure 2.8). The speaker emitted linear frequency modulated sweeps between 96-20 kHz, with durations of $170 \mu\text{s}$, 1 ms and 2 ms at 92 dB rms SPL re 20 Pa at 1 m. The speaker was driven by a custom-built amplifier with input from a soundcard (Player 216H, Avisoft Bioacoustics, 1 MHz sampling rate). The microphone signal was recorded simultaneously with an attenuated version of the speaker signal on a multichannel soundcard (USG 416H, Avisoft Bioacoustics, 500 kHz sampling rate). The microphone had a noise floor of 24 dB rms SPL re 20 Pa. All echoes were recorded at ≥ 22 dB signal-to-noise ratio. The experiment was performed in the middle of a large empty room ($\sim 4 \times 4 \times 2$ m) to temporally separate bat echoes from background echoes.

We ensonified the bat from front (0°) to back (180°) in steps of 45° . We assumed that the bat was symmetrical and thus did not ensonify angles from $180\text{-}360^\circ$. The angular separation between the speaker and the microphone was also altered in steps of 45° between -180° to $+180^\circ$. This resulted in 40 target strength measurements (5 sound directions x 8 angular separations).

The integrated target strength (Cook, 1985) of the recordings were calculated by subtracting the energy of recordings with the bat from those without the bat at the expected time window of echo arrival. The echo level was calculated in rms by taking the square root of the energy.

Results: The monostatic target strength of a flying stuffed *Myotis myotis* bat at various orientations was between -43 and -34 dB at 1 m distance, matching the general range of previously published values (Götze et al., 2016).

The bistatic target strength, which was used to calculate the received level of the secondary echoes, was between -44 and -10 dB across all combinations of emitter-receiver locations. For further details on experimental protocol, raw data and reproduction of generated results, please refer to the archived Jupyter notebooks at this link: <https://doi.org/10.5281/zenodo.3469845>.

2.6.1.9 Acoustic shadowing in bat groups with varying number of bats and inter-bat spacing

As multiple bats fly together in a group, the bats themselves will block all sounds travelling between an emitter and a receiver. Essentially, the bats themselves act as obstacles that cause acoustic shadowing, reducing the received sound pressure level at the focal bat. In a large group, multiple bats may shadow a sound as it moves from the emitter and to the receiver. We quantified acoustic shadowing in a series of playback experiments that varied the inter-bat spacing (0.5 and 1.0 m) and the number of bats (1 – 6) in a line.

Methods: A microphone (CM16/CMPA, Avisoft Bioacoustics) and speaker (Polaroid, custom-built) were placed at a fixed distance of 9.9 m apart, facing each other. We hung 1 to 6 “model bats” made of foam with paper wings at 0.5 or 1.0 m distance to each other from a string running above the speaker to the microphone. The designed model bat showed acoustic shadowing similar to that of the stuffed *Myotis myotis* used in the target strength measurements described in 2.6.1.8.

The speaker was placed as far as possible from the microphone to calculate acoustic shadowing without the effects of speaker directionality. The speaker played back a variety of 7 ms Tukey windowed signals consisting of pure tones (20, 35, 50, 100 kHz) and a downward modulated linear sweep (100-15 kHz). Each signal type was played back 15 times at ~4% duty cycle. Multiple signal types were used to obtain a generalized estimate of shadowing across a wide range of call peak frequencies and call types. The playback signals and recordings are available here: <https://doi.org/10.5281/zenodo.3469845>. Additionally, we also recorded the same playback without model bats being present. We calculated acoustic shadowing as the reduction in received level by subtracting the received level (in dB rms) without bats from the received level with bats. We performed a linear regression of attenuation as a function of factors number of bats and inter-bat-distance, to estimate the amount of acoustic shadowing caused per bat and the spacing between them.

Results: Bats effectively shadowed the sound, with strong effects of the inter-bat-spacing and the number of bats. Bats at 0.5 m distance in front of the receiving bat (=microphone) reduced the received SPL by 5.17 dB (SEM=0.44, t=-11.639, 95% CI =-6.05,-4.30), while bats at 1.0 m interbat-spacing reduced the received SPL by 1.85 dB (SEM=0.44, t=-4.164, 95% CI =-2.72,-0.98). Each bat reduced received SPL by 0.83 dB (SEM=0.08, t=-9.852, 95% CI =-0.99,-0.66). For further details on experimental protocol, raw data and reproduction

of generated results, please refer to the archived Jupyter notebooks at this link: <https://doi.org/10.5281/zenodo.3469845>.

2.6.2 Model implementation

Here, we present how we implemented the parameters described before into our final model, and how our model was initialized and run.

2.6.2.1 Model idea

The idea of our model is to analyze the relative timing and sound pressure level of target echoes, masking calls and secondary echoes at a focal bat flying in a group of other bats. Each model iteration thus analyzed one single interpulse interval, i.e., the time after emission of one call until the emission of the next call by the focal bat. Within that interpulse interval, the focal bat received the echoes from its own call that reflected off the neighboring bats, the calls of those neighboring bats, and the secondary echoes which originate from the calls of the neighboring bats reflecting off other neighboring bats (Figure 2.1, main text).

We placed groups of bats in a 2D plane with various inter-bat distances and heading directions. We then calculated received timing and SPL of all sounds based on realistic assumptions about call properties, mammalian auditory characteristics and sound physics.

All echoes, calls and secondary echoes were considered to be equal in duration, amplitude envelope, and frequency composition. Frequency composition was not explicitly specified, which is a conservative modelling choice that maximizes masking potential (Yost, 2007) and makes our model generalizable to multiple bat species. All sounds were treated as having a constant amplitude envelope (i.e., no amplitude modulations), but they differed in the sound pressure level received by the focal bat.

2.6.2.2 Model initialization

Each model iteration consisted of distributing bats in a 2D plane and assigning each bat a heading direction. This spatial distribution was used to calculate the arrival times and received level of target echoes, masker calls and secondary echoes within the interpulse interval (Figure 2.1). The interpulse interval was discretized into time bins of $1 \mu\text{s}$ duration. Each received target echo corresponded to one neighbor. The arrival time of each echo was calculated using twice the distance between focal and neighboring bat.

The arrival time of masker calls and secondary echoes were chosen randomly. The random arrival time assignment of masker calls and secondary echoes is supported by the finding that groups of *Miniopterus fuliginosus* (Hase, 2018)

do not coordinate their calling behavior, and seem to echolocate independently. Moreover, at large group sizes beyond a few bats it is unlikely that bats could effectively co-ordinate their call emission times.

2.6.2.3 Target echo properties

Target echoes are the echoes that the focal bat receives in response to its own echolocation call. In our model, the target echoes are echoes reflected off the neighboring bats. When a focal bat hears a target echo it means it has detected the corresponding neighboring bat.

An echo was defined as a sound occupying a block of time within the interpulse interval (Figure 2.9A). Echoes were simulated to arrive at delays corresponding to the distance to the neighboring bat they reflected off, e.g. if a neighboring bat was at 1 m distance to the focal bat, then its echo arrived at a delay of 6.06 ms (at 330 m/s sound propagation).

The received level of the returning echo was calculated based on emitted call source level into the direction of the neighboring bat, our monostatic target strength measurements of a bat, and geometric attenuation over the sound travel distance. If acoustic shadowing and atmospheric absorption were included in a simulation run, the received level was reduced based on the number of bats in the path and the atmospheric attenuation for the overall distance travelled by the echo. Echo arrival direction was determined based on the position of each neighboring bat.

2.6.2.4 Masker call properties

Masker calls arrived at random time points with uniform probability within the interpulse interval (Figure 2.9B), based on the observed lack of call synchronization in groups of *Miniopterus fuliginosus* (Hase, 2018). Call directionality was based on the directionality function in Giuggioli et al., 2015, who fit a cosine based function to describe the overall call directionality of *Myotis daubentonii* echolocating in the field. We set the asymmetry parameter A to 7.0. We calculated the angle of call emission towards the focal bat for each conspecific bat based on its angular position (heading) and distance. We then calculated the effective source level into the direction of the focal bat by reducing the call's on-axis source level (Table 2.1 in main text) according to the call directionality function and the focal bat's relative position to the conspecific (Figure 2.7). This reduced level was the final received level of the conspecific masker call. If acoustic shadowing and atmospheric attenuation were included in a simulation run, the received level was reduced based on the number of bats in the path and the overall distance travelled by the call.

2.6.2.5 Secondary echo properties

Like the masking calls, secondary echoes arrived randomly with uniform probability in the interpulse interval (Figure 2.9C). The received level of a secondary echo was based on the emitted call source level into the direction of the neighboring bat, our bistatic target strength measurements of a bat, and geometric attenuation over the sound travel distance. If acoustic shadowing and atmospheric absorption were included in a simulation run, the received level was reduced based on the number of bats in the path and the overall distance travelled by the secondary echo.

2.6.2.6 Obtaining the masker sound pressure level profile

All sounds were treated as having a fixed received level (no envelope modulations). For each target echo, we calculated its unique masker SPL profile based on the relative timing, relative arrival directions and received levels of all asking calls and secondary echoes. This masker SPL profile was different for each target echo because the temporal and spatial properties of the masking sounds differ for each echo, resulting in different received levels and spatial unmasking (see main text for details). We first calculated the effective masker SPL for each masking sound by correcting for spatial unmasking based on the angular separation between the echo and the masker. All effective masker SPLs of all masking sounds together over time represent the complete masker sound pressure profile for each target echo (Figure 2.9E). When two or more maskers overlapped in time, we added their linear sound pressures to obtain their joint masking SPL. This approach assumes that overlapping maskers are coherent sound sources that constructively interfere. This is a conservative assumption that will maximize masking.

2.6.2.7 Determining neighbor detection: the temporal masking envelope

The masker SPL profile for each echo describes the received masking SPL over time. From the masker SPL profile, we created an echo-to-masker ratio profile by normalizing the SPL of the target echo to the masker SPL profile:

$$\text{echo-to-masker ratio profile (dB)} = \text{echo level (dB SPL)} - \text{masker sound pressure level (dB SPL)}$$

The echo-to-masker-ratio profile is comparable to a signal-to-noise-ratio: at 0 dB, echo and masker have the same SPL. The masker is louder than the echo for negative values, and the echo is louder than the masker for positive values.

To determine whether a given target echo was heard or not, we compared the echo-to-masker ratio for this echo with the temporal masking envelope (see 2.6.1.1). The temporal masking envelope describes the echo-to-masker ratio at

which masking occurs as a function of relative timing between echo and masker. Using the temporal masking envelope is important because masking does not only occur when the masker coincides with the echo, but also when the masker does not overlap with the echo and arrives before (forward masking) or after (backward masking) it. In our case, echo and masker had durations of only 1-2.5 ms, while a masker arriving at up to ~25 ms before and up to ~1 ms after the echo still causes some amount of masking. Thus, our temporal masker envelope had a duration of either ~27 or ~28.5 ms (Figure 2.9F). We compared the echo-to-masker ratio profile to the temporal masking envelope. The echo was considered not heard if the echo-to-masker ratio profile lay below the temporal masking envelope, i.e., the echo-to-masker ratio was lower than required for echo detection. Alternatively, the echo was considered heard if the echo-to-masker ratio profile lay above the temporal masking function, i.e., the echo-to-masker SPL ratio was higher than required for echo detection.

However, as the echo-to-masker ratio continuously fluctuates over time, it is possible that it is not fully above or below the temporal masking envelope throughout the envelope's duration. We thus defined an echo to be masked (= not heard), if it was masked for more than 25% of its duration (of 1 or 2.5 ms). To calculate the total duration of masking, we analyzed the total duration that the echo-to-masker ratio was below the temporal masking envelope. As long as the total masking duration was shorter than 25% of the echo duration (of 1-2.5 ms) the echo was considered detected. If this duration was longer than 25% of the echo duration, the echo was masked and the corresponding neighboring bat was considered not detected. This 25% threshold was set to make the simulated auditory system immune to short spikes in masking sound pressure level occurring during the temporal masking function.

2.6.3 Open-source software used in the research

All simulation code, experimental data and results were made possible through the use of the NumPy (Oliphant, 2006a), SciPy (Virtanen, 2019), Pandas (McKinney, 2010), Matplotlib (Hunter, 2007a), Statsmodels (Seabold and Perktold, 2010), sounddevice (Geier, 2015a), Anaconda (Anaconda, 2016) and CPython (Van Rossum, 1991) open-source projects.

2.6.4 Acknowledgements

We thank Renate Heckel and Felix Hartl for contributing to and building the ensonification setup, Magnus Wahlberg for helpful discussions on the ensonifications, Henrik Brumm for permission to use Raum 1.03 for the ensonification experiments, and Mihai Valcu for facilitating simulation runs on the in-house server facility.

2.6.5 Supplementary Schematics, Tables and Figures

2.6.5.1 Echo detection pseudo-code

Pseudo-code of the steps in a simulation run to determine the detected neighbors per call emission.

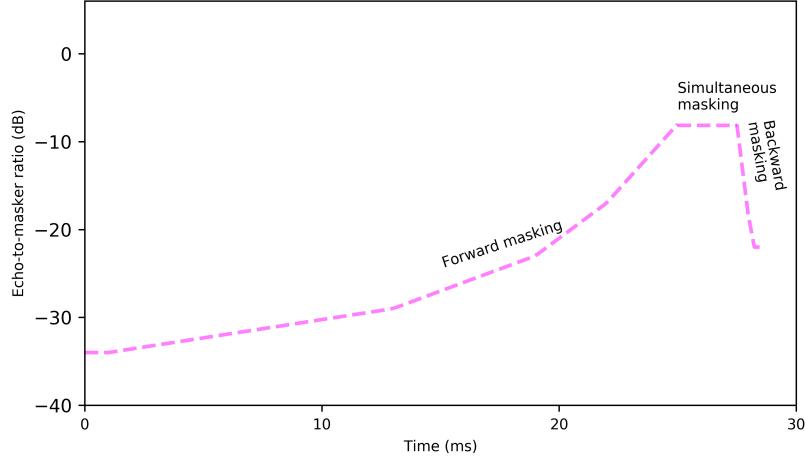
1. Place N bats in group with minimum inter-bat distance
2. Choose bat closest to the center of the group as the focal bat
3. Populate interpulse intervals with maskers and echoes:
 - () Propagate maskers (calls and secondary echoes) and calculate their received levels according to the position and orientation of the source neighbors. Assign random timing within interpulse interval.
 - () Propagate echoes from focal bats' own call and calculate their arrival time and received levels according to the position and orientation of the neighbors.
4. Implement hearing directionality of the focal bat: amplify the received level of all sounds according to their relative angle of arrival
5. Per echo, determine if it was heard:
 - () Implement spatial unmasking by reducing the effective received level of all masking sounds based on their angular separation to the echo
 - () Combine all maskers over time to form a 'masker profile'
 - () Calculate the 'echo-to-masker profile', with reference to the echo level
 - () Implement temporal masking by checking if the relative echo-masker profile lies below the temporal masking envelope centered on the echo's location in the interpulse interval.

2.2: Target-detection studies in echolocating bats used to extract echo-masker SPL ratio for our model. The time delay is the time between the edge of a masker and the target echo. A positive time delay indicates forward masking (masker arrives before the target), a negative time delay indicates backward masking (masker arrives after the target).

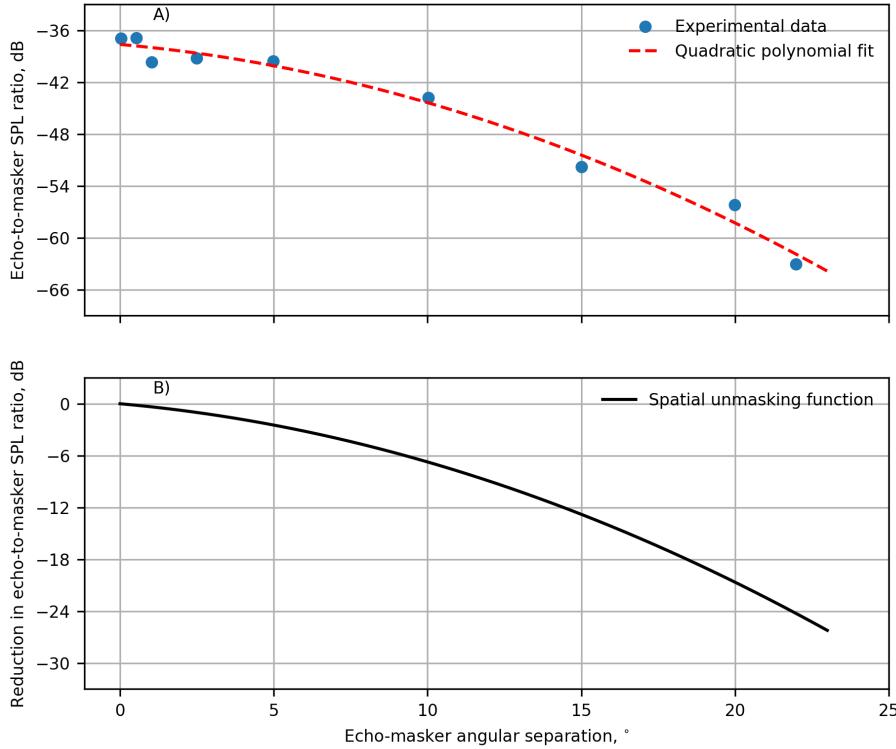
Publication	Species	Time delay (ms)	Masking condition	Echo-masker SPL (dB)
Siewert et al. 2004	<i>Megaderma lyra</i>	3	Forward	-17
Siewert et al. 2004	<i>Megaderma lyra</i>	6	Forward	-23
Siewert et al. 2004	<i>Megaderma lyra</i>	12	Forward	-29
Siewert et al. 2004	<i>Megaderma lyra</i>	24	Forward	-34
Sümer et al. 2009	<i>Eptesicus fuscus</i>	-0.65	Backward	-22.3

2.3: Results of the logistic regression to quantify the effect of different parameter values on the odds ratio to detect at least one neighbor. Odds ratio values >1 indicate a higher probability of neighbor detection, while odds ratios <1 indicate a lower probability of neighbor detection.

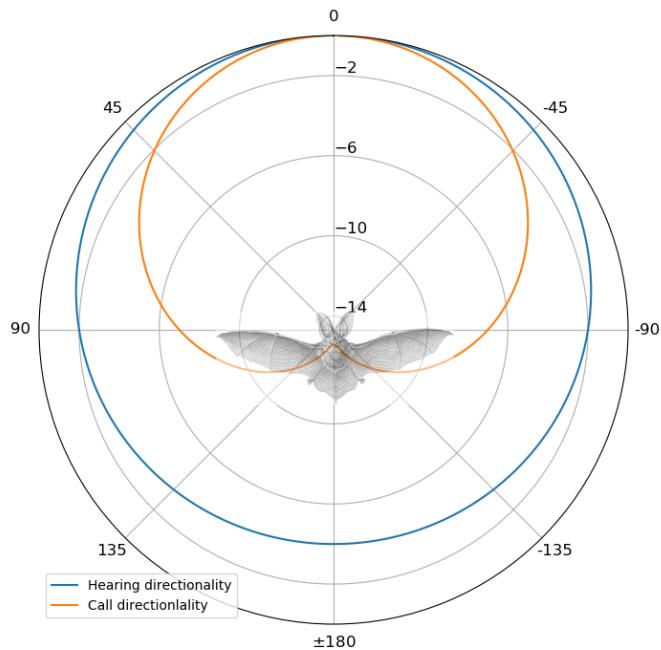
Parameter	Value tested	Reference value	Odds Ratio	Odds Ratio - 2.5 CI	Odds Ratio - 97.5 CI	Log Odds Ratio	Log Odds Ratio SEM	Z (log odds ratio estimate)	P > z
Intercept			0.32	0.31	0.35	-1.11	0.027	-40.51	0
Heading variation (°)	±90	±10	1.32	1.28	1.36	0.28	0.015	18.35	0.0
Acoustic shadowing	Yes	No	0.75	0.73	0.78	-0.28	0.015	-18.63	0.0
Interpulse interval (ms)	25	100	0.001	9e-04	0.001	-6.84	0.073	-93.52	0
	50	100	0.048	0.046	0.05	-3.04	0.023	-	0
								134.14	
	200	100	14.6	13.995	15.228	2.68	0.022	124.46	0
	300	100	74.68	70.497	79.122	4.31	0.029	146.49	0
Minimum interbat distance (m)	1	0.5	0.31	0.301	0.321	-1.17	0.016	-72.83	0
Sound duration (ms)	1	2.5	34.66	33.172	36.206	3.55	0.022	158.83	0
Source level (dB SPL re 20 μ Pa at 1m)	94	100	0.99	0.941	1.034	-0.01	0.024	-0.57	0.57
	106	100	1.01	0.966	1.061	0.01	0.024	0.52	0.6
	112	100	1.01	0.966	1.061	0.01	0.024	0.5	0.62
	120	100	0.98	0.938	1.03	-0.02	0.024	0.73	0.47
Atmospheric attenuation (dB/m)	0	-1	1.05	1.01	1.086	0.05	0.019	2.48	0.01
	-2	-1	1.01	0.97	1.042	0.01	0.018	0.27	0.78



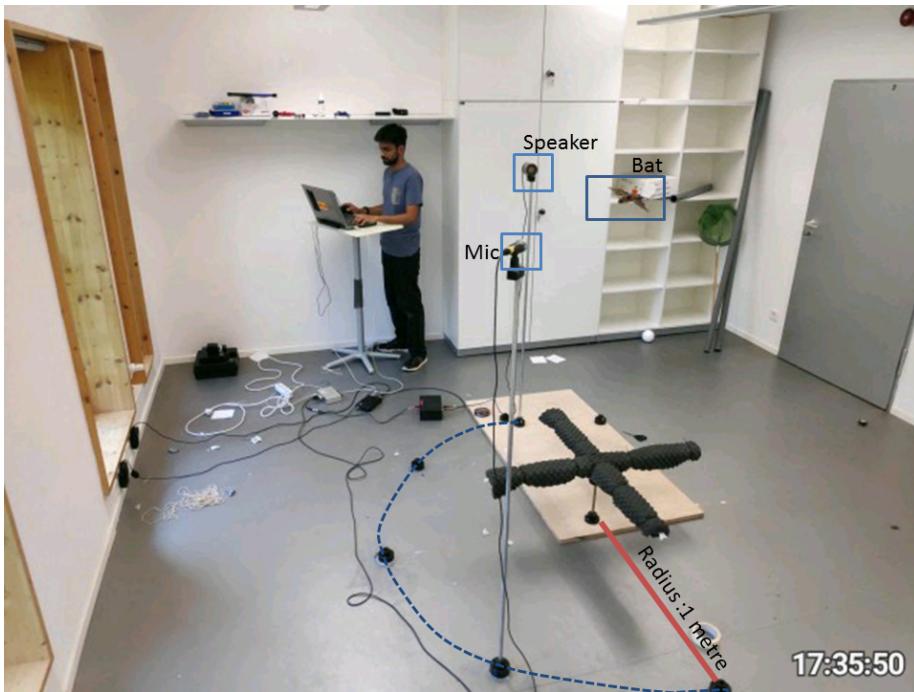
2.5: The 'temporal masking envelope' used to simulate temporal masking. The envelope represents the lower echo-to-masker ratios at which a bat can detect echoes for various echo-masker delays. The envelope is the equivalent of the lowest signal-to-noise ratios at which echoes can be detected over different time delays. The envelope is centered on the position of the echo, and has a long forward masking section (at times prior to the echo), and a short backward masking section (at times after the echo). The simultaneous masking region is equal to the length of the echo itself. If the echo-to-masker ratio profile is above the temporal masking envelope for most of its duration (i.e., the echo-to-masker SPL ratio was higher than required for echo detection), we considered an echo to be heard. If the echo-to-masker ratio profile is below the envelope for more than 25% of the echo's duration, the echo was considered not heard. Here the temporal masking envelope is shown for a 2.5 ms echo. Data and sources used to construct the temporal masking envelope are given in Table 2.2



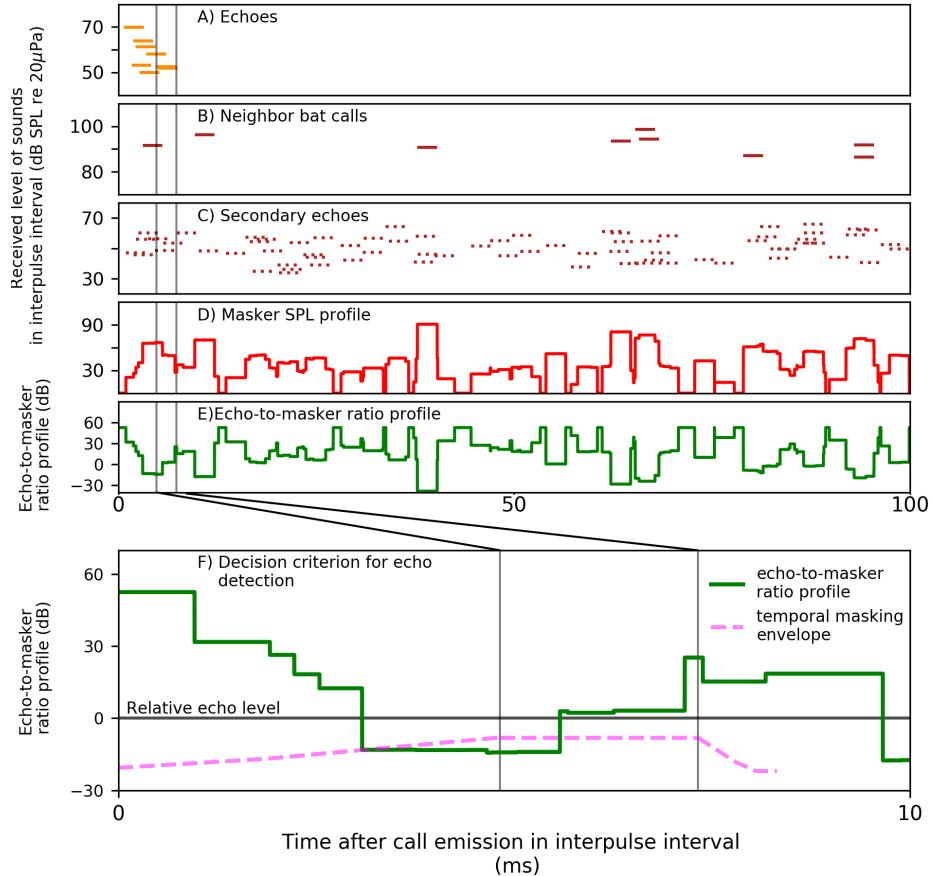
2.6: The ‘spatial unmasking function’ describes the reduction in echo-to-masker SPL ratio at echo detection as a function of angular separation between echo and masker. A) The original data set of Sümer et al. (2009) (blue dot) and our digitized and interpolated dataset (red line). The error between the data and our interpolation is less than 2 dB. B) The final spatial unmasking function as used in our simulations was derived from the interpolated fit in A), which was normalised to the echo-to-masker ratio at zero degrees angular separation. This final spatial unmasking function describes the reduction in required echo-to-masker SPL ratio relative to the co-localized case: when echo and masker are co-localized, the reduction is 0 dB, while the reduction becomes greater with increasing angular separation.



2.7: Calling and hearing directionality of bats. Call directionality (orange) is directional, with a difference of up to -14 dB in source level from front to back. Calls emitted to the front of a bat result in higher received levels of calls, echoes and secondary echoes. Hearing directionality (blue) is less directional, with a difference of up to -4 dB from front to back. Hearing directionality causes sounds arriving from the back to be perceived fainter than sounds arriving from the front. Bat drawing from *Kunstformen der Natur* (Ernst Haeckel, 1899).



2.8: The ensonification setup used to measure the monostatic and bistatic target strength of a bat as an acoustic target. A stuffed *Myotis myotis* bat was hung at the same height as the speaker and microphone. The bat could be rotated in the azimuth. The microphone and speaker were placed at 1 m radius around the bat at various positions (black rounded plastic molds on floor) with a separation of 45° from each other. By a combination of bat orientation, microphone and speaker positions all possible incoming and outgoing relative angles were measured. Here the positions of the speaker and microphone for a bistatic target strength measurement with 135° angle between microphone and speaker are shown.



2.9: Schematic representation of the sounds arriving in the interpulse interval for a simulation with a group of 10 bats (A-C) and how echo detection was determined (D-F). A-E show the full interpulse interval of 100 ms, while F shows an enlargement of the first 10 ms. A) Timing and received SPL of the individual echoes reflected off neighbors. The echoes arrive at delays corresponding to the neighbors' distance from the focal bat. The vertical lines single out one specific target echo to illustrate the simulated auditory system (see F). B) Timing and received SPL of the calls from neighboring bats arriving randomly with a uniform probability over the interpulse interval. C) Timing and received SPL of the secondary echoes in the interpulse interval, arriving randomly with uniform probability over the interpulse interval. D) The masker SPL profile obtained by adding the effective masker SPL of all maskers (calls and secondary echoes) over time for the chosen echo. The effective masker SPL is the received SPL corrected for spatial unmasking based on the angular separation between the echo and the masker, using the spatial unmasking function in Fig. S2. E) The echo-to-masker ratio profile obtained by normalizing the echo SPL to the masker SPL profile. 0 dB is the echo's relative SPL. F) Determining whether an echo was heard or not, by comparing the echo-to-masker ratio profile (solid green) to the temporal masking envelope (dashed pink). If the echo-to-masker ratio is above the temporal masking envelope, then the echo was not masked. In contrast, if the echo-to-masker ratio is cumulatively below the temporal masking envelope for more than 25% of the echo's duration (of 1 or 2.5 ms), then the echo was considered masked. The vertical lines indicate the actual temporal location of the example echo from A). The temporal masking envelope is centered on the chosen echo. Here, the echo-to-masker ratio is below the temporal masking envelope for almost a whole echo duration, meaning that this echo was masked.

3

Horseshoe bat craziness paper

Here we describe a whole series of crazy experiments and results with horseshoe bats.

4

ushichka

Some *significant* applications are demonstrated in this chapter.

4.1 Example one

4.2 Example two

5

Robust Self-Calibration of Constant Offset Time-Difference-of-Arrival

This chapter was published as a peer-reviewed paper in the conference proceedings of the International Conference on Acoustics, Speech, and Signal Processing:

Batstone, K., Flood, G., Beleyur, T., Larsson, V., Goerlitz, H. R., Oskarsson, M., & Åström, K. (2019, May). Robust self-calibration of constant offset time-difference-of-arrival. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (pp. 4410-4414). IEEE.

Abstract

In this paper we study the problem of estimating receiver and sender positions from time-difference-of-arrival measurements, assuming an unknown constant time-difference-of-arrival offset. This problem is relevant for example for repetitive sound events. In this paper it is shown that there are three minimal cases to the problem. One of these (the five receiver, five sender problem) is of particular importance. A fast solver (with run-time under $4 \mu s$) is given. We show how this solver can be used in robust estimation algorithms, based on RANSAC, for obtaining an initial estimate followed by local optimization using a robust error norm. The system is verified on both real and synthetic data.

5.1 Introduction

The problem of estimating receiver-sender node positions from measured arrival times of radio or sound signals is a key issue in different applications such as microphone array calibration, radio antenna array calibration, mapping and positioning. This field is well researched but in this paper we will focus on the anchor-free sensor network calibration both in terms of time-of-arrival measurements (TOA) and time-difference-of-arrival measurements (TDOA). For time-of-arrival the planar case of three receivers and three senders (3R/3S) was solved in Stewénius (2005). For the full 3D case the over-determined problem (10R/4S) was studied in Pollefeys and Nister (2008), where a solver for this non-minimal case was provided. There are actually three minimal cases for the 3D case, namely (4R/6S), (5R/5S) and (6R/5S). A practical solver was presented in Kuang et al. (2013). There are in general 38, 42 and 38 solutions respectively for the three different set ups. Faster solvers for these minimal cases were provided in Larsson et al. (2017).

In this paper we study the constant offset TDOA self-calibration problem. It is a problem that naturally arises e.g. when signals are emitted with a known period. As an estimation problem it lies between TOA and full TDOA. In the paper we study the minimal (5R/5S) problem and provide a fast (few μs) solver. Robust parameter estimation often use the hypothesize and test paradigm, e.g. using random sampling consensus, Fischler and Bolles (1981) or one of its many variants (Chum et al. (2003); Raguram et al. (2012); Korman and Litman (2018)). In these frameworks minimal solvers are important building blocks for generating model hypotheses, and we show in the paper how a minimal solver can be used for robust parameter estimation of sender positions, receiver positions and unknown offset. The system is capable of handling missing data, outliers and noise. The algorithms are tested on synthetic data as well as real data, in an office environment and in a cave. The methods are straightforward to generalize for degenerate configurations which arise if senders or receivers are restricted to a plane or to a line.

5.2 Time-difference-of-arrival self calibration

The problem we are considering involves m receiver positions $\mathbf{r}_i \in \mathbb{R}^3$, $i = 1, \dots, m$ and n sender positions $\mathbf{s}_j \in \mathbb{R}^3$, $j = 1, \dots, n$. This could for example represent the microphone positions and locations of sound emissions, respectively. Assume that the arrival time of a sound j to receiver i is t_{ij} and that the time that sound j is emitted is T_j . Multiplying the travel time $t_{ij} - T_j$ with the speed v of the signal we obtain the distance between senders and receiver,

$$v(t_{ij} - T_j) = \|\mathbf{r}_i - \mathbf{s}_j\|_2, \quad (5.1)$$

where $\|\cdot\|_2$ is the l^2 -norm. The speed v is throughout the paper assumed to be known and constant.

In many settings the times of emissions T_j are unknown, but regular, *e.g.*

$$T_j = k_1 j + k_0, \quad (5.2)$$

where the interval k_1 is known. Inserting (5.2) into (5.1) we obtain

$$v(t_{ij} - k_1 j - k_0) = \|\mathbf{r}_i - \mathbf{s}_j\|_2. \quad (5.3)$$

Assuming an erroneous (but regular) emission time $\tilde{T}_j = k_1 j + \tilde{k}_0$ and introducing (the measured) $z_{ij} = v(t_{ij} - \tilde{T}_j)$ and (the unknown) $o = v(k_0 - \tilde{k}_0)$ yields the following expression

$$z_{ij} = \|\mathbf{r}_i - \mathbf{s}_j\|_2 + o. \quad (5.4)$$

Note that this is a simplified variant of the general time-difference-of-arrival problem (see *e.g.* Kuang and Astrom (2013)), which allows for a different offset o for every j ,

$$z_{ij} = \|\mathbf{r}_i - \mathbf{s}_j\|_2 + o_j. \quad (5.5)$$

Problem 1 (Constant Offset Time-Difference-of-Arrival Self-Calibration)
Given measurements \tilde{z}_{ij}

$$\tilde{z}_{ij} = \|\mathbf{r}_i - \mathbf{s}_j\|_2 + o + \epsilon_{ij}, \quad (5.6)$$

for a subset $W \subset I$ of all the receiver-sender index pairs $I = \{(i, j) | i = 1, \dots, m, j = 1, \dots, n\}$ determine receiver positions \mathbf{r}_i , $i = 1, \dots, m$ and sender positions \mathbf{s}_j , $j = 1, \dots, n$ and offset o . Here the errors ϵ_{ij} are assumed to be either **inliers**, in which case the errors are small ($\epsilon_{ij} \in N(0, \sigma)$) or **outliers**, in which case the measurements are way off.

Here we will use the set W_{in} for the indices (i, j) corresponding to the inlier measurements and W_{out} for the indices corresponding to the outlier set.

5.3 Local optimization and the low rank relaxation

If an initial estimate of the parameters $\theta_1 = \{R, S, o\}$ is given and if the set of inliers is known, then refinement of the estimate can be found by optimization methods, *e.g.* Levenberg-Marquardt (LM) (Levenberg (1944); Marquardt (1963)),

$$\min_{\theta_1} f(\theta_1) = \sum_{(i,j) \in W_{\text{in}}} (z_{ij} - (\|\mathbf{r}_i - \mathbf{s}_j\|_2 + o))^2. \quad (5.7)$$

There is an interesting relaxation to the problem, that exploits the fact that the matrix with elements $(z_{ij} - o)^2$ is rank 5, (Pollefeys and Nister (2008)). Further

simplifications use the double compaction method (Kuang and Astrom (2013)). The double compaction matrix M is defined as the matrix with elements

$$M_{ij} = (z_{ij} - o)^2 - a_i - b_j, \quad (5.8)$$

and it can be shown to have rank 3, i.e. $M = U^T V$, where U is of size $3 \times m$ and V is of size $3 \times n$. The relaxed problem involves a set of parameters $\theta_2 = \{U, V, b, a, o\}$. Here the constraints can be written as

$$z_{ij} = \sqrt{u_i^T v_j + a_i + b_j} + o, \quad (5.9)$$

where u_i denotes column i of U and v_j denotes column j of V . Refinement of parameters can be done by performing local optimization on

$$\min_{\theta_2} f(\theta_2) = \sum_{(i,j) \in W_{\text{in}}} \left(z_{ij} - (\sqrt{u_i^T v_j + a_i + b_j} + o) \right)^2. \quad (5.10)$$

5.4 Minimal problems and solvers

By counting equations and unknowns, one finds that there are three minimal problems. The first two are the symmetric case when $m = 4, n = 7$ or $m = 7, n = 4$. This case is not addressed in this paper, but we believe it to be difficult to solve. The other case is $m = n = 5$. Here, we first present a solver for the constant offset and then discuss how to solve for sender and receiver positions.

Given a 5×5 matrix, Z , with time-difference-of-arrival measurements z_{ij} , the rank 3 constraint on the double compaction matrix in (5.8) can be written as

$$f(o) = \det(C^T(Z - o)^{\circ 2}C) = 0, \quad (5.11)$$

where

$$C = \begin{pmatrix} -1 & -1 & -1 & -1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (5.12)$$

and \circ^2 denotes element-wise squaring (Hadamard power). Although the elements of $(Z - o)^{\circ 2}$ are of degree 2 in o , the quadratic terms cancel out after multiplication with C^T and C . Thus the elements of $C^T(Z - o)^{\circ 2}C$ are linear in o . Since the determinant is linear in each column, the determinant $f(o)$ is a polynomial of degree four in the offset o . This can be summarized as

Theorem 1 *Given time-difference-of-arrival measurements from five receivers to five senders, there are four possible offsets o , given as the roots to the fourth degree polynomial $f(o)$, counting complex roots and multiplicity of roots.*

5.1: Execution times for 5×5 minimal solvers steps. Notice that the steps of calculating o and the relaxed solution is significantly faster than upgrading to the full solution

Implementation	Matlab	C++
Calculation of o	$38 \mu s$	$3.7 \mu s$
Calculation of $\theta_2 = \{U, V, a, b, o\}$	$100 \mu s$	N/A
Calculation of $\theta_1 = \{R, S, o\}$	$600 ms$	$22 ms$

For each solution o it is possible to generate a solution θ_2 to the relaxed problem, according to

$$b = ((z_{11} - o)^2 \ (z_{12} - o)^2 \ (z_{13} - o)^2 \ (z_{14} - o)^2 \ (z_{15} - o)^2),$$

$$a = \begin{pmatrix} 0 \\ (z_{21} - o)^2 - (z_{11} - o)^2 \\ (z_{31} - o)^2 - (z_{11} - o)^2 \\ (z_{41} - o)^2 - (z_{11} - o)^2 \\ (z_{51} - o)^2 - (z_{11} - o)^2 \end{pmatrix}, \quad (5.13)$$

$$U = (0 \ u_2 \ u_3 \ u_4 \ u_5), \quad (5.14)$$

$$V = (0 \ v_2 \ v_3 \ v_4 \ v_5), \quad (5.15)$$

where $(u_2 \ u_3 \ u_4 \ u_5)^T (v_2 \ v_3 \ v_4 \ v_5)$ is any rank 3 factorization of the matrix $C^T(Z - o)^{\circ 2} C$.

From a solution θ_2 to the relaxed problem it is possible to upgrade to a solution θ_1 to the original problem. This involves solving a system of polynomial equations. The procedure was first described in Kuang et al. (2013), where an algorithm for solving this was presented. Recently, a faster algorithm was presented in Larsson et al. (2017).

An efficient implementation for calculating the four solutions of the offset o given the measurements z takes $4 \mu s$ for a C++-implementation. Generating the solution θ_2 to the relaxed problem adds a few μs . However, calculating a solution θ_1 to the original problem takes another $22 ms$. Thus, it is advantageous to estimate the parameters of the relaxed problem and postpone the upgrade from θ_2 to θ_1 as a final step, see Table 5.1.

5.5 Using RANSAC for five rows

We propose the use of the fast minimal solver in an hypothesize and test framework to obtain (i) a initial estimate on the offset o and (ii) an initial inlier set. The steps are described in Algorithm~1

Algorithm 1 Offset RANSAC

- 1: Randomly select 5 rows and columns. Find the four solutions on o given the time-difference-of-arrival measurements.
 - 2: For each solution o , calculate the relaxed solution $\theta_2 = \{U, V, a, b, o\}$.
 - 3: For selected rows and for each remaining column, check for inliers according to the residuals in (5.10).
-

5.6 Robust estimation of parameters

We use these minimal solvers with RANSAC as described in the previous section to find one or several initial estimates of the parameters θ_2 for a subset of five receivers and k senders. The solution is extended to additional rows and/or columns using robust techniques as described in Batstone et al. (2016). During this process it is useful to keep the errors down by occasionally refining the solutions using local optimization. This has shown to reduce failures, see e.g. (Engels et al. (2006); Klein and Murray (2007)). In the proposed estimation algorithm we postpone the upgrade from θ_2 to θ_1 until we have found a good solution involving a large portion of the receiver and sender positions.

5.7 Experimental Validation

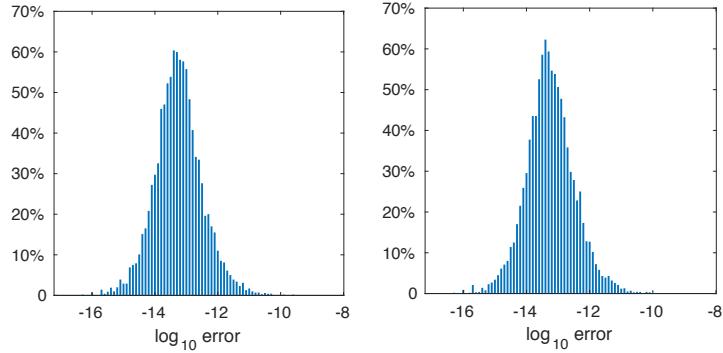
5.7.1 Minimal Solver

To test the numerical accuracy and robustness of our minimal solver we conducted an experiment using simulated data without noise. We generated a large number of instance problems (10,000) with known offsets. We then ran our solvers and compared the returned solutions with the ground truth solution. For each instance problem we recorded the distance to the closest solution. In Figure 5.1 the resulting histogram of the logarithm of the absolute errors are shown. As can be seen, both implementations get close to machine precision.

5.7.2 Experimental Setup for Real Data

We have tested our system on (i) experiments made in an office environment and (ii) experiments made at the Orlova Chuka cave, Bulgaria.

For the office experiments, 12 microphones (8x t.bone MM-1, 4x Shure SV100) were positioned around a room ($\sim 3 \times 5 \text{ m}^2$) and measured using a laser to obtain ground truth positions of the microphones with an error of $\pm 2 \text{ mm}$. The space was cleared of most the furniture to create an open space to conduct the experiment in. The sound recordings were captured using a Roland UA-1610 Sound Capture audio interface and automatically amplified. The recordings were made using the open source software Audacity 2.3.0 with a sampling frequency



5.1: Left shows the histogram of the logarithm of the absolute errors, for the Matlab implementation of our minimal solver. To the right the corresponding histogram for the C++ implementation.

of 96 $k\text{Hz}$ on a laptop. A synthetically generated chirp was then played using a simple loudspeaker every half second for 30 s while moving the speaker around in the room.

For the cave experiments, 12 microphones (4x Sanken CO-100K, 8x Knowles SPU0410) were positioned in a section of the cave, four microphones were placed on an inverted T array near one wall, while the other eight microphones were placed on the adjacent wall. The sound recordings were captured using pre-amplifiers (Quadmic, RME) and two synchronised Fireface 800 (RME) audio interfaces running at a sampling frequency of 192 $k\text{Hz}$. Recording and playback were controlled via a custom written script based on the sound device library (Geier (2015b)) in Python 2.7.12 (Van Rossum and Drake Jr (1995)). Ultrasonic chirps (8 ms, 16 – 96 $k\text{Hz}$ upward hyperbolic sweep) were played every second via one of the audio interfaces, amplified (Basetech AP-2100) and presented through a Peerless XT25SC90-04 loudspeaker. The speaker was attached to a 3-m-long pole and slowly waved in the approximately $5 \times 9 \times 3 \text{ m}^3$ recording volume. Playbacks were done past 6:00 am to prevent disturbing the resident bat population.

5.7.3 Experimental Evaluation for Real Data

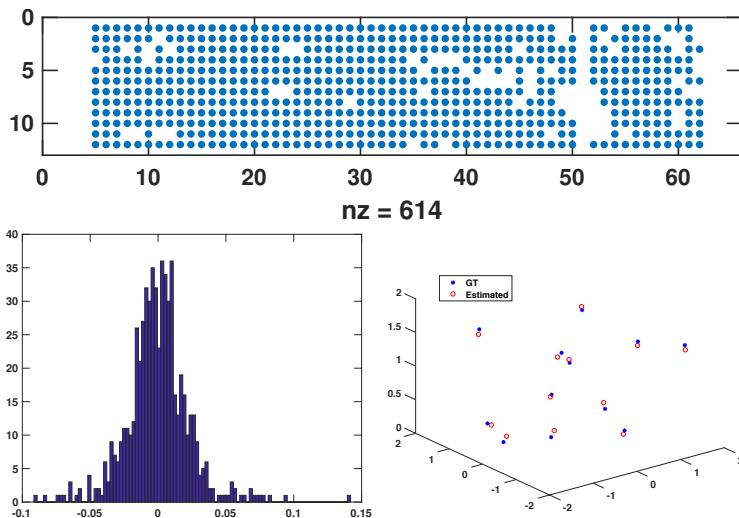
Once the office recordings were taken, an algorithm was used to find the chirps in the captured sound recordings and the algorithm then outputs the z_{ij} matrix. This can then be used in our RANSAC scheme, Algorithm~1. For this experiment we used the (5R/5S) minimal solver. A fixed number of iterations was used; 100 iterations for the initial selection of 5 receivers and senders, then the extension to more columns and rows was allowed until there was no better solution. The tolerance was set to $T = 0.01$ for the initial selection and extension of rows and column.

Once the initial values have been estimated, it underwent l^2 optimization on the inlier set. The results of the estimated microphone positions after the optimization are shown in Figure 5.2.

This produced an Euclidean distance error between each of the microphones calculated position and its ground truth position as (0.2016,0.0587,0.1444,0.1153,0.2017,0.1326,0.1407, 0.1198,0.2041,0.2010,0.1908,0.2110) m.

For graphical purposes, a Procrustes fitting was used on the microphone positions to spread the total error over all 12 microphones. In the Procrustes fitting only rotation and translation were allowed.

For the cave experiment a similar scheme was devised and the results are shown in Figure 5.3.

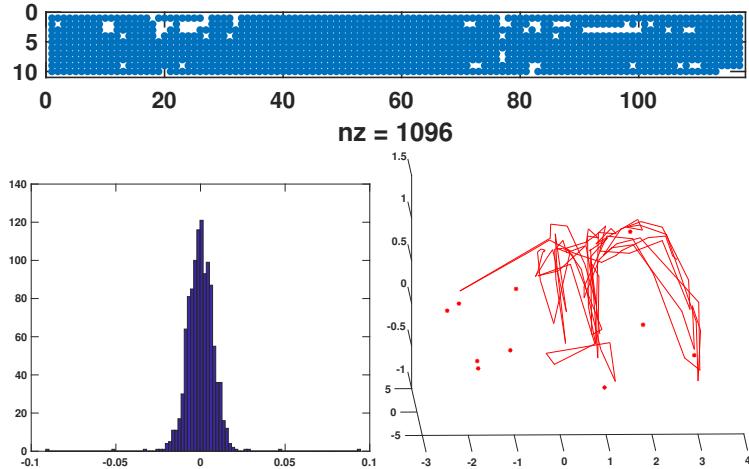


5.2: For the office experiment the figure shows detected inliers W_{in} (top), inlier residual histogram (bottom left), and estimated and ground truth microphone positions (bottom right).

5.8 Conclusions

In this paper, a novel method has been constructed to efficiently solve a TDOA problem with a constant offset. This has been verified using simulated data to test the solver and real experimental data to test our algorithms in realistic scenarios.

Looking at Figure 5.1 and Table 5.1, it can be seen that the calculation of the offsets and the calculation of the relaxed form θ_2 are very fast solvers without loss in numerical accuracy. The advantage of this is that when using a RANSAC approach, the iterations are performed quickly, giving a good initial estimate in



5.3: For the cave experiment the figure shows detected inliers W_{in} (top), inlier residual histogram (bottom left) and estimated microphone and sound source positions, red dots and line respectively (bottom right).

which to optimize over, which is important in highly non-linear systems such as this.

Looking at the results from the office experiment, Figure 5.2, we can see that the calculated microphone positions are accurate and the residuals are small, mostly in the range $\pm 0.04 \text{ m}$. Further to this our inlier set appears to be accurate. The first and last few columns (corresponding to sound emissions) are not used in our initialisation. This is correct because the recording started before the chirps were sounded and ended after, so the chirp detection algorithm falsely determined that they were also chirps but our method decided that the data in those regions do not fit the model. A comparison of the calculated microphone positions were made to a solution from a Full TDOA system, (Kuang and Astrom (2013)), which produced similar results and very similar residuals. This provided a sanity check that the chirp detection was working correctly and that from this dataset a better solution could not be found.

For the cave experiment, similar conclusions can be made, since the residuals are very low, we can conclude that we have an accurate model. This gives a real life example of how algorithms such as the one proposed can be used.

For future work, the study of the number of inliers could be of use. At the moment our algorithm may not extend to more rows and columns if the initial solution is poor, perturbing our final solution. Perhaps a method which could adapt the initial selection in order to give a required amount of inliers could be more advantageous.

6

tacost: Testing and simulating the performance of acoustic tracking systems

This chapter was published as a preprint on *biorXiv*:

Beleyur, T. (2020). tacost: Testing and simulating the performance of acoustic tracking systems. bioRxiv 2020.06.22.165308; doi: <https://doi.org/10.1101/2020.06.22.165308>

Abstract

`tacost` is a Python package to allow the testing of acoustic tracking systems. While many microphone array systems have been characterised analytically and experimentally - these are time-intensive methods. `tacost` provides a simulation based framework to rapidly assess the tracking behaviour of multiple array geometries, and the dissection of other relevant parameters. This paper explains briefly the design of the package and highlights two example use cases in which the tracking accuracy of different microphone geometries are characterised.

6.1 Introduction

Acoustic tracking is a common method used to study vocalising animals such as birds, bats and cetaceans (Suzuki et al., 2017; Aubauer, 1996; Møhl et al., 2000; Hügel et al., 2017; Holderied and Von Helversen, 2003; Rhinehart et al., 2020; Blumstein et al., 2011). Using acoustic tracking, biologists can detect the position of the animal and track it through space as it moves over time. The localisation accuracy of an acoustic tracking system depends on a variety of factors. There are *internal* factors such as microphone array geometry, signal processing routines, and the mathematical formulations used to localise sounds (time-of-arrival, time-of-arrival-difference, angle-of-arrival, power-steering). The *external* factors include aspects related to the actual signal itself, ie. signal-to-noise ratio, and spectro-temporal properties of the emitted sound (noise, linear/hyperbolic sweep) (Wahlberg, 1999). While experiments and analytical modelling may be the definitive way to determine a tracking system's end accuracy, simulations allow a quick and systematic method to estimate the source of tracking errors. **tacost** provides a flexible workflow to manipulate and study the effect of both internal and external factors. **tacost** generates audio files for source positions and array geometries specified by the user. This allows the user to analyse the efficacy of their tracking system's baseline performance.

6.2 Statement of need

Generating simulated audio for a set of source sounds, positions and a given array configuration is a relatively simple task. However, to my knowledge, there are no publicly available, tested and documented packages for this task published to date. Codebases that are publicly available have the advantage of being used by a larger user-base and can thus benefit from bug discoveries much faster than in-house or individually written one-time use scripts. **tacost** provides a robust and well-documented software workflow (Taschuk and Wilson, 2016) with user and developer friendly documentation hosted online. **tacost** contributes to the Python scientific ecosystem in the hope of promoting the growth of acoustics and bioacoustic research in open-source languages like Python. In particular, **tacost** will help researchers working in the field of acoustics and bio-acoustics (eg. de Framond-Bénard et al., 2020) plan and examine the behaviour of their acoustic tracking systems.

6.3 Design

The design of **tacost** focusses on a reproducible and user-friendly method (Wilson et al., 2012) to generate WAV files that form the input for acoustic tracking softwares. Users may interact with **tacost** through custom-written

Python scripts by calling it as a Python package with `import tacost` or in the ‘no-coding’ mode. The ‘no-coding’ mode is especially suitable for users unfamiliar with Python. The no-coding mode is based around a parameter file that is used to specify various parts of the WAV file to be created. Through the parameter file the user can specify the emitted sound, source positions, inter-sound-intervals, sampling rate and other relevant variables to customise the test scenario.

6.4 Examples

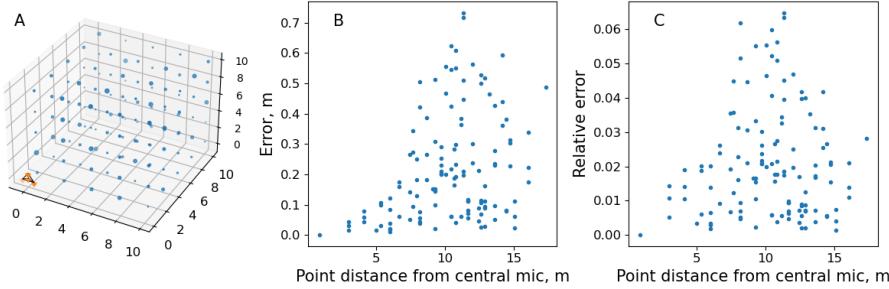
The localisation accuracy of a microphone array may not be uniform over 3D space (Aubauer, 1996; Wahlberg, 1999). This accuracy is independent of the actual signal and recording conditions of the input data, but rather dependent on the array geometry and mathematical formulations used to record and calculate sound source position.

The accuracy of a few microphone array configurations has been characterised analytically (Aubauer, 1996) and experimentally (Wahlberg, 1999). While reflecting the system’s capabilities, analytical and experimental characterisations are often time-intensive. In contrast, simulation uncovers the intrinsic accuracy of an array relatively quickly through the use of audio files with simulated emission points spread across the recording volume of interest. `tacost` can be used to characterise the maximal localisation accuracy of an acoustic tracking system with novel array geometries and recording scenarios. In Example 1, I show how `tacost` can be used to verify known trends in localisation error with the tristar60, a commonly used array system. In Example 2, I show how `tacost` can be used to estimate the expected localisation error in a multi-microphone array with a novel and field-friendly geometry.

6.4.1 Localisation accuracy of the tristar60 system

The tristar60 array is a commonly used array geometry (Aubauer, 1996; Holderied and Von Helversen, 2003; Hügel et al., 2017; Lewanzik and Goerlitz, 2018) with 4 microphones in a plane on an inverted T array. Three peripheral microphones are placed 120° to each other at 60 cm radial distance from the central mic on the inverted T-array.

A series of emission points spanning the upper right quadrant of the array were simulated. The emitted sound was set to a linear sweep. The output WAV files from `tacost` were run through the TOADSuite package (Goerlitz, 2019; Stilz et al., 2019), a software package that localises sounds using the time-of-arrival-differences across channels. Figure 6.1 shows the localisation accuracy map for the tristar60 microphone array. It can be seen (Figure 6.1) that localisation error increases with increasing radial distance from the central microphone, and remains <7% of the radial distance.



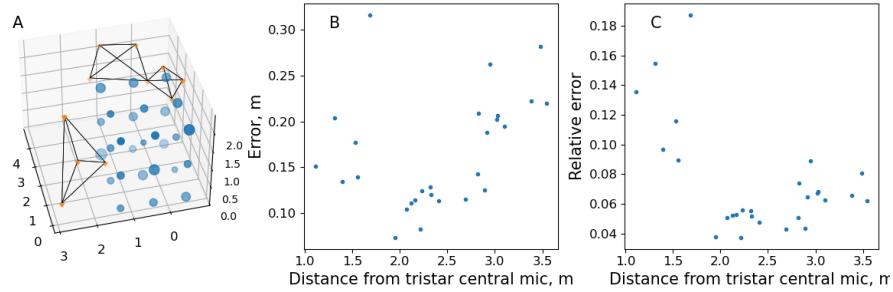
6.1: Accuracy of a sound source localisation based on time-of-arrival-differences with a tristar60 array. A) The tristar60 microphone array is placed at the origin of the coordinate system (bottom left, orange dots connected by black lines). The blue points are the simulated source positions which form a 'calibration grid'. The size of each dot is proportional to the localisation error B) The localisation error increases with increasing radial distance of source from the central microphone. The error is the euclidean distance between the predicted and simulated source point. The errors range between 0-0.7m. C) The relative error of localisation, defined as $\frac{\text{Localisation error}}{\text{Distance from central microphone}}$. Even though absolute localisation error tends to increase with distance from array, all localisation happens with <7 % relative error.

6.4.2 Localisation accuracy of a multi-microphone array in the field

While recording in the field, it may be difficult to use fixed arrays mounted on stands. Arrays on stands are difficult to carry and may also influence the behaviour of the animals being recorded. It is thus advantageous to use less obtrusive microphone geometries, for instance by placing microphones are placed on pre-existing structures such as the walls of a cave or trees. These microphone geometries are field-friendly, but their localisation accuracy is hard to characterise analytically. `tacost` is an ideal tool to explore the tracking performance of such flexibly placed microphone arrays.

(Figure 6.2A) shows the microphone array geometry and recording system described in (Batstone et al., 2019). In short, the array consisted of 11 microphones, 4 of them on a 120cm tristar, and the remaining 7 microphones attached to the walls of a cave. A series of sound emission points were created simulating points in the volume enclosed by the array. The points matched the volume echolocating bats flew within. The simulated sound was set to a linear sweep, which mimicked that of a bat call. The `tacost` output WAV files were analysed with the TOADSuite. The resulting accuracy map reveals that overall, the localisation error is between 7-30 centimetres for the given emission points. This corresponds to a maximum error of upto 30cm in tracking the position, and of upto 19% relative error. In contrast to the previous example highlighting the increase in tracking error with increasing source sound distance, these results show a somewhat different trend. The relative error is also much higher, and it may have to do with the positioning of the sound sources in the volume with reference to the array. The relative location of the sound source affects

the tracking accuracy (Aubauer, 1996).



6.2: Localisation accuracy of a multi-microphone array in the field, localised with time-of-arrival-differences. A) The line-connected points (orange) represent the microphone array consisting of 11 microphones. Four microphones are in a tristar 120 array (left, tristar array with 120cm radial distance from central mic), and the remaining 7 mics are placed on the walls of the cave (top right, two quadrilateral outlines joined at a common vertex represent the 7 mics on the cave walls). The free-standing points (blue) are the simulated emission points which form a 'calibration grid'. Each simulated point is shown as a dot, and the size of the dot is proportional to the tracking error. B) The distribution of localisation error. The error is the euclidean distance between the predicted and simulated point. The localisation error is between 0.07-0.32 m for the given points. C) The relative localisation error with reference to the central mic of the 120 cm tristar array microphone. The 95%ile bounds of tracking error lie between 3.7-16.6%, with a maximum of 18.8% error. Even points that are nearby seem to be localised with a higher relative error. This higher relative error may be the result of sound source position with reference to the microphone array.

6.5 Future directions

tacost as it stands is currently written to implement a first-order assessment of a tracking system's accuracy. The package has been primarily written keeping acoustic signals propagating through air where the velocity of sound is assumed to be constant. It may also be used to test tracking in radar or underwater sonar systems, contingent on how uniform the medium of wave propagation is over the distances being studied. As of version 0.1.0 ,straight line propagation of signals are simulated, without spherical spreading or atmospheric absorption implemented. Future releases may include such propagation losses. Another important aspect affecting all tracking systems is the directionality of the sensors (microphones) and emitted signals (animal vocalisations, calibration speakers). A common problem in acoustic tracking with bats and cetaceans is not being able to track animals because their echolocation calls can be very directional (Matsuta et al., 2013; Surlykke, 2012; Koblitz et al., 2016). Implementing sensor and source sound directionality will help assessing how many microphones might be required to successfully track animals in their surroundings, and which array geometries are best able to do so.

6.6 Acknowledgements

This work was supported by a doctoral fellowship from the German Academic Exchange Service (DAAD) and the International Max Planck Research School for Organismal Biology. I would like to thank Léna de Framond for generating the acoustic localisation output, Holger R Goerlitz for helpful comments on this manuscript and discussions on the topic of tracking, and the IT team at the Max-Planck Institute for Ornithology for their support.

7

itsfm, an open-source package to reliably segment and measure sounds by frequency modulation

This chapter was published as a preprint on *biorXiv*:

Beleyur, T. (2020). *itsfm, an open-source package to reliably segment and measure sounds by frequency modulation* bioRxiv 2021.01.09.426033; doi: <https://doi.org/10.1101/2021.01.09.426033>

```

## 
## Attaching package: 'flextable'

## The following objects are masked from 'package:kableExtra':
## 
##     as_image, footnote

```

Abstract

Analysing animal vocalisations in detail provides insights into the biomechanics, decision making and sensory processes behind their behaviours. Echolocating bats, and in particular, the CF-FM calls of high-duty cycle bats serve as a convenient model system to illustrate this point. The CF component in the CF-FM call is used for prey detection and the FM component is used in target ranging. According to the behavioural context at hand such as flight with conspecifics or prey capture, bats choose to increase the duration, intensity or spectral range of the components differently. Studying the call component alterations requires an objective methodology that first segments the components and then allows measurements on them. Studies till now have segmented the call components manually, or automatically using what I term the ‘peak-frequency’ method. Manual segmentation is error prone, while the ‘peak-frequency’ method requires on-axis recordings for good results. Despite multiple papers using a peak-frequency based segmentation, there remain no publicly available software implementations. `itsfm` is an open-source package that fills this gap with two implemntations that can segment CF-FM calls, one of them being an implementation of the peak-percentage method. `itsfm` additionally introduces the ‘pseudo-Wigner-Ville distribution’ (PWVD) method for call segmentation, thus allowing the segmentation of calls captured under a wider variety of recording conditions. I create a synthetic dataset and assess the performance of the PWVD method and the ‘peak-frequency’ method. The PWVD performs consistently well in call component segmentation in comparison to the peak-percentage method. I also discuss the supporting methods in the `itsfm` package that can help the further automatic segmentation, measurement and analysis of sounds. Though originally developed for the segmentation and measurement of CF-FM bat calls, the methods in `itsfm` are species-agnostic, and may be used for vocalisations of any type.

7.1 Introduction

Vocalisations are a window into the sensory, behavioural and biomechanical states of an animal (Green and Marler, 1979; Metzner and Müller, 2016). Echolocating bats present a unique model system where vocalisations play a fundamental role in the animal's sensorimotor decisions. Echolocating bats emit loud calls and listen for returning echoes to detect objects around them (Griffin, 1958b). Bats are known to flexibly alter various aspects of their calls to optimise echo detection, and thus their own sensory input. For instance, bats flying in the open emit long calls with a narrow bandwidth, and switch to short high-bandwidth sweeps as they are about attack an insect prey (Fenton, 2013b).

Among echolocating bats, the so-called CF-FM bats are a particularly interesting model system to study sensorimotor decisions. CF-FM calls (Figure 7.1) consist of a constant-frequency (CF) component and upto two frequency-modulated (FM) components. The CF component is used in the detection of prey wing-flutter (Schnitzler and Denzinger, 2011) while the FM component is used in target ranging (Tian and Schnitzler, 1997). Bats are known to independently alter the CF and FM components depending on the presence of echolocating conspecifics (Fawcett et al., 2015), artificial playbacks (Lu et al., 2020; Hage et al., 2013, 2014) or during flight manuevers (Tian and Schnitzler, 1997; Schoeppler et al., 2018). Studying how CF-FM bats alter their call components requires an objective method that can reliably segment the components, and thus facilitate accurate acoustic parameter measurement.

7.1.1 State of the art: CF-FM call segmentation

Manual segmentation of calls into CF and FM is the most intuitive and direct one can take, and has been an approach used in publications to date (Vater et al., 2003; Fawcett et al., 2015; Gessinger et al., 2019). Manual segmentation however doesn't scale with sample size, are not very reproducible and can be biased (Brumm et al., 2017). Tian and Schnitzler (1997) is to my knowledge, the first publication to attempt a semi-manual segmentation of the CF and FM call components, and their method has formed the founding basis for further work. I hereby refer to methods based on their approach as the 'peak percentage' approach. Tian and Schnitzler (1997) method relies on the fact that the CF component is at the highest frequency and forms a large part of the call. By filtering below and above a threshold frequency close to the CF frequency, the FM and CF components can be separated. In *Rhinolophus ferrumequinum*, Tian and Schnitzler (1997) define the threshold frequency at 0.8kHz below the 2nd harmonic of 'CF component'. 0.8 kHz corresponds to around 1% of the CF peak frequency (~80 kHz), and is also equivalent to filtering at 99 % of the CF peak frequency. In their semi-manual method Tian and Schnitzler (1997) measured the CF peak frequency using an FFT frequency

analyser to separate FM and CF components. Schoeppler et al. (2018) further automate the method of Tian and Schnitzler (1997) by using 99% of the CF frequency to define FM components in a spectrogram based method run on a computer. Lu et al. (2020) follow on the methodology of Schoeppler et al. (2018) and Tian and Schnitzler (1997), and set the FM to begin at 97% of the CF peak frequency. Peak-percentage type approaches allow a straightforward segmentation and measurement, however the method was developed keeping on-axis, high signal-to-noise ratio recordings in mind, such as those that are obtained in flight room experiments. For instance, a pre-requisite for the peak-percentage method to work is a spectrally dominant CF component, in the absence of which the threshold frequency is not identified correctly, leading to poor segmentation. The peak-percentage method also requires setting a reasonable peak-percentage to define the threshold frequency that determines where the CF ends and FM begins. Previous studies have used percentages between 97-99% of the CF peak frequency, and the exact percentage is likely to play a big role in segmentation accuracy.

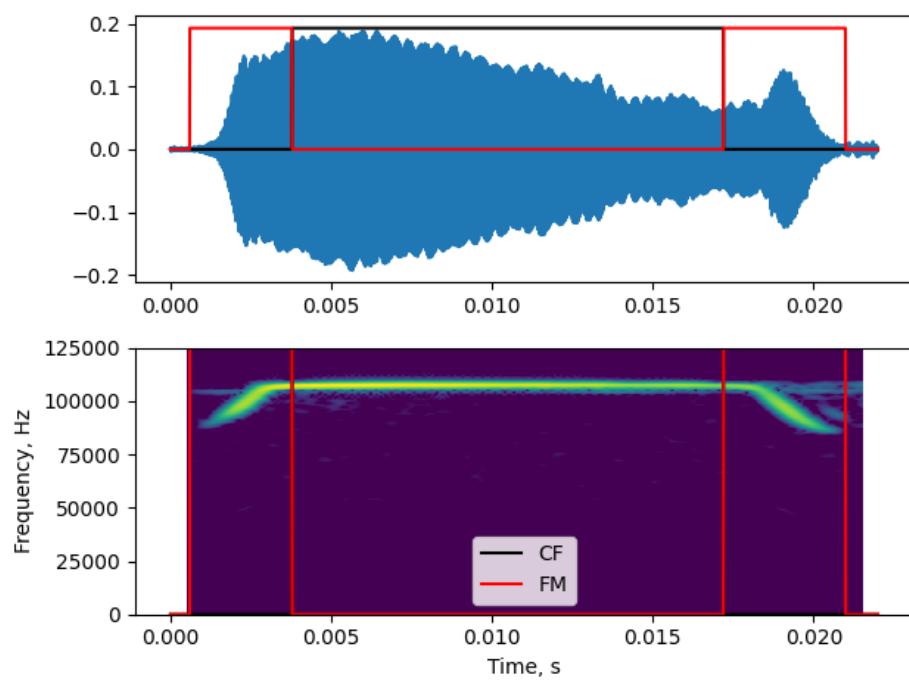
Despite the apparent popularity of the peak-percentage method there are no openly available code implementations that have been tested for their performance against synthetic data. While code descriptions help explaining the principles behind design, it is not sufficient to ensure uniformity or correctness in implementation. Differences in implementation may lead to differences in scientific results (Baker and Vincent, 2019; McFee et al., 2018). Publishing code as publicly available packages allows for external code inspection and improvements. `itsfm` fills the gap by implementing the peak-percentage method and introducing an alternate segmentation method. The segmentation methods are tested against synthetic datasets, with a open-source code base written in a non-proprietary language, and supported by a detailed user-guide online.

7.2 Package description

`itsfm` currently provides two main approaches to segment the CF and FM components of a sound (Figure 7.1), the ‘peak-percentage’ and ‘pwvd’ methods.

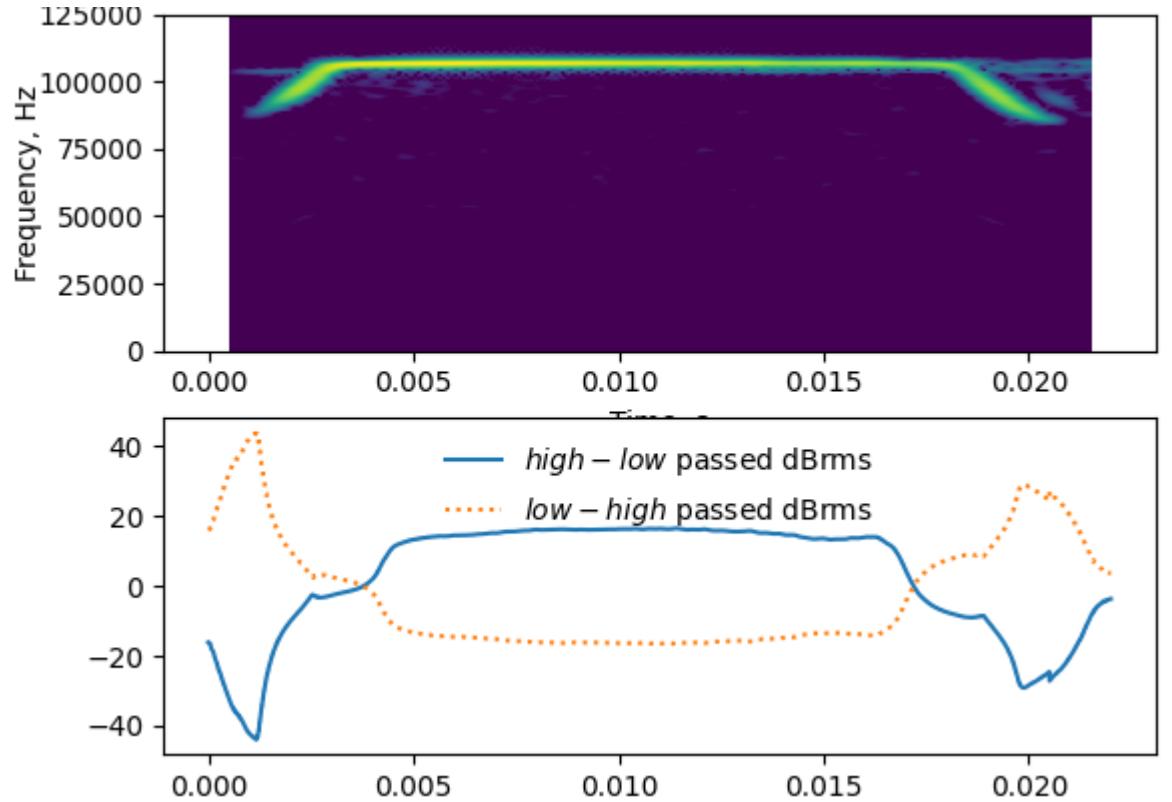
7.2.1 Peak-percentage segmentation

The **peak-percentage** method is best for sounds with one or more dominant CF components of the same frequency, and FM components that are below the CF component’s frequency (Figure 7.2). A typical rhinolophid/hipposiderid CF-FM call is the simplest example for which this method works. This method’s implementation is inspired by previously published efforts to segment CF-FM calls into their respective components (Lu et al., 2020; Tian and Schnitzler, 1997; Schoeppler et al., 2018). The approach implemented here creates two



7.1: Diagnostic plot showing the CF/FM segmentation output of a *Rhinolophus euryale/mehelyi* call

versions of the raw audio that are low and high passed at a threshold frequency. The threshold frequency is calculated as a fixed percentage of the raw audio's peak frequency, eg. 99%. The dB rms profile of the low and high passed audio are then calculated and compared by subtraction. Continuous regions where the low-passed audio is greater than the high-passed audio are considered FM regions, and CF regions where it is vice-versa (Figure 7.2).



7.2: Diagnostic output showing the underlying basis of the peak-percentage CF-FM segmentation method. Top: A spectrogram representation of the call shown in Figure 1, Below: the high/low-passed dB rms profiles of the call. The peak frequency of the entire call is taken and a high and low-passed version of the sound if created at 99% of the call peak frequency. The dB rms profile differences of the high and low-passed sounds are calculated and subtracted from each other. The region where the low-passed dB rms profile is higher are labelled FM and vice-versa as CF.

The peak-percentage method is relatively easy to parameterise as it accepts two intuitive input parameters, the `peak_percentage` (peak percentage value between 0-1) and `window_size` (the number of samples for the window used to calculate the dB rms profile of the high/low passed audio). A set of additional optional parameters may also be specified. The default low/high pass filter is a second order elliptic filter with 3dB ripple (pass band) and 10dB minimum

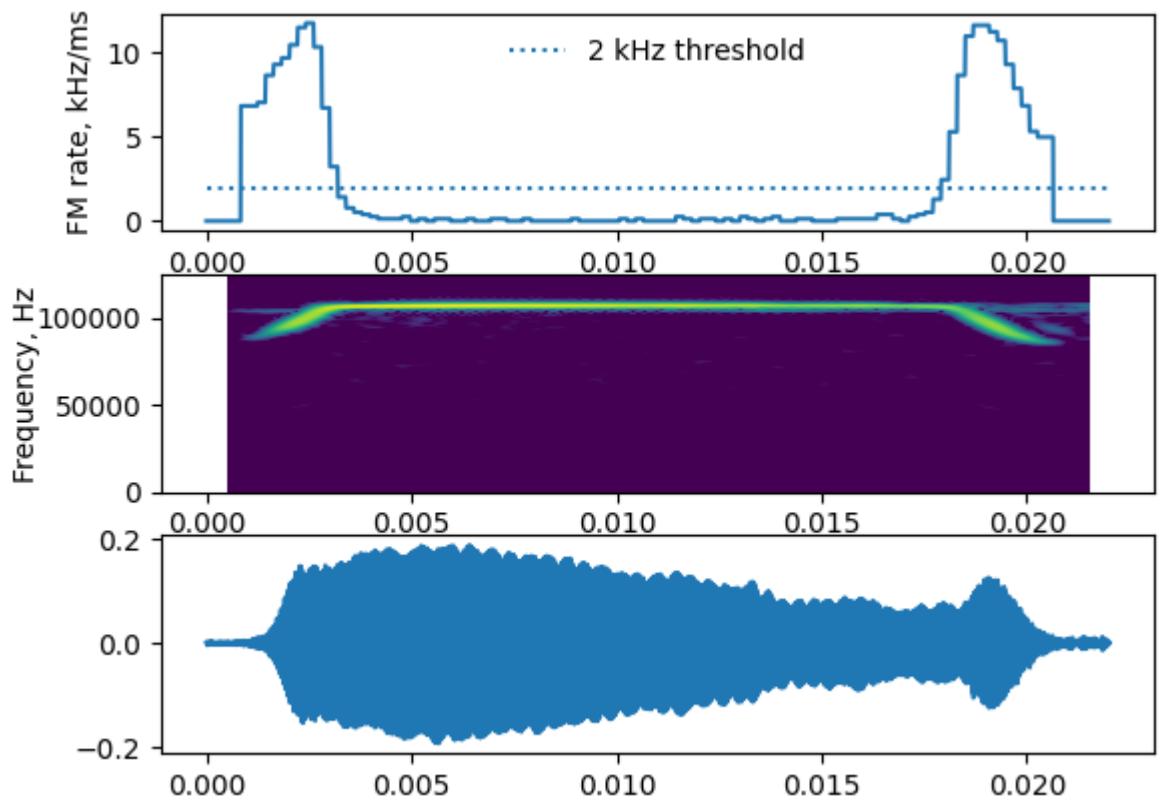
attenuation in the stop band. The user may also optionally specify their own recursive filter coefficients.

A major drawback in the peak-percentage method is its limited use-cases. Sounds must be sufficiently similar to the ‘ideal’ spectro-temporal shape of a classic CF-FM call, or they will be mis-segmented. Not even all CF-FM calls are likely to be segmented properly, eg. CF-FM calls emitted during landing or approach with short CF segments and longer FM segments. If the CF segment of the input sound does not contribute majorly to the spectrum, then the peak-percentage method fails. Experience with field recordings having off-axis CF-FM bat calls shows that the peak-percentage method also fails here because the CF component may not be as dominant as in on-axis recordings of the same call. Aside from CF-FM echolocation calls, the peak-percentage method may also be used for certain types of bird calls with long CF and short FM calls (eg. those emitted by the *Pachycephala* genus)

7.2.2 PWVD segmentation

The `pwvd` method (Figure 7.3) tracks the frequency modulation over the course of the input sound. Regions with an above threshold frequency modulation are considered FM regions, and those below are considered CF regions. The frequency modulation over the course of a sound is estimated by first generating a sample-level ‘frequency profile’ through the use of the Pseudo Wigner-Ville Distribution (PWVD). The PWVD is a relatively underutilised method in bioacoustics (but see Fu and Kloepper, 2018; Kopsinis et al., 2010) which generates time-frequency representations with high spectro-temporal resolution (Boashash, 2015). The first derivative of the frequency profile is used to generate a sample-level estimate of frequency modulation and thus segment regions that are above or below the threshold.

The `pwvd` method requires somewhat more parametrisation and methodological understanding than the `peak-percentage` method. The `pwvd` method’s effectiveness is dependent on the `fmrace_threshold` (frequency modulation threshold, in kHz/ms), `pwvd_window` size (number of samples used to form the ‘slices’ of the time-frequency representation) and `tfr_cliprange` (permitted dynamic range in dB, used to clip the time-frequency representation and remove noise). In addition to these primary parameters, the `pwvd` method can be further fine-tuned to improve segmentation. The frequency profile is currently generated by tracking the dominant frequency over each slice of the PWVD representation. The dominant frequency approach is susceptible to noise and changes in sound levels over time, and thus requires additional correction routines that interpolate between problematically tracked regions. The problematic regions are identified by measuring the acceleration (second derivative) of the sound’s frequency profile. Regions above a user-set threshold are considered ‘spiky’ and are interpolated or extrapolated based on neighbouring regions frequency estimates.



7.3: Diagnostic plots of the pwvd method. Top: Sample-level frequency modulation rate estimates. All regions \geq the threshold FM rate (here 2kHz/ms) are considered FM regions, while all regions below this are considered CF regions. Middle-Bottom: spectrogram and waveform of the original sound for comparison.

Even though the `pwd` method requires some initial effort to parameterise, the flexibility it provides allows the analysis of a much wider-range of sounds than the `peak-percentage` method. The CF/FM segmentation in the `pwd` method is independent of the actual call shape, and even complex sounds such as bat social calls and bird songs could be segmented through this method. A major drawback of the current `pwd` implementation is its inability to reliably segment multi-harmonic sounds. Multi-harmonic sounds present a challenge for the simple frequency tracking in place currently, and alternative algorithms will be a focus of future development.

7.2.3 Supporting `itsfm` methods

Along with the primary segmentation methods, `itsfm` has a collection of supporting methods that allow quantification, visualisation and batch-processing. A series of inbuilt measurement functions allow acoustically relevant measurements such as duration, rms, peak-frequency, or terminal frequency. Custom measurements may also be specified by the user. A sound analysed with the `pwd` method generates more than the identified CF/FM regions. Raw data on the frequency profile of the sound and the rate of frequency modulation over time are of interest to researchers studying the speed at which vocalisations can be modulated from a behavioural and biomechanical viewpoint (Metzner and Müller, 2016; Hage et al., 2013). Along with the background data used to form the segmentations, `itsfm` also provides a series of inbuilt visualisation functions to visualise the input sound itself (`visualise_sound`) and generate diagnostic plots of the segmentation output through the `itsFMIInspector` class and `visualise_cffm_segmentation` (Figure 7.1).

Handling audio recordings made in the field calls for the individual handling of each recording. To aid the reproducible processing of multiple files with unique input parameters `itsfm` can also be called through a command-line interface that accepts batch files in the CSV format. To facilitate iterative parameter optimisation, the user can choose to select only a few audio recordings or the entire set of files defined in the batch file. For each processed audio file, the diagnostic plot and measurements are saved in the working directory.

The `itsfm` package also comes bundled with a series of field recordings of bat calls of various *hipposiderid*, *rhinolophid* and *noctilionid* species. These field recordings allow the user to test the utility of the methods in the package, and gain familiarity with setting correct parameters.

7.3 Methods evaluation

7.3.1 Synthetic dataset creation and segmentation

To test the accuracy of the segmentation methods implemented in the `itsfm` package, I generated a set of synthetic CF-FM calls with known segment durations and spectral properties. Synthetic calls were generated based on calls broadly based on the structure of rhinolophid and hipposiderid call parameters using the package's inbuilt `make_cffm_call` function. A set of 324 synthetic calls were made through a combination of parameters in Table 7.1. Each synthetic call consisted of an iFM, CF and tFM component (naming as per (Tian and Schnitzler, 1997)), and is Tukey windowed without any padded silent samples or background noise. All synthetic calls were generated at a sampling rate of 250kHz.

7.1: Parameter values used to generate synthetic CF-FM calls. The parameters broadly reflect the call shape of a rhinolophid/hipposiderid CF-FM bat calls. iFM and tFM regions were generated from the same FM parameter set. 9 CF x 6 iFM x 6 tFM combinations = 324 calls

Parameter name	Values
CF duration (ms)	5, 10, 15
CF peak frequency (kHz)	40, 60, 90
i/tFM duration (ms)	1,2
i/tFM bandwidth (kHz)	5,10,20

The synthetic calls were segmented according to method-specific parameters that were optimised based on trial-and-error on a smaller representative batch. The parameter values used for both segmentations are shown in Table 7.2.

7.2: Segmentation method specific parameters used to analyse the synthetic data.

Method	Parameters				
pwvd	Window size (samples)	FM rate threshold (kHz/ms)	Accelaration threshold (kHz/ms ²)	Extrapolation window(s)	
	125	2	10	75x 10 ⁻⁶	

Method	Parameters		
peak percentage	Window size (samples)	Peak percentage 0.99	Double pass True
	125		

The accuracy of segmentation was determined by comparing the duration of the obtained call components and the original values used to make the synthesised calls. The accuracy of other parameters eg. CF peak frequency, FM bandwidth was not assessed. It follows directly that if the call components have been poorly segmented, any measurements made from the underlying audio will also be unrepresentative of the actual call parameters. Some calls appeared to have more than three components due to false positive CF/FM identifications, and were not included in the accuracy calculations.

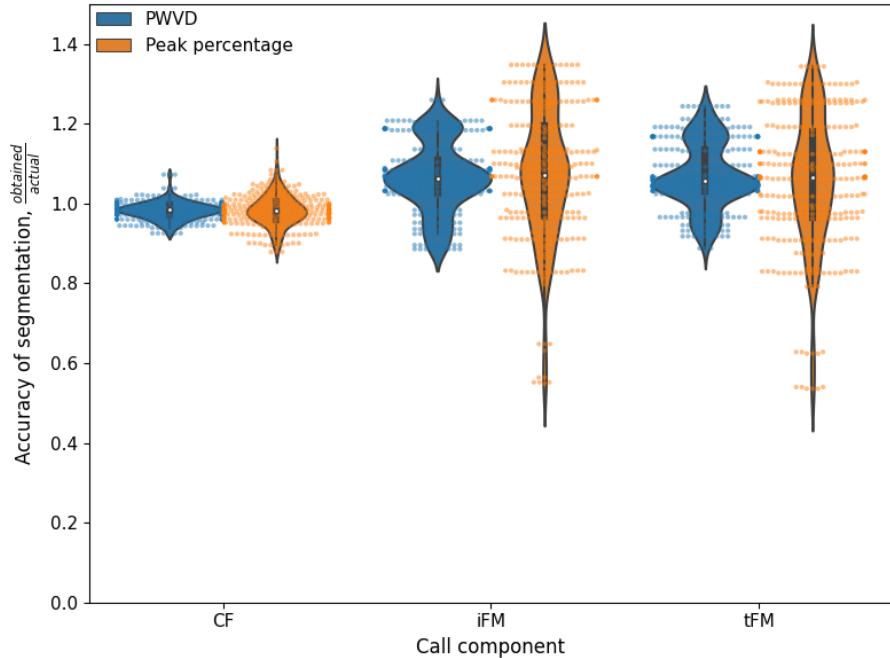
7.3.2 Results

The `pwd` method correctly identified 99% of all calls (322/324) as having only 3 components. The `peak_percentage` method correctly identified 94% of all calls as having 3 components (306/324). Both segmentation methods achieved a satisfactory performance. The `pwd` method was superior in its segmentation accuracy to the `peak_percentage` method across all the parameter combinations and call components tested (Table 7.4, 7.3.2).

The relatively lower overall performance of the `peak_percentage` method can be specifically attributed to the call properties of certain synthetic calls. A further inspection of calls with lower than 0.8 accuracy in component duration revealed that calls with a high CF frequency (60 and 90 kHz) and at least one low bandwidth FM component (5kHz) were segmented with lower accuracy. This is explained by the fact that the peak-percentage of the recursive filter is set at 0.99 of the peak frequency. A low FM bandwidth call with a high CF frequency will have its cutoff frequency much below the actual CF frequency (600 and 900 Hz below peak frequency here). The lower cutoff frequency will thus lead to a shorter duration estimate of the low-bandwidth FM component. The accuracy of component durations was above 0.8 for all calls segmented with the `pwd` method.

One percent of all `pwd` segmented calls and six percent of all `peak_percentage` segmented calls had more than three detected call components. What caused the false positive call component detections in the `pwd` and `peak_percentage` methods? In the `peak_percentage` method, the false component detections consisted of very short ($\leq 0.1\text{ms}$) falsely detected CF and FM segments located

next to one another. These neighbouring CF and FM segments were caused by brief alterations in the dB rms levels of the high and low-passed audio. The brief alterations in the dB rms levels are likely due to the combination of windowing function applied on the synthetic calls and edge effects during high/low pass filtering. Such edge effects may not necessarily occur during the processing of experimentally recorded calls, which may have smoother roll-offs in call level. The two cases where false components were detected with the `pwd` method were borderline cases where false CF components were detected in what should have been an FM region of the call. On further inspection it was shown that the frequency tracking of these false CF components was indeed accurate, but the action of the error-correction routines caused a slight drop in the frequency modulation rate to 1.9 kHz/ms, just slightly below the threshold of 2.0 kHz/ms. The error-correction routines in `pwd` are typically required when low signal-level at the beginning and ends of the call causes jumps in the frequency tracking.



7.4: Accuracy of call component segmentation of the synthetic test data set shown with raw data overlaid on violinplots. The accuracy is calculated as the measured call component duration by the original duration. Blue violinplots: accuracy of the `pwd` method, orange violinplots: accuracy of the peak-percentage method. The `pwd` method is superior to the peak-percentage method in its segmentation performance across call components.

Call component | Segmentation method | Segmentation accuracy ($\frac{\text{measure duration}}{\text{original duration}}$)
 |
 | | : 95%ile range |

CF peak percentage 0.9-1.06	CF pwvd 0.94-1.02	tFM peak percentage 0.65-1.35	tFM pwvd 0.89-1.21
iFM peak percentage 0.62-1.3	iFM pwvd 0.92-1.24		

Table: Summary statistics describing the performance of the two segmentation methods on the synthetised test data set. The pwvd method performs better than the peak-percentage method over the tested parameter space and for all call components (iFM,tFM and CF). The relative segmentation

7.4 Discussion

Software based automation in acoustic analysis is an important step in ensuring reproducible results, which in turns spurs the growth of the research field (McFee et al., 2018; Baker and Vincent, 2019). The **itsfm** package written in the Python (Van Rossum and Drake Jr, 1995) language is an open-sourced method which may be used in the analysis of animal vocalisations such as CF-FM bat calls, and other vocalisations. The **itsfm** package has already been successfully used to segment and measure call parameters in an upcoming publication on group echolocation in CF-FM bats (Mysuru Rajagopalachari et al., 2020). The package introduces a new method the ‘pwvd’ method to segment CF and FM components based directly on the rate of frequency modulation. The ‘pwvd’ method also performs consistently better than the ‘peak-percentage’ method, and is thus the recommended segmentation method to use, at least for sounds that resemble CF-FM calls.

The use of **itsfm** in the analysis of other types of vocalisations still needs further explored. For instance, bird calls have been analysed (See online user-guide). The current ‘pwvd’ frequency tracking implementation only tracks a single frequency per point of time, and thus is not able to handle multi-harmonic sounds with equal harmonic emphasis very well. Future implementations of frequency tracking need to apply more sophisticated problem-region detection and also frequency tracking (eg. Viterbi path).

7.5 Open-source software and packages used

itsfm is written in the Python language (Van Rossum and Drake Jr, 1995), and relies on the numpy, scipy, pandas, matplotlib and tftb (Oliphant, 2006b; Virtanen et al., 2020; Hunter, 2007b; McKinney et al., 2010; Deshpande, 2019). The Jupyter Notebook and Rmarkdown projects (Kluyver et al., 2016; Xie et al., 2018) were used in the analysis of data and writing of this paper.

7.6 Supporting information

The `itsfm` package can be installed from the Python package index (PyPi) with the command `pip install itsfm`. The latest versions of the package and drafts of this paper are accessible at <https://github.com/thejasvibr/itsfm>. Online documentation with detailed examples and troubleshooting guides can be accessed at <https://itsfm.readthedocs.io>

7.7 Acknowledgements

I would like to thank Diana Schoeppler for sharing know-how on analysing CF-FM calls and Neetash MR for helpful discussions. This work was funded by the DAAD and the IMPRS for Organismal Biology. I'd like to thank the following people for contributing to the call recording library Aditya Krishna, Aiqing Lin, Gloria Gessinger, Klaus-Gerhard Heller, Laura Stidsholt and Neetash MR.

8

General Discussion

8.1 Active sensing : a tricky and yet manageable situation.

- The ‘cocktail party nightmare’ may be apt for how we perceive it, but for bats the term may perhaps be a bit hyperbolic.
- All models are wrong, but some are useful - while I would alter this statement to say that all biological models are caricatures, and some of them are closer to the complex network of processes in play.
- So many simplifications I’ve done in the conceptual framework itself - there is scope for a more systematic experiments to actually recreate the ‘auditroy masking function of bats’, and also estimate spatial unmasking in more species. Does the ability of a species to detect echoes in noise depend on the typical group size it lives in over the course of its life — or is the ability to detect echoes just a by-product of noise rejection that individuals need to echolocate to orient in complex environments (echo-echo masking, call-echo masking, etc.)
- Bats do collide into each other, but it’s not like they suffer major consequences??

8.2 *Ushichka* as a model dataset for group active sensing

-
- Developing the methodology and investing in the equipment is not an easy task. eg. how I see Ushichka as an astronomy-type dataset where large multi-disciplinary collaborations can be built around it.

- Implications of group echolocation on collective behaviour

8.3 Lowering the technical barriers to study group echolocation

- measurement-free, array-free acoustic tracking setups. The more scaleable your mic array is - the better you can track echolocating bats.
- package to estimate and simulate the acoustic tracking performance pre-hoc or post-hoc
- package to
- What we need is

8.4 Sensory systems

- What other animal systems could be studied using the modelling and experimental approaches described here: cave swiftlets, lingual-echolocating bats, cetaceans.
- As pointed out by (Nelson and MacIver, 2006) In some sense lingual-echolocating bats + cetaceans may not suffer from overlaps as such because the calls are so short – but given how short the sounds are - maybe they suffer more from informational masking??
- Informational masking - or ‘attentional processes’ are another very important part of the whole process that hasn’t been answered before. Auditory masking has mainly been studied in terms of ‘energetic masking’ in echol. bats. Controlled playback studies looking into informational masking still need to be there (eg. as Lutz pointed out - moth clicks are not v ‘loud’, but may be good starting point?).

8.5 Inverse problems all the way

9

Acknowledgements

*“An alle erschienenen und unerschienenen Erscheinungen
Menschen, Tiere, Meinungen, Gefühle, Fraktale
Zikaden, Mutanten, Schamanen, Fackelträger
Geistreisende und mikroencyklopen-jagende Weichorganismen
Es ist an der Zeit einen Kreis zu bilden”¹*
- Käptn Peng & Die Tentakel von Delphi, Der Anfang ist nah (The beginning is nigh)

And what a journey it has been, this Phd'fying for the past five plus years. Sitting in front of the screen right now, I never really realised how much I've relied on all the amazing people and things around me to make it this far. Let's begin, starting from the office outwards in an ever growing radius from where I sit right now.

To my office-mates: Daniel who taught me the value of a nice silent office and the joy of handling bats, Theresa - your willingness to help and chat about all things science and random, and for introducing me to Käptn Peng. Danke Danke. To all other former and current AFEG and associated members down the corridor (Aiqing, Erin, Klemen, Lena, Leo, Stefan, Laura, Toni, Neetash, Verena + both Claires) thank you all for forming a great and supportive and enthu environment to come to work every day. A big shout-out to Toni for making the whole 2018 field season possible at all! Holger, it has been awesome working with you - and I owe you for teaching me the value of good presentation and rigour in science!

Haus 11, the out-house where all the cool people are, here's to all of you current and former Haus 11'ers for the random corridor conversations and barbecue evenings! Without the Narada/Hermes/Loki messenger-like figures (who roam across many worlds and are bearers of all kinds of knowledge) Nicole Drenkard

¹To all manifested and unmanifested manifestations, humans, animals, opinions, feelings, fractals, cicadas, mutants, shamans, torchbearers, spirit-travelling and micro-encyclic hunting molluscs, it is about time to form a circle.

and Diana Werner, many things would not have been possible logistically and administratively - here's a big thanks to both of you. Joerg at the Pfore : quatsch'ing with you has always been a joy every time I came to pick up chocolate or a parcel. The Max-Planck Institute for Ornithology, my place of work, I thank for having the best work-place conditions (and location) and support systems I have ever experienced. I always looked forward to arriving at work. From the people around the instit now to all things and beings around: the peaceful snow outside my window, the lake, the beautiful cycle-path to the institute, and the occasional deer eating away peacefully - I am in some way thankful to these experiences and many more that have directly or indirectly come to this point in time and space.

To my more distantly located colleagues in my graduate school, the IMPRS for Organismal Biology and all of you at the Ghetto - I thank you all for forming a supportive net to share the joy and vent the frustrations of a life in science. Special thanks to the grad school coordinators Maeggi Hieber-Ruiz, Fransisca Rosa Mende and Corinna Loes for readily providing support whenever asked for.

Neetash, for simultaneously playing the multiple roles of colleague, friend and wife over the past few years - I can not say enough, and perhaps words will not do justice! I can only be thankful that you have been by my side through the ups and downs as I tried to phd my way through to this point. Reni, for providing an oasis to retreat from work while Qwirkl-ing away, and Dieter for being the amazing conversationalist and story-teller he is.

My extended family in Germany: Apurva-Sindhu, Deepu-Ravi. My gratitude to you for creating a home away from home and a second base-camp over the past many years with their hospitality, food and love! Parts of this thesis literally wouldn't be written without all encouragement that came from giving me the space and time to write away in your homes along with all spiritual nourishment from the hot cups of chai and bonda-bajji's that came with it.

Amma-Appa, for constantly supporting all my weird ideas of becoming, at various points, a vet and/or a 'zoologist' (even before I actually knew what the word meant), etc! Praj - for being my go-to person in times of joy and confusion for work and life related matters. I am here only because of the unquestioning 'I-have-your-back' attitude you all have shown at each decision point that has come till now.

And finally, I would not be here without all the bats flying through the night and in the caves far far away in Bulgaria. All the bats that hung around as I fumbled with cables with numb hands on cold autumn nights, as I learnt to patiently deal with (hundreds of?) metres of tangled cable every night, and set up equipment hoping they'd land up. When they did it was always a pleasure to watch and hear them. So here's to all you hovercraft Daubis, hyper pipistrelles and gliding Mouse-eared bats, none of this would have been possible without all of you - and this is looking to spend more years to come understanding you

weird creatures!

To everyone and everything mentioned above, and especially to all those who I have not specifically mentioned - I convey my gratitude: thank you, Danke,

,

10

Author contributions

Summary, General introduction and Discussion : written by Thejasvi Beleyur.
Zusammenfassung written by ZZASDFASDFASDFZZASDFASDFASDFZZASDFASDFASDFASDFZZ

Chapter 2: Thejasvi Beleyur (TB) and Holger R Goerlitz (HRG) conceived the study. TB formulated the computational model, wrote the code, analysed data and presented the results. TB wrote the first draft, and HRG provided input for later drafts.

Chapter 3: TB and Neetash Mysuru Rajagoplachari (NMR) conceived the study and carried out field data collection. NMR and Aditya Krishna formulated and annotated bat flights in video data. TB wrote the code to synchronise audio-video data, analysed audio, performed statistical analysis, and prepared figures. TB and NMR wrote the first draft, and HRG provided input for later drafts.

Chapter 4: TB designed and executed field data collection and wrote code to control the recording system. HRG conceived and designed the recording system. TB wrote the first draft with observations and figures, HRG provided input for later drafts.

Chapter 5: TB initiated the inter-disciplinary collaboration, carried out field data collection and wrote code to control the recording system. HRG conceived and designed the recording system. TB and HRG wrote the section describing data collection in the Orlova Chuka cave system. All other authors (Kenneth Batstone, Gabrielle Flood, Viktor Larsson, Magnus Oskarsson, Kalle Åström) were involved in the writing of other paper sections, analysis and presentation of results.

Chapter 6: TB wrote the code, analysed data, presented results and wrote the manuscript.

Chapter 7: TB wrote the code, analysed data, presented results and wrote the manuscript.

- Adams, A., Davis, K., and Smotherman, M. (2017). Suppression of emission rates improves sonar performance by flying bats. *Sci. Rep.*, 7:41641.
- Amichai, E., Blumrosen, G., and Yovel, Y. (2015). Calling louder and longer: how bats use biosonar under severe acoustic interference from other bats. *Proceedings of the Royal Society B: Biological Sciences*, 282(1821):20152064.
- Anaconda (2016). Anaconda software distribution.
- Aodha, O. (2018). Bat detective-deep learning tools for bat acoustic signal detection. *PLOS Comput. Biol.*, 14:1005995.
- Aubauer, R. (1994). Dreidimensionale flugbahnverfolgung von fledermäusen fortschritte der akustik-daga 94. *DPG-Verlag, Bad Honnef, Germany*.
- Aubauer, R. (1996). *Acoustical flight path tracking of echolocating bats*. PhD thesis, Acoustical Society of America.
- Baker, E. and Vincent, S. (2019). A deafening silence: a lack of data and reproducibility in published bioacoustics research? *Biodiversity Data Journal*, 7:e36783.
- Ballerini, M., Cabibbo, N., Candelier, R., Cavagna, A., Cisbani, E., Giardina, I., Lecomte, V., Orlandi, A., Parisi, G., Procaccini, A., 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):1232–1237.
- Bates, M., Stamper, S., and Simmons, J. (2008). Jamming avoidance response of big brown bats in target detection. *J. Exp. Biol.*, 211:106–113.
- Batstone, K., Floor, G., Beleyur, T., Larsson, V., Goerlitz, H. R., Oskarsson, M., and Åström, K. (2019). Robust self-calibration of constant offset time-difference-of-arrival. *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, United Kingdom, 2019*, 2019(2):4410–4414.

- Batstone, K., Oskarsson, M., and Åström, K. (2016). Robust time-of-arrival self calibration with missing data and outliers. In *Signal Processing Conference (EUSIPCO), 2016 24th European*, pages 2370–2374. IEEE.
- Bee, M. and Micheyl, C. (2008). The cocktail party problem: What is it? how can it be solved? and why should animal behaviorists study it? *J. Comp. Psychol*, 122:235–251.
- Belden, J., Mansoor, M. M., Hellum, A., Rahman, S., Meyer, A., Pease, C., Pacheco, J., Koziol, S., and Truscott, T. T. (2019). How vision governs the collective behaviour of dense cycling pelotons. *Journal of the Royal Society Interface*, 16(156):20190197.
- Blumstein, D. T., Mennill, D. J., Clemins, P., Girod, L., Yao, K., Patricelli, G., Deppe, J. L., Krakauer, A. H., Clark, C., Cortopassi, K. A., Hanser, S. F., McCowan, B., Ali, A. M., and Kirschel, A. N. G. (2011). Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. *Journal of Applied Ecology*, 48(3):758–767.
- Boashash, B. (2015). *Time-frequency signal analysis and processing: a comprehensive reference*. Academic Press.
- Bode, N., Franks, D., and Wood, A. (2011). Limited interactions in flocks: Relating model simulations to empirical data. *J. R. Soc. Interface*, 8:301–304.
- Bradbury, J. W. and Vehrencamp, S. L. (1998). *Principles of animal communication*. Sinauer Associates, Sunderland, MA.
- Bridson, R. (2007). Fast poisson disk sampling in arbitrary dimensions. *ACM SIGGRAPH 2007 sketches - SIGGRAPH*, 07(22).
- Brinkløv, S., Fenton, M., and Ratcliffe, J. (2013). Echolocation in oilbirds and swiftlets. *Front. Physiol*, 4:123.
- Brumm, H. and Slabbekoorn, H. (2005). Acoustic communication in noise. *Adv. Stud. Behav.*, 35:151–209.
- Brumm, H., Zollinger, S. A., Niemelä, P. T., and Sprau, P. (2017). Measurement artefacts lead to false positives in the study of birdsong in noise. *Methods in Ecology and Evolution*, 8(11):1617–1625.
- Brungart, D. S. and Iyer, N. (2012). Better-ear glimpsing efficiency with symmetrically-placed interfering talkers. *The Journal of the Acoustical Society of America*, 132(4):2545–2556.
- Bullock, T., Hopkins, C., Popper, A., and Fay, R., editors (2005). *Electroreception*. Springer Science+Business Media, Inc, New York.

- Bullock, T. H., Hamstra, R. H., and Scheich, H. (1972). The jamming avoidance response of high frequency electric fish. In *How do Brains Work?*, pages 509–534. Springer.
- Cherry, E. (1953). Some experiments on the recognition of speech, with one and with two ears. *J. Acoust. Soc. Am.*, 25:975–979.
- Chum, O., Matas, J., and Kittler, J. (2003). Locally optimized ransac. In *Joint Pattern Recognition Symposium*, pages 236–243. Springer.
- Cook, J. (1985). Target strength and echo structure. In HG, U., editor, *Adaptive Methods in Underwater Acoustics*, page 155–172. Reidel Publishing Company.
- Corcoran, A. and Moss, C. (2017). Sensing in a noisy world: Lessons from auditory specialists, echolocating bats. *J. Exp. Biol.*, 220:4554–4566.
- Couzin, I., Krause, J., James, R., Ruxton, G., and Franks, N. (2002). Collective memory and spatial sorting in animal groups. *J. Theor. Biol.*, 218:1–11.
- Culling, J. and Stone, M. (2017). Energetic masking and masking release. In Middlebrooks, J., Simon, J., Popper, A., and Fay, R., editors, *The Auditory System and the Cocktail Party*. ASA Press.
- Cvikel, N., Berg, K. E., Levin, E., Hurme, E., Borissov, I., Boonman, A., Amichai, E., and Yovel, Y. (2015a). Bats aggregate to improve prey search but might be impaired when their density becomes too high. *Current Biology*, 25(2):206–211.
- Cvikel, N., Levin, E., Hurme, E., Borissov, I., Boonman, A., Amichai, E., and Yovel, Y. (2015b). On-board recordings reveal no jamming avoidance in wild bats. *Proceedings of the Royal Society B: Biological Sciences*, 282(1798):20142274.
- de Framond-Bénard, L., Beleyur, T., Lewanzik, D., and Goerlitz, H. R. (2020). Low-amplitude echolocation in a free-flying gleaning bat as a pre-adaptation for aerial-hunting of eared insects. *manuscript in preparation*.
- Deshpande, J. (2019). tftb.
- Dyer, J. R., Johansson, A., Helbing, D., Couzin, I. D., and Krause, J. (2009). Leadership, consensus decision making and collective behaviour in humans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1518):781–789.
- Ebata, M. (2003). Spatial unmasking and attention related to the cocktail party problem. *Acoust. Sci. Technol.*, 24:208–219.
- Engels, C., Stewenius, H., and Nister, D. (2006). Bundle adjustment rules. In *PCV06*.

- F, D. M., J, R., H, P., M, O., and U, F. (2008). Simulated head related transfer function of the phyllostomid bat *phyllotomus discolor*. *J Acoust Soc Am*, 124(4).
- Falk, B., Jakobsen, L., Surlykke, A., and Moss, C. (2014). Bats coordinate sonar and flight behavior as they forage in open and cluttered environments. *J. Exp. Biol.*, 217:4356–4364.
- Fawcett, K., Jacobs, D. S., Surlykke, A., and Ratcliffe, J. M. (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.
- Fay, R. (2008). Sound source perception and stream segregation in nonhuman vertebrate animals. In Yost, W., Popper, A., and Fay, R., editors, *Auditory Perception of Sound Sources*, page 307–323. Springer Science+Business, Media, LLC.
- Fenton, M. (2013a). Questions, ideas and tools: Lessons from bat echolocation. *Anim. Behav.*, 85:869–879.
- Fenton, M. B. (2013b). Questions, ideas and tools: lessons from bat echolocation. *Animal Behaviour*, 85(5):869–879.
- Firzlaff, U. and Schuller, G. (2003). Spectral directionality of the external ear of the lesser spear-nosed bat, *phyllotomus discolor*. *Hear Res*, 185(1–2).
- Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–95.
- Fletcher, H. (1940). Auditory patterns. *Rev. Mod. Phys*, 12:47–66.
- Fu, Y. and Kloepper, L. N. (2018). A systematic method for isolating, tracking and discriminating time-frequency components of bat echolocation calls. *The Journal of the Acoustical Society of America*, 143(2):716–726.
- Geier, M. (2015a). Available at: [https://github.com/mgeier/python-sounddevice](#)
- Geier, M. (2015b). sounddevice 0.3.5. https://python-sounddevice.readthedocs.io/en/0.3.5/_modules/sounddevice.html.
- Gessinger, G., Gonzalez-Terrazas, T. P., Page, R. A., Jung, K., and Tschapka, M. (2019). Unusual echolocation behaviour of the common sword-nosed bat *lonchorhina aurita*: an adaptation to aerial insectivory in a phyllostomid bat? *Royal Society open science*, 6(7):182165.
- Ghose, K., Horiuchi, T., Krishnaprasad, P., and Moss, C. (2006). Echolocating bats use a nearly time-optimal strategy to intercept prey. *PLoS Biol*, 4:108.

- Ghose, K. and Moss, C. (2006). Steering by hearing: a bat's acoustic gaze is linked to its flight motor output by a delayed, adaptive linear law. *J. Neurosci*, 26:1704–1710.
- Gillam, E. H., Hristov, N. I., Kunz, T. H., and McCracken, G. F. (2010). Echolocation behavior of brazilian free-tailed bats during dense emergence flights. *Journal of Mammalogy*, 91(4):967–975.
- Gillam, E. H. and Montero, B. K. (2016). Influence of call structure on the jamming avoidance response of echolocating bats. *Journal of Mammalogy*, 97(1):14–22.
- Giuggioli, L., McKetterick, T., and Holderied, M. (2015). Delayed response and biosonar perception explain movement coordination in trawling bats. *PLoS Comput Biol*, 11(3).
- Goerlitz, H. (2018). Weather conditions determine attenuation and speed of sound : Environmental limitations for monitoring and analyzing bat echolocation. *Ecol Evol*, 8(March).
- Goerlitz, H. R. (2019). Toadsuite manual.
- Green, S. and Marler, P. (1979). The analysis of animal communication. In *Social behavior and communication*, pages 73–158. Springer.
- Greenfield, M. D. (2015). Signal interactions and interference in insect choruses: singing and listening in the social environment. *Journal of Comparative Physiology A*, 201(1):143–154.
- Grégoire, G. and Chaté, H. (2004). Onset of collective and cohesive motion. *Physical review letters*, 92(2):025702.
- Griffin, D. (1958a). *Listening in the Dark: The Acoustic Orientation of Bats and Men*. Yale University Press, New Haven, CT.
- Griffin, D. and Galambos, R. (1941). The sensory basis of obstacle avoidance by flying bats. *J. Exp. Zool*, 86:81–506.
- Griffin, D. R. (1958b). Listening in the dark: the acoustic orientation of bats and men.
- Götze, S., Koblitz, J., Denzinger, A., and Schnitzler, H. (2016). No evidence for spectral jamming avoidance in echolocation behavior of foraging pipistrelle bats. *Sci Rep*, 6(July).
- Habersetzer, J. (1981). Adaptive echolocation sounds in the bat rhinopoma hardwickei. *J. Comp. Physiol*, 144:559–566.
- Hafter, E., Sarampalis, A., and Loui, P. (2008). Auditory attention and filters. In Yost, W., Popper, A., and Fay, R., editors, *Auditory Perception of Sound Sources*, page 115–142. Springer Science+Business, Media, LLC.

- Hage, S. R., Jiang, T., Berquist, S. W., Feng, J., and Metzner, W. (2013). Ambient noise induces independent shifts in call frequency and amplitude within the lombard effect in echolocating bats. *Proceedings of the National Academy of Sciences*, 110(10):4063–4068.
- Hage, S. R., Jiang, T., Berquist, S. W., Feng, J., and Metzner, W. (2014). Ambient noise causes independent changes in distinct spectro-temporal features of echolocation calls in horseshoe bats. *Journal of Experimental Biology*, 217(14):2440–2444.
- Hartbauer, M. and Römer, H. (2016). Rhythm generation and rhythm perception in insects: the evolution of synchronous choruses. *Frontiers in neuroscience*, 10:223.
- Hase, K. (2018). Bats enhance their call identities to solve the cocktail party problem. *Commun Biol*, 1:39.
- Hase, K., Miyamoto, T., Kobayasi, K., and Hiryu, S. (2016). Rapid frequency control of sonar sounds by the fm bat, miniopterus fuliginosus, in response to spectral overlap. *Behav. Processes*, 128:126–133.
- Heiligenberg, W. (2012). *Principles of electrolocation and jamming avoidance in electric fish: a neuroethological approach*, volume 1. Springer Science & Business Media.
- Henninger, J., Krahe, R., Sinz, F., and Benda, J. (2020). Tracking activity patterns of a multispecies community of gymnotiform weakly electric fish in their neotropical habitat without tagging. *Journal of Experimental Biology*, 223(3).
- Herbert-Read, J. E. (2016). Understanding how animal groups achieve coordinated movement. *Journal of Experimental Biology*, 219(19):2971–2983.
- Holderied, M. and Von Helversen, O. (2003). Echolocation range and wingbeat period match in aerial-hawking bats. *Proceedings of the Royal Society B: Biological Sciences*, 270(1530):2293–2299. cited By 143.
- Hunter, J. (2007a). Matplotlib: A 2d graphics environment. *Comput Sci Eng*, 9(3).
- Hunter, J. D. (2007b). Matplotlib: A 2d graphics environment. *Computing in science & engineering*, 9(3):90–95.
- Hügel, T., van Meir, V., Muñoz-Meneses, A., Clarin, B. M., Siemers, B. M., and Goerlitz, H. R. (2017). Does similarity in call structure or foraging ecology explain interspecific information transfer in wild Myotis bats? *Behavioral Ecology and Sociobiology*, 71(11).
- Jakobsen, L., J.M., R., and A, S. (2013). Convergent acoustic field of view in echolocating bats. *Nature*, 493(7430).

- Jakobsen, L. and Surlykke, A. (2010). Vespertilionid bats control the width of their biosonar sound beam dynamically during prey pursuit. *Proc Natl Acad Sci*, 107(31).
- Jarvis, J., Jackson, W., and Smotherman, M. (2013). Groups of bats improve sonar efficiency through mutual suppression of pulse emissions. *Front. Physiol*, 4:140.
- Jeffery, W. R. (2001). Cavefish as a model system in evolutionary developmental biology. *Developmental biology*, 231(1):1–12.
- Jones, G., Griffin, D., and Galambos, R. (1999). Scaling of echolocation call parameters in bats. *J. Exp. Biol*, 202:3359–3367.
- Jones, G., Morton, M., Hughes, P., and Budden, R. (1993). Echolocation, flight morphology and foraging strategies of some west african hipposiderid bats. *J. Zool*, 230:385–400.
- Klein, G. and Murray, D. (2007). Parallel tracking and mapping for small ar workspaces. In *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*, pages 225–234. IEEE.
- Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., and Willing, C. (2016). Jupyter notebooks – a publishing format for reproducible computational workflows. In Loizides, F. and Schmidt, B., editors, *Positioning and Power in Academic Publishing: Players, Agents and Agendas*, pages 87 – 90. IOS Press.
- Koblitz, J. C., Stilz, P., Rasmussen, M. H., and Laidre, K. L. (2016). Highly directional sonar beam of narwhals (*Monodon monoceros*) measured with a vertical 16 hydrophone array. *PLoS ONE*, 11(11).
- Kopsinis, Y., Aboutanios, E., Waters, D. A., and McLaughlin, S. (2010). Time-frequency and advanced frequency estimation techniques for the investigation of bat echolocation calls. *The Journal of the Acoustical Society of America*, 127(2):1124–1134.
- Korman, S. and Litman, R. (2018). Latent ransac. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6693–6702.
- Kuang, Y. and Astrom, K. (2013). Stratified sensor network self-calibration from tdoa measurements. In *Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European*, pages 1–5. IEEE.
- Kuang, Y., Burgess, S., Torstensson, A., and Åström, K. (2013). A complete characterization and solution to the microphone position self-calibration problem. In *The 38th International Conference on Acoustics, Speech, and Signal Processing*.

- Kunz, T., editor (1982). *Ecology of Bats*. Plenum Press.
- Larsson, V., Åström, K., and Oskarsson, M. (2017). Polynomial solvers for saturated ideals. In *International Conference on Computer Vision (ICCV)*. IEEE.
- Lawrence, B. and Simmons, J. (1982). Measurements of atmospheric attenuation at ultrasonic frequencies and the significance for echolocation by bats. *J Acoust Soc Am*, 71(3).
- Lee, W. J., Falk, B., Chiu, C., Krishnan, A., Arbour, J. H., and Moss, C. F. (2017). Tongue-driven sonar beam steering by a lingual-echolocating fruit bat. *PLoS Biol*, 15(12).
- Levenberg, K. (1944). A method for the solution of certain non-linear problems in least squares. *Quarterly of applied mathematics*, 2(2):164–168.
- Lewanzik, D. and Goerlitz, H. R. (2018). Continued source level reduction during attack in the low-amplitude bat *Barbastella barbastellus* prevents moth evasive flight. *Functional Ecology*, 32(5):1251–1261.
- Lin, Y. and Abaid, N. (2015). Modeling perspectives on echolocation strategies inspired by bats flying in groups. *J. Theor. Biol.*, 387:46–53.
- Lin, Y., Abaid, N., and Müller, R. (2016). Bats adjust their pulse emission rates with swarm size in the field. *J. Acoust. Soc. Am.*, 140:4318–4325.
- Lu, M., Zhang, G., and Luo, J. (2020). Echolocating bats exhibit differential amplitude compensation for noise interference at a sub-call level. *Journal of Experimental Biology*, 223(19).
- Luo, J., Goerlitz, H., Brumm, H., and Wiegreb, L. (2015). Linking the sender to the receiver: Vocal adjustments by bats to maintain signal detection in noise. *Sci. Rep.*, 5:18556.
- Luo, J. and Moss, C. (2017). Echolocating bats rely on audiovocal feedback to adapt sonar signal design. *Proc. Natl. Acad. Sci. U.S.A.*, 114:10978–10983.
- Marquardt, D. W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the society for Industrial and Applied Mathematics*, 11(2):431–441.
- Matsubara, J. and Heiligenberg, W. (1978). How well do electric fish electrolocate under jamming? *J. Comp. Physiol.*, 125:285–290.
- Matsuta, N., Hiryu, S., Fujioka, E., Yamada, Y., Riquimaroux, H., and Watanabe, Y. (2013). Adaptive beam-width control of echolocation sounds by cf-fm bats, *rhinolophus ferrumequinum nippon*, during prey-capture flight. *Journal of Experimental Biology*, 216(7):1210–1218.

- McCracken, G. F. (1984). Communal nursing in mexican free-tailed bat maternity colonies. *Science*, 223(4640):1090–1091.
- McFee, B., Kim, J. W., Cartwright, M., Salamon, J., Bittner, R. M., and Bello, J. P. (2018). Open-source practices for music signal processing research: Recommendations for transparent, sustainable, and reproducible audio research. *IEEE Signal Processing Magazine*, 36(1):128–137.
- McKinney, W. (2010). Data structures for statistical computing in python. *Proc 9th Python Sci Conf*.
- McKinney, W. et al. (2010). Data structures for statistical computing in python. In *Proceedings of the 9th Python in Science Conference*, volume 445, pages 51–56. Austin, TX.
- Metzner, W. and Müller, R. (2016). Ultrasound production, emission, and reception. In *Bat Bioacoustics*, pages 55–91. Springer.
- Miller, G. and Licklider, J. (1950). The intelligibility of interrupted speech. *J. Acoust*, 22:167–173.
- Møhl, B., Wahlberg, M., Madsen, P. T., Miller, L. A., and Surlykke, A. (2000). Sperm whale clicks: Directionality and source level revisited. *The journal of the Acoustical Society of America*, 107(1):638–648.
- Moussaïd, M., Helbing, D., and Theraulaz, G. (2011). How simple rules determine pedestrian behavior and crowd disasters. *Proceedings of the National Academy of Sciences*, 108(17):6884–6888.
- Mysuru Rajagopalachari, N., Beleyur, T., Krishna, A., and Goerlitz, H. R. (2020). title to be decided.
- Møhl, B. and Surlykke, A. (1989). Detection of sonar signals in the presence of pulses of masking noise by the echolocating bat, *eptesicus fuscus*. *J. Comp. Physiol. A*, 165:119–124.
- Nelson, M. and MacIver, M. (2006). Sensory acquisition in active sensing systems. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol*, 192:573–586.
- Obrist, M., Fenton, M., Eger, J., and Schlegel, P. (1993). What ears do for bats: a comparative study of pinna sound pressure transformation in chiroptera. *J. Exp. Biol*, 180,119–152.
- Oliphant, T. (2006a). *A guide to NumPy*. Trelgol Publishing USA.
- Oliphant, T. E. (2006b). *A guide to NumPy*, volume 1. Trelgol Publishing USA.
- Ortega, J., editor (2016). *Sociality in Bats*. Springer International Publishing Switzerland.

- Perkins, M., Frank, H., Pauly, J., and Hadly, E. (2017). Frequency shifting reduces but does not eliminate acoustic interference between echolocating bats: A theoretical analysis. *J. Acoust. Soc. Am.*, 142:2133–2142.
- Petrites, A., Eng, O., Mowlds, D., Simmons, J., and DeLong, C. (2009). Interpulse interval modulation by echolocating big brown bats (*epitesicus fuscus*) in different densities of obstacle clutter. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.*, 195:603–617.
- Pita, D., Collignon, B., Halloy, J., and Fernández-Juricic, E. (2016). Collective behaviour in vertebrates: a sensory perspective. *Royal Society open science*, 3(11):160377.
- Pollefeyns, M. and Nister, D. (2008). Direct computation of sound and microphone locations from time-difference-of-arrival data. In *Proc. of International Conference on Acoustics, Speech and Signal Processing*.
- Raguram, R., Chum, O., Pollefeyns, M., Matas, J., and Frahm, J.-M. (2012). Usac: a universal framework for random sample consensus. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):2022–2038.
- Reynolds, C. (1987). Flocks, herds and schools : A distributed behavioral model. *ACM SIGGRAPH Comput. Graph*, 21:25–34.
- Rhinehart, T. A., Chronister, L. M., Devlin, T., and Kitzes, J. (2020). Acoustic localization of terrestrial wildlife: Current practices and future opportunities. *Ecology and Evolution*, n/a(n/a).
- Richards, M., J., S., and Holm, W. (2010). *Principles of Modern Radar*. SciTech Publishing, Edison, New Jersey.
- Rohatgi, A. (2015). WebPlotDigitizer. Available at: [http://arohatgi.info/WebPlotDigitizer/](#)
- Schnitzler, H.-U. and Denzinger, A. (2011). Auditory fovea and doppler shift compensation: adaptations for flutter detection in echolocating bats using cf-fm signals. *Journal of Comparative Physiology A*, 197(5):541–559.
- Schoeppler, D., Schnitzler, H.-U., and Denzinger, A. (2018). Precise doppler shift compensation in the hipposiderid bat, *hipposideros armiger*. *Scientific reports*, 8(1):1–11.
- Seabold, S. and Perktold, J. (2010). Econometric and statistical modeling with python. In *Proc 9th Python Sci Conf*.
- Seibert, A., Koblitz, J., Denzinger, A., and Schnitzler, H. (2015). Bidirectional echolocation in the bat *barbastella barbastellus*: Different signals of low source level are emitted upward through the nose and downward through the mouth. *PLoS One*, 10:0135590.

- Siewert, I., Schillinger, T., and Schmidt, S. (2004). Forward masking and the consequence on echo perception in the gleaning bat, *megaderma lyra*. In Thomas, J., Moss, C., and Vater, M., editors, *Echolocation in Bats and Dolphins*, page 266–268. University of Chicago Press.
- Simmons, J. and Gaudette, J. (2012). Biosonar echo processing by frequency-modulated bats. *IET Radar Sonar and Navigation*, 6:556–565.
- Sonyia Jamal, F. (2017). Bulla ki jaana main kaun – bulla! i know not who i am. Online; accessed last 9-January-2021.
- Speaks, C. (1996). *Introduction to Sound: Acoustics for the Hearing and Speech Sciences*. Singular Publishing Group, Inc, San Diego, California.
- Stewénius, H. (2005). *Gröbner Basis Methods for Minimal Problems in Computer Vision*. PhD thesis, Lund University.
- Stidsholt, L., Johnson, M., Beedholm, K., Jakobsen, L., Kugler, K., Brinkløv, S., Salles, A., Moss, C. F., and Madsen, P. T. (2019). A 2.6-g sound and movement tag for studying the acoustic scene and kinematics of echolocating bats. *Methods in Ecology and Evolution*, 10(1):48–58.
- Stilz, P., Koblitz, J., and Goerlitz, H. R. (2019). Toadsuite acoustic tracking software.
- Strandburg-Peshkin, A., Twomey, C. R., Bode, N. W., Kao, A. B., Katz, Y., Ioannou, C. C., Rosenthal, S. B., Torney, C. J., Wu, H. S., Levin, S. A., et al. (2013). Visual sensory networks and effective information transfer in animal groups. *Current Biology*, 23(17):R709–R711.
- Sumpter, D. (2006). The principles of collective animal behaviour. *philos. Lond. B Biol. Sci*, 361:5–22.
- Surlykke, A. (1992). Target ranging and the role of time-frequency structure of synthetic echoes in big brown bats, *eptesicus fuscus*. *J. Comp. Physiol. A*, 170:83–92.
- Surlykke, A. (1993). Echolocation in two very small bats from thailand *craseonycteris thonglongyai* and *myotis siligorensis*. *Behav. Ecol. Sociobiol*, 33:1–12.
- Surlykke, A. (2012). Echolocating bats emit a highly directional sonar sound beam in the field. *Proceedings of the Royal Society B: Biological Sciences*, 276(1658):853–860.
- Surlykke, A., Nachtigall, P. E., Fay, R. R., and Popper, A. N. (2014). *Biosonar*, volume 51. Springer.
- Surlykke, A., Pedersen, S., and Jakobsen, L. (2012). Echolocating bats emit a highly directional sonar sound beam in the field. *Proc R Soc B Biol Sci*, 276(1658).

- Suzuki, R., Matsabayashi, S., Hedley, R. W., Nakadai, K., and Okuno, H. G. (2017). Harkbird: Exploring acoustic interactions in bird communities using a microphone array. *Journal of Robotics and Mechatronics*, 29(1):213–223.
- Sümer, S., Denzinger, A., and Schnitzler, H.-U. (2009). Spatial unmasking in the echolocating big brown bat, *epetesicus fuscus*. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.*, 195:463–472.
- Taschuk, M. and Wilson, G. (2016). Ten Simple Rules for Making Research Software More Robust. pages 1–10.
- Theriault, D., Wu, Z., Hristov, N., Swartz, S., Breuer, K., Kunz, T., and Betke, M. (2010). *Reconstruction and Analysis of 3D Trajectories of Brazilian Free-tailed Bats in Flight*. Bost Univ Comput Sci Technical Rep.
- Tian, B. and Schnitzler, H.-U. (1997). Echolocation signals of the greater horseshoe bat (*rhinolophus ferrumequinum*) in transfer flight and during landing. *The Journal of the Acoustical Society of America*, 101(4):2347–2364.
- Tressler, J. and Smotherman, M. (2009). Context-dependent effects of noise on echolocation pulse characteristics in free-tailed bats. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.*, 195:923–934.
- Tunstrøm, K., Katz, Y., Ioannou, C. C., Huepe, C., Lutz, M. J., and Couzin, I. D. (2013). Collective states, multistability and transitional behavior in schooling fish. *PLoS Comput Biol*, 9(2):e1002915.
- Turner, A. and Penn, A. (2002). Encoding natural movement as an agent-based system: an investigation into human pedestrian behaviour in the built environment. *Environment and planning B: Planning and Design*, 29(4):473–490.
- Ulanovsky, N., Fenton, M., Tsoar, A., and Korine, C. (2004). Dynamics of jamming avoidance in echolocating bats. *Proc. Biol. Sci.*, 271:1467–1475.
- Ulanovsky, N. and Moss, C. (2008). What the bat’s voice tells the bat’s brain. *Proc. Natl. Acad. Sci. U.S.A.*, 105:8491–8498.
- Van Rossum, G. (1991). Interactively testing remote servers using the python programming language. *CWI Q*, 4(December).
- Van Rossum, G. and Drake Jr, F. L. (1995). *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam.
- Vater, M., Kossl, M., Foeller, E., Coro, F., Mora, E., and Russell, I. (2003). Development of echolocation calls in the mustached bat, *pteronotus parnellii*. *Journal of Neurophysiology*, 90(4):2274–2290.
- Vicsek, T., Czirók, A., Ben-Jacob, E., Cohen, I., and Shochet, O. (1995). Novel type of phase transition in a system of self-driven particles. *Phys. Rev. Lett.*, 75:1226–1229.

- Vicsek, T. and Zafeiris, A. (2012). Collective motion. *Phys. Rep.*, 517:71–140.
- Virtanen, P. (2019). Scipy 1.0—fundamental algorithms for scientific computing in python. arXiv Prepr:1–22.
- Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Jarrod Millman, K., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and Contributors, S. . . (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272.
- Wahlberg, M. (1999). Positioning accuracy of a large-aperture hydrophone array for sperm whale research. *The Journal of the Acoustical Society of America*, 105(2):1318–1318.
- Wilson, G., Aruliah, D. A., Brown, C. T., Hong, N. P. C., Davis, M., Guy, R. T., Haddock, S. H. D., Huff, K., Mitchell, I. M., Plumley, M., Waugh, B., White, E. P., and Wilson, P. (2012). Best Practices for Scientific Computing. 12(1).
- Wisniewska, D. (2015). Range-dependent flexibility in the acoustic field of view of echolocating porpoises (*phocoena phocoena*). *eLife*, 2015(4).
- Xie, Y., Allaire, J., and Golemund, G. (2018). *R Markdown: The Definitive Guide*. Chapman and Hall/CRC, Boca Raton, Florida. ISBN 9781138359338.
- Yost, W. (2007). *Fundamentals of Hearing: An Introduction*. Academic Press, London.
- Yovel, Y., Melcon, M., Franz, M., Denzinger, A., and Schnitzler, H. (2009). The voice of bats :how greater mouse-eared bats recognize individuals based on their echolocation calls. *PLOS Comput. Biol.*, 54:1000400.