

Beleyur-Thesis

Thejasvi Beleyur

2020 – 11 – 03

Contents

1	Introduction	5
2	Modeling active sensing in groups of bats reveals echo detection even in large groups	7
	Abstract	8
2.1	Introduction	9
2.2	Materials and methods	10
2.3	Results	13
2.4	Discussion	17
2.5	Conclusion	20
2.6	Supplementary Information	22
3	Horseshoe bat craziness paper	39
4	ushichka	41
4.1	Example one	41
4.2	Example two	41
5	Robust Self-Calibration of Constant Offset Time-Difference-of-Arrival	43
	Abstract	44
5.1	THINGS TO DO	45
5.2	Introduction	45
5.3	Time-difference-of-arrival self calibration	45
5.4	Local optimization and the low rank relaxation	46

5.5	Minimal problems and solvers	47
5.6	Using RANSAC for five rows	48
5.7	Robust estimation of parameters	49
5.8	Experimental Validation	49
5.9	Conclusions	51
6	tacost: Testing and simulating the performance of acoustic tracking systems	53
6.1	Introduction	53
7	itsfm	55
7.1	Example one	55
7.2	Example two	55

Chapter 1

Introduction

This is the part where I say something broad and clever. And then proceed to make further connections into the topic of interest.

In chapter 1, I present miaow miaow miaow.

In Chapter 2, I present bow bow bow

In Chapter 3, I do kat kat kat kat

In Chapter 4, I do BOW BOW BOW BOW .

1. About the SFS methodology : Multi-microphone arrays have been an important and indispensable tool in the study of echolocation in the laboratory ever since the work of [Aubauer XXXX]. Working with microphone arrays in the field however comes with a series of challenges. Handling multi-microphone arrays in the field can be logistically and physically straining: the long cables involved must be brought to order each time, the arrays themselves can be bulky and must be assembled on site each time, and 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 of microphone positions. The use of TotalStation type surveying systems that provide direct XYZ measurements are not common in the field (they are themselves another piece of bulky equipment), and instead it is more common to measure inter-microphone distances with a tape measure or laser range finders. Having measured the inter-microphone distances in an array, the final XYZ positions of each microphone can be reconstructed. In my search for a method thatan academic ‘Christmas present’.

Figures and tables with captions will be placed in **figure** and **table** environments, respectively.

Chapter 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:

- 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, 2013, Griffin (1958)) 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 (2015), Gillam (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.

2.2.1 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

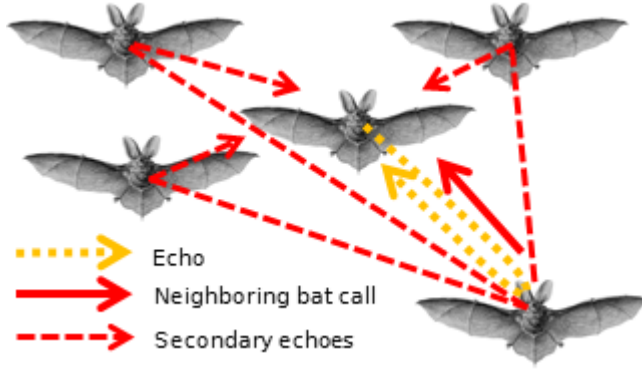


Figure 2.1: Schematic of the cocktail party nightmare. Arrows indicate the different types of sounds received by a focal bat: it needs to hear the echoes returning from its own calls (orange) to detect its neighbors, despite the masking by the calls of neighboring bats (solid red) and their secondary 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.

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 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 1) while keeping group size constant ($n=100$, 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).

2.2.2 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, sections 1.6 and 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, section 1.9). For target and secondary echoes, we also considered monostatic and bistatic target strengths measured in this paper (SI Appendix, section 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), @suemer2009a). 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 (2007a), 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, section 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).

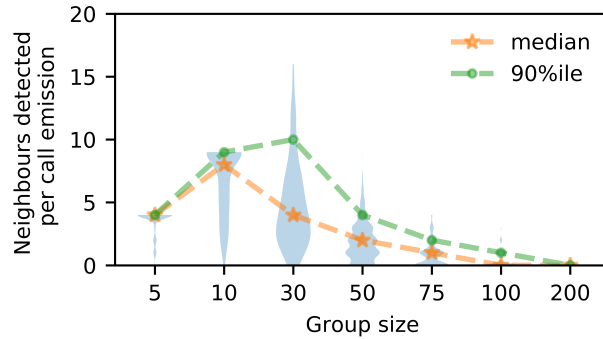


Figure 2.2: Number of detected neighbors per call by a focal bat in the center of a group. The initial rise in the number of detected neighbors is because there are indeed more neighbors and the degree of masking is low. However, with increasing group size, most of the neighbors cannot be detected anymore, 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).

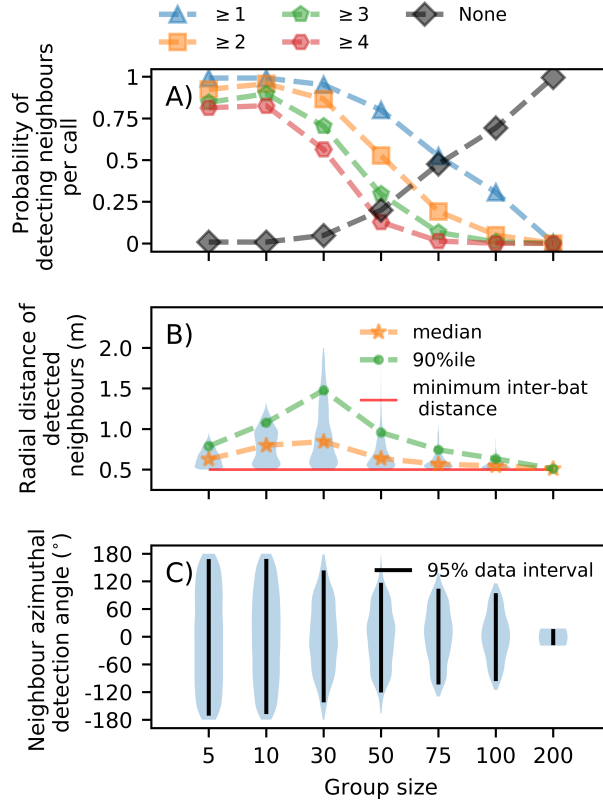


Figure 2.3: Characterization of the focal bat's perception. (A) The probability of detecting $\leq X$ neighbors per call ($X = 1, 2, 3, 4$, or none). Even in groups of up to 100 bats, the focal bat has a 0.3 probability of detecting at least 1 neighbor per call. In even larger groups (200 bats), no neighbors 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.

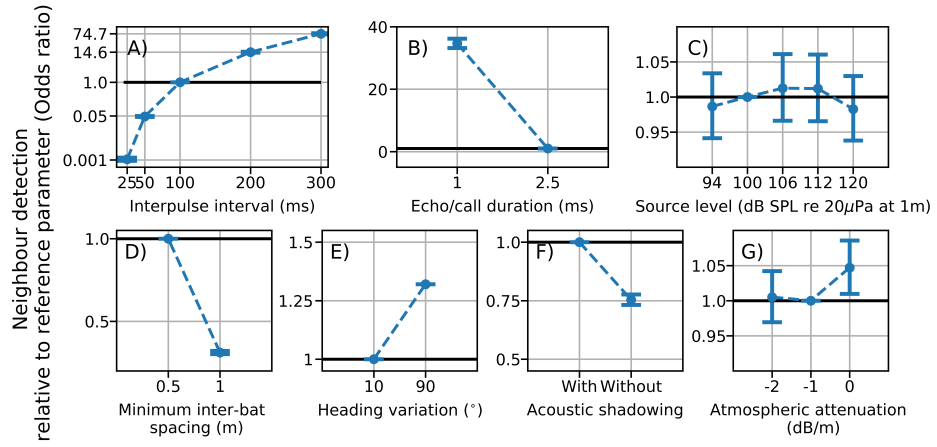


Figure 2.4: Effect of call parameters (A–C), group geometry (D and E), and acoustic parameters (F and G) on neighbor detection. Each plot shows the probability of neighbor detection (model estimate and 95% confidence interval of odds ratio) when changing model parameters relative to the reference parameter used in the simulations of Scenario 1 (Table 1). 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)

2.3.2 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 (90%ile, 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 1).

The call parameters interpulse interval and call duration showed the strongest effect (Figure 2.4A and B and SI Appendix, TableS2). 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 to 0.05 times (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 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 (griffin1958a). 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), G. Grégoire (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. (2015), 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 (42). 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 (11, 23, 43, 44). 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 for 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 flight. Interindividual 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 interbat 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 (9). 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 (45)

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 F (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 (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 (Haft 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. (2015), 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 S1). 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 the target object's 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 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 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 $\pm 180^\circ$. 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 ($0^\circ \leq \theta \leq 180^\circ$).

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/COMPA, Avisoft Bioacoustics, Glienicke, 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 μ 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 - 360° . 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 \times 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/COMPA, 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 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, 2007b) 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 μ 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 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, 2006), SciPy (Virtanen, 2019), Pandas (McKinney, 2010), Matplotlib (Hunter, 2007), 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 Schematic 1

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:
 - (a) 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.
 - (b) 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:
 - (a) Implement spatial unmasking by reducing the effective received level of all masking sounds based on their angular separation to the echo
 - (b) Combine all maskers over time to form a 'masker profile'
 - (c) Calculate the 'echo-to-masker profile', with reference to the echo level
 - (d) 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.

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

Table 2.1: 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).

Parameter	Value tested	Reference value	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			0.31	0.35	-1.11	0.027	-40.51	0.0
Intercept			0.31	0.35	-1.11	0.027	-40.51	0.0
Intercept			0.31	0.35	-1.11	0.027	-40.51	0.0
Intercept			0.31	0.35	-1.11	0.027	-40.51	0.0

Table 2.2: TO BE FIXXXXXXXXXXXXX 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.

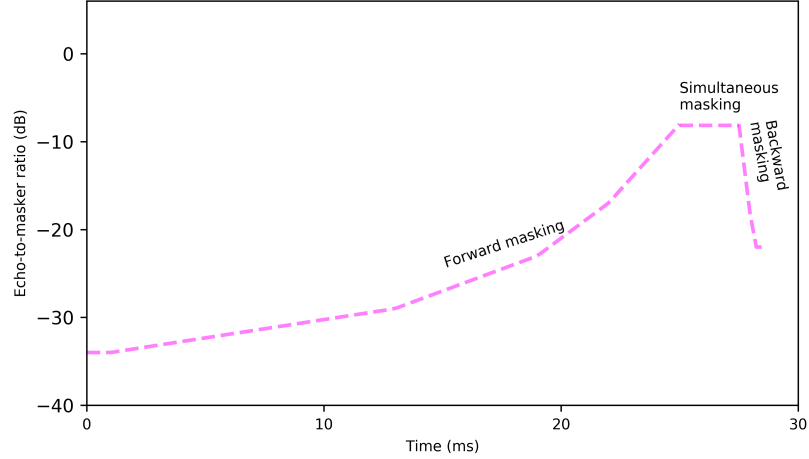


Figure 2.5: The 'temporal masking envelope' used to simulate temporal masking. The envelope represents the lower echo-to-masker ratios at which a bat can detect echoes for various echo-masker delays. The envelope is the equivalent of the lowest signal-to-noise ratios at which echoes can be detected over 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 S1.

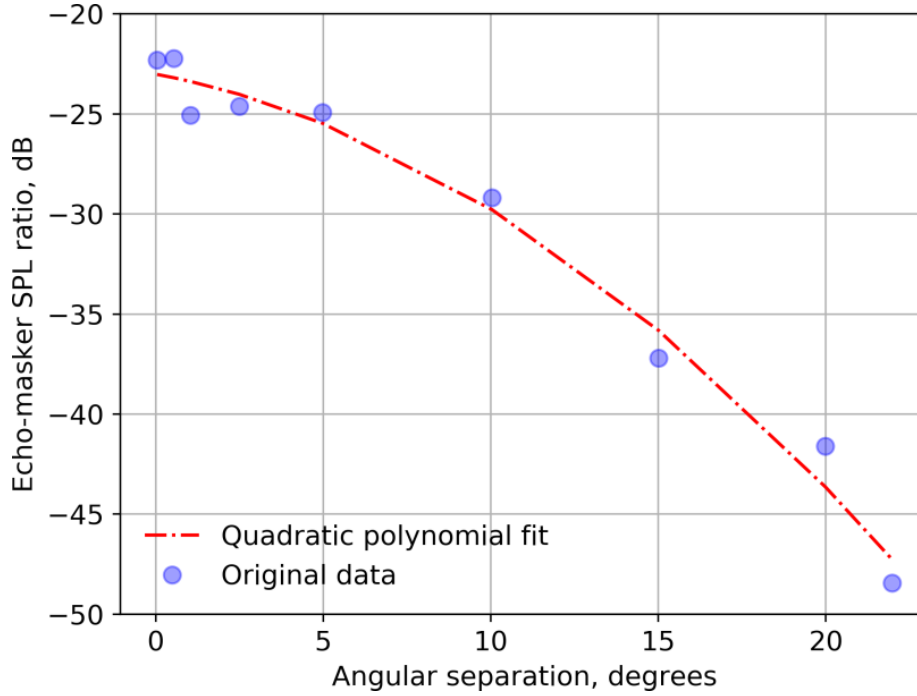


Figure 2.6: The ‘spatial unmasking function’ describes the reduction in echo-to-masker SPL ratio at echo detection as a function of angular separation between echo and masker. A) The original data set of Sümer et al. (2009) (blue dot) and our digitized and interpolated dataset (red line). The error between 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.

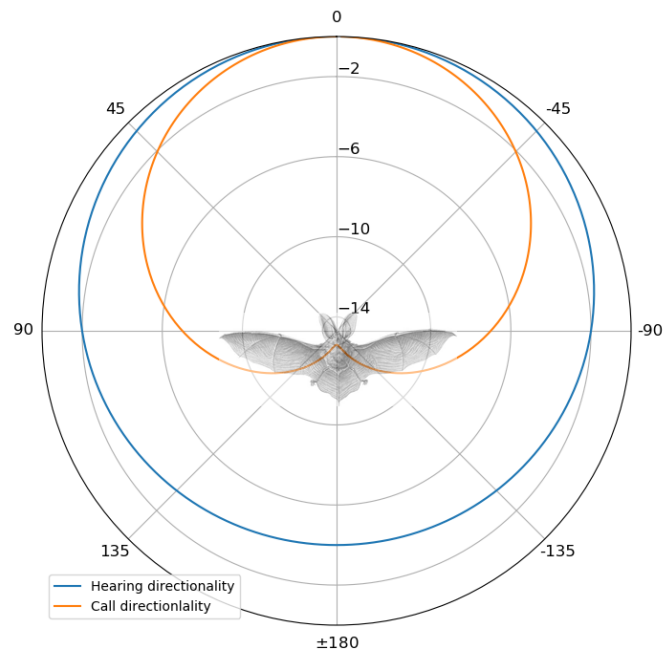


Figure 2.7: Calling and hearing directionality of bats. Call directionality (orange) is directional, with a difference of up to -14 dB in source level from front to back. Calls emitted to the front of a bat result in higher received levels of calls, echoes and secondary echoes. Hearing directionality (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).

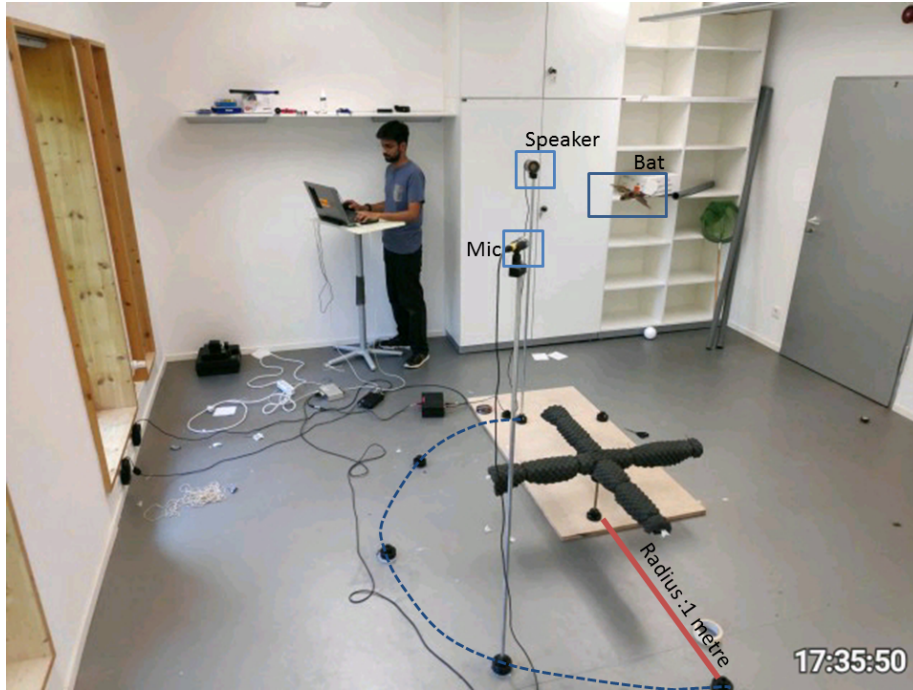


Figure 2.8: The ensouffication setup used to measure the monostatic and bistatic target strength of a bat as an acoustic target. A stuffed *Myotis myotis* bat was hung at the same height as the speaker and microphone. The bat could be rotated in the azimuth. The microphone and speaker were placed at 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.

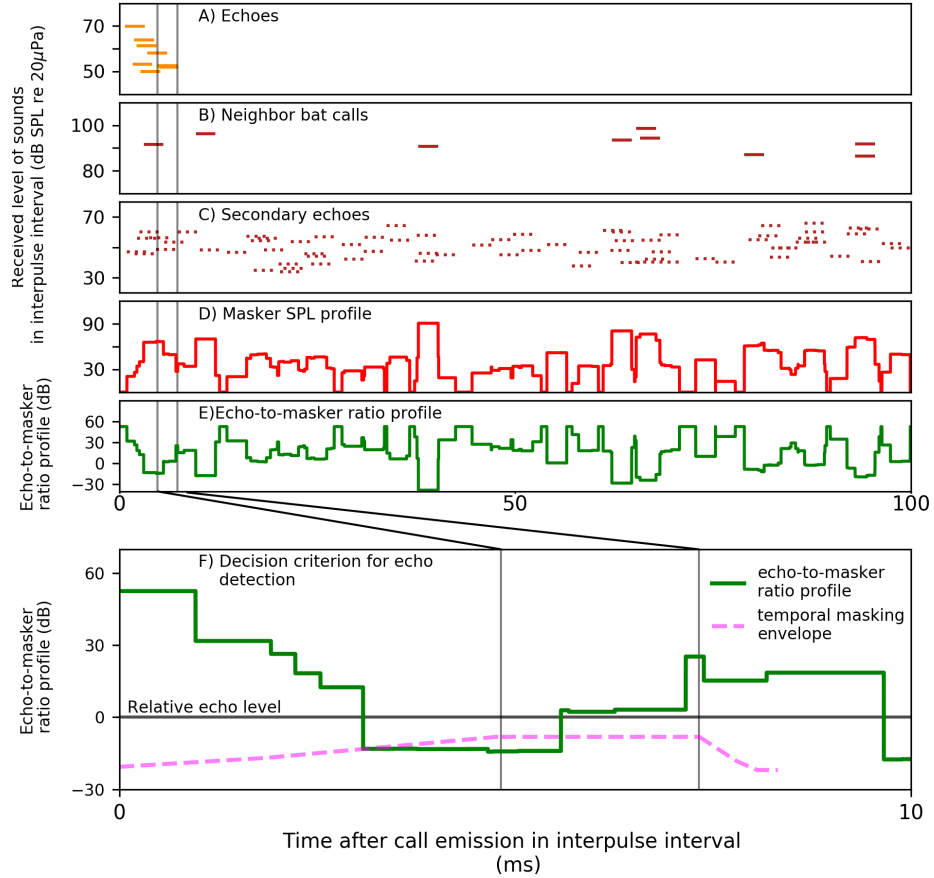


Figure 2.9: Schematic representation of the sounds arriving in the interpulse interval for a simulation with a group of 10 bats (A-C) and how echo detection was determined (D-F). A-E show the full interpulse interval of 100 ms, while F shows an enlargement of the first 10 ms. A) Timing and received SPL 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.

Chapter 3

Horseshoe bat craziness paper

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

Chapter 4

ushichka

Some *significant* applications are demonstrated in this chapter.

4.1 Example one

4.2 Example two

Chapter 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 μs) 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 THINGS TO DO

1. FIX THE FIGURE SIZES – THEY'RE REALLY SMALL RIGHT NOW....

5.2 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. (2013); 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.

5.3 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.

5.4 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

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

5.5 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

Table 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 $\theta_1 = \{R, S, o\}$	$600 ms$	$22 ms$

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.*

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

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

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

$$U = \begin{pmatrix} 0 & u_2 & u_3 & u_4 & u_5 \end{pmatrix}, \quad (5.14)$$

$$V = \begin{pmatrix} 0 & v_2 & v_3 & v_4 & v_5 \end{pmatrix}, \quad (5.15)$$

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 $C^T(Z - o)^2 C$.

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.

5.6 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).
-

%There are two key questions here. First, how many inliers do we expect for the null hypothesis, *i.e.* if all data is random noise. Secondly, how many iterations do we typically need to find all the inliers, when the data is

5.7 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.

5.8 Experimental Validation

5.8.1 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.

5.8.2 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.

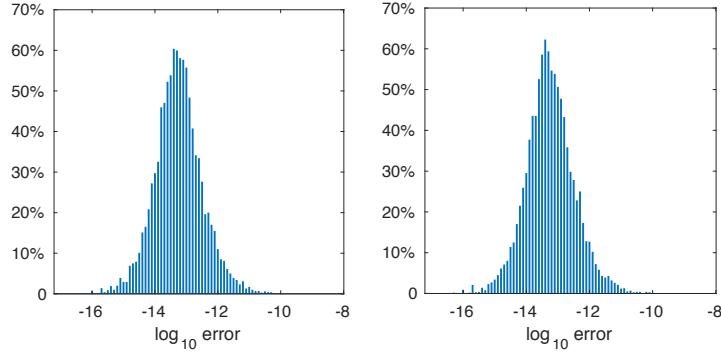


Figure 5.1: Left shows the histogram of the logarithm of the absolute errors, for the Matlab implementation of our minimal solver. To the right the corresponding histogram for the C++ implementation.

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 of 96 kHz 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 kHz . 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 , $16 - 96 \text{ kHz}$ 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.

5.8.3 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.

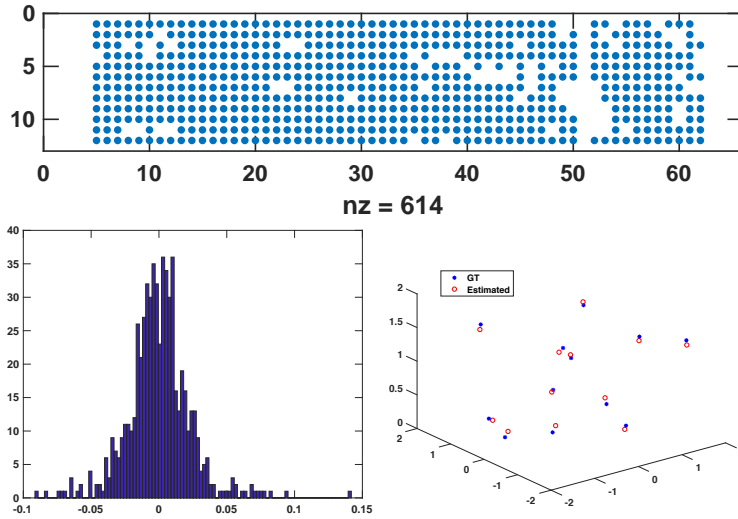


Figure 5.2: For the office experiment the figure shows detected inliers W_{in} (top), inlier residual histogram (bottom left), and estimated and ground truth microphone positions (bottom right).

5.9 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

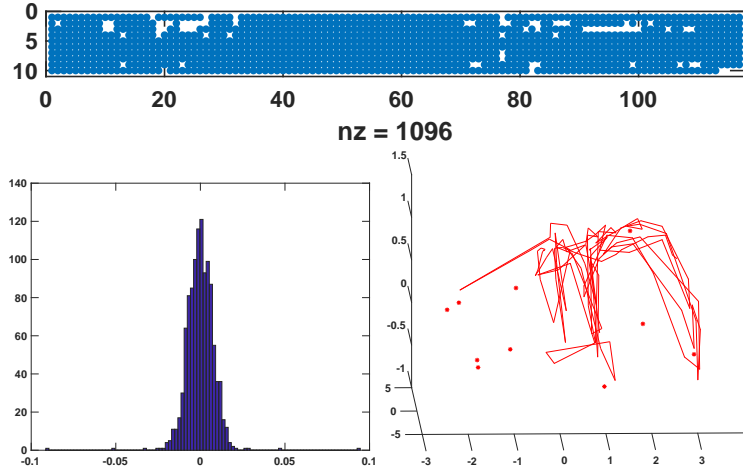


Figure 5.3: For the cave experiment the figure shows detected inliers W_{in} (top), inlier residual histogram (bottom left) and estimated microphone and sound source positions, red dots and line respectively (bottom right).

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 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 ± 0.04 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.

Chapter 6

tacost: Testing and simulating the performance of acoustic tracking systems

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.

Chapter 7

itsfm

Some *significant* applications are demonstrated in this chapter.

7.1 Example one

7.2 Example two

Bibliography

- 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 Y. (2015). Yovel calling louder and longer: How bats use biosonar under severe acoustic interference from other bats. *Proc. R Soc. B Biol. Sci.*, 282:20152064.
- 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. (1996). *Acoustical flight path tracking of echolocating bats*. PhD thesis, Acoustical Society of America.
- 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., 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 Michéyl, 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.
- Blumstein, D. T., Mennill, D. J., Clemens, 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.

- 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.
- 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.
- Bullock, T., Hopkins, C., Popper, A., and Fay, R., editors (2005). *Electroreception*. Springer Science+Business Media, Inc, New York.
- 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 The Auditory System and the Cocktail Party*. ASA Press.
- Cvikel, N. (2015). Bats aggregate to improve prey search but might be impaired when their density becomes too high. *Curr. Biol*, 25:206–211.
- Cvikel, N., Levin, E., Hurme, E., Borissov, I., Boonman, A., Amichai, E., and Yovel, Y. (2015). On-board recordings reveal no jamming avoidance in wild bats. *Proceedings of the Royal Society B: Biological Sciences*, 282(1798):20142274.
- 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 *phyllostomus 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.
- Fay, R. (2008). *Sound source perception and stream segregation in nonhuman vertebrate animals” in Auditory Perception of Sound Sources*. Springer Science+Business, Media, LLC.
- Fenton, M. (2013). Questions, ideas and tools: Lessons from bat echolocation. *Anim. Behav*, 85:869–879.
- Firzlaff, U. and Schuller, G. (2003). Spectral directionality of the external ear of the lesser spear-nosed bat, *phyllostomus 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.
- G. Grégoire, H. C. (2004). Onset of collective and cohesive motion. *Phys. Rev. Lett*, 92:025702.
- Geier, M. (2015a). Available at:.
- Geier, M. (2015b). sounddevice 0.3.5. https://python-sounddevice.readthedocs.io/en/0.3.5/_modules/sounddevice.html.
- Ghose, K. and F, C. (2006). Moss, 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.
- 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.
- Gillam, E. (2010). Echolocation behavior of brazilian free-tailed bats during dense emergence fights. *J. Mammal*, 91:967–975.
- Gillam, E. and Montero, B. (2016). Influence of call structure on the jamming avoidance response of echolocating bats. *J. Mammal*, 97:14–22.
- 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).
- Griffin, D. (1958). *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.
- 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.
- Haftner, E., Sarampalis, A., and Loui, P. (2008). *Auditory attention and filters” in Auditory Perception of Sound Sources*. Springer Science+Business, Media, LLC.
- Hase, K. (2018). Bats enhance their call identities to solve the cocktail party problem. *Commun Biol*, 1:39.
- 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.
- 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).
- Hunter, J. (2007). Matplotlib: A 2d graphics environment. *Comput Sci Eng*, 9(3).
- 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.
- 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.

- Korman, S. and Litman, R. (2018). Latent ransac. 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). 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.
- 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.
- Luo, J., Goerlitz, H., Brumm, H., and Wiegerebe, 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.
- McKinney, W. (2010). Data structures for statistical computing in python. *Proc 9th Python Sci Conf*.

- 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.
- Møhl, B. and Surlykke, A. (1989). Detection of sonar signals in the presence of pulses of masking noise by the echolocating bat, *Myotisotis 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. (2006). *A guide to NumPy*. Trelgol Publishing USA.
- Ortega, J., editor (2016). 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). Inter-pulse interval modulation by echolocating big brown bats (*Myotisotis fuscus*) in different densities of obstacle clutter. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol*, 195:603–617.
- Pollefeys, 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., Pollefeys, M., Matas, J., and Frahm, J.-M. (2013). Usac: a universal framework for random sample consensus. 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:.
- 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*. 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.
- 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.
- 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., Nachtigall, P., and R. (2014). *R*. Springer-Verlag, New York.
- 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., Matsubayashi, 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, *eptesicus fuscus*. *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol*, 195:463–472.
- 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.

- 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.
- 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.
- 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.
- 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.
- Wisniewska, D. (2015). Range-dependent flexibility in the acoustic field of view of echolocating porpoises (phocoena phocoena. *eLife*, 2015(4).
- Yost, W. (2007a). *Fundamentals of Hearing: An Introduction*. Academic Press, London.
- Yost, W. (2007b). *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.