Three-Dimensional Localization of Bats: Visual and Acoustical

Klaus Hochradel, Timm Häcker, Tino Hohler, Andreas Becher, Stefan Wildermann, and Alexander Sutor

***Abstract*— We introduce a multi-sensor array consisting of multiple microphones and a stereo-vision system that enables to track three-dimensional (3D) flight paths of bats. Field applicability is a major design factor which led to a portable battery powered system. Eight microphones form the acoustic array that allows the detection and localization of ultrasonic echolocation calls. Two near infrared sensitive cameras provide a visual identification of the three dimensional flight path. We minimized costs of this multi-sensor arrays by developing the system from ground up, including a novel acoustic signal recording and processing hardware. Our acoustic data acquisition system is the first to sample eight microphones simultaneously at 1 MS/s/CH with 16 Bit resolution without the necessity of computers or laptops. All channels are processed in real time in the frequency and time domain to evaluate the occurrence of echolocation calls. Hereby, long term deployments of the system are achieved since only relevant signals are recorded. Parallel to the acoustic acquisition, the system records images of two near-infrared (IR) sensitive cameras at 12 Hz frame rate each. Combined with IR illumination 3D flight paths can be extracted in post-processing. In addition, the system is able to log several environmental parameters, such as temperature and humidity. Recorded datasets can be associated with a global position and time-stamp through a GPS module and transformed to GPS coordinates with a three-axis accelerometer to determine the angle. The overall system (microphones, cameras, and sensors) was designed to be affordable, user-friendly, battery powered and without the necessity of an additional laptop.**

***Index Terms*— Microphone, array, stereo camera, infrared, automation, ultrasound, source localization, bat.**

I. Introduction

HEARING and vocalization are fundamental aspects of communication for most animals as well as humans. In general, hearing is not only part of communication but also vital for survival which is why many mammals developed outstanding sensorial abilities. Multiple nocturnal birds evolved keen hearing senses to the precision where they are able to localize prey based on their emitted acoustic signals. In terrestrial mammals, bats perfected their acoustic sensory system to navigate and hunt in the absence of daylight. Despite the fascinating abilities of bats, they are also the second largest order of mammals and poorly studied compared to other groups. Modern technology facilitated by means of microphones and infrared- and thermal cameras the observation of this nocturnal mammal. Specialized equipment enables studies on foraging and echolocation behavior, habitat usage and many more. Increase in knowledge is important to understand and reduce wildlife conflicts, for example the bat mortality at wind turbines [4].

Studying the biosonar sense [23] demands different fields of research to collaborate. Neurophysiology, sensory ecology, acoustics and engineering together will hopefully decipher this sense. Acoustic engineering enables by the use of ultrasonic microphones the investigation of their echolocation calls. Multiple microphones at a priori known positions allow to determine the location in space of each call by means of their time differences of arrival at the microphones. These arrays of microphones are necessary to understand the sonar systems of bats since they enable not only to study the acoustic characteristics of an echolocation call [19] but also to correlate the call characteristic with surrounding parameters [15]. So far, microphone arrays have been used in specific echolocation studies [15], [17], [25] (overview over several arrays can be found [21] but on the other hand have been neglected in other studies (e.g. behavior related studies [4], [11]) due to their complexity and the lack of commercial availability. Thus, only researchers with access to or knowledge of engineering are able to utilize the benefits of localization systems [18], [21]. Consequently, all applied arrays are part of research projects and not commercially available. In addition to the lack of availability, most of the existing arrays share one further disadvantage, they are either only designed for laboratory measurements [9] or, in the field, an expert is needed [8], [12], [13]. This also limits the possible amount of recordings and localizations due to the effort to install and operate.

Core of this research is not to present another research array but to focus on a low-cost and easy to use localization system for a variety of researches. To achieve this goal, we developed new hardware and software from scratch. The following article will give a brief overview of the developments starting from the novel microphones and the high performance recording and processing platform. Only having information on the location where each echolocation call was emitted might not be sufficient enough for specific studies. To overcome this limitation we equipped our microphone array with two infrared sensitive cameras which enable visual localization. This offers

Manuscript received February 4, 2019; revised March 11, 2019; accepted March 11, 2019. Date of publication April 2, 2019; date of current version June 19, 2019. The associate editor coordinating the review of this paper and approving it for publication was Prof. Kazuaki Sawada. *(Corresponding author: Klaus Hochradel.)*

K. Hochradel and A. Sutor are with the Institute of Measurement and Sensor Technology, UMIT-Private University for Health Sciences, Medical Informatics and Technology GmbH, 6060 Hall in Tirol, Austria (e-mail: klaus.hochradel@umit.at).

T. Häcker is with the Chair of Sensor Technology, University of Erlangen-Nuremberg, 91054 Erlangen, Germany.

T. Hohler, A. Becher, and S. Wildermann are with the Department of Computer Science 12, University of Erlangen-Nuremberg, 91054 Erlangen, Germany.

Digital Object Identifier 10.1109/JSEN.2019.2907399

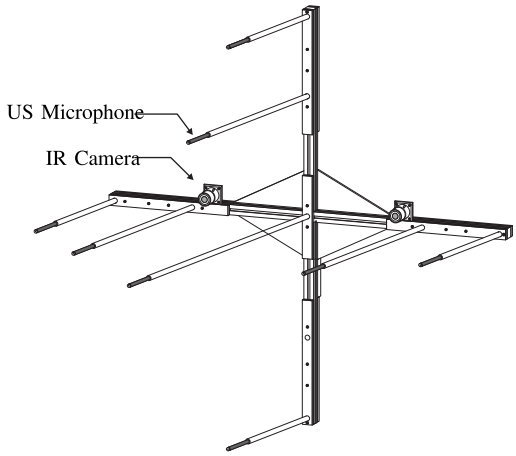

Fig. 1. Design of the microphone array based on the silicon microphone SPU0410 by *Knowles* and the stereo infrared camera system.

a whole new set of research possibilities. For example we would be able to study social behavior in big roosts where we can record the behavior of the bats with the cameras and at the same time make audio recordings where we can identify the recordings of individual bats. Also studies on flight behavior can be widened as we can in post processing visually determine the distance to obstacles or structures.

II. Setup

This chapter introduces a multi-sensor array consisting of eight ultrasonic microphones and two IR sensitive cameras (Fig. 1). To achieve a portable and modular system we designed a cross-shaped array with the microphones protruding out of plane and two cameras in-plane. This design allows to disassemble the array into a small size package of 500 x 200 x 200 mm (length, width, height). The microphones are designed with respect to minimizing echoes which is why they are out of the cross-plane. All parts are manufactured with precise fitting to reduce possible errors during assembly and to receive reproducibility in the accuracy of the microphone positions.

A. Microphones

Microphones commonly used for bioacoustic ultrasonic applications are Knowles FG type capsules. They are robust, omnidirectional up to 80 kHz and offer suitable high ultrasonic sensitivity but commercially available versions feature relatively high costs of approximately 400 EUR [1]. The FG type microphone capsule itself is with 35 EUR comparatively inexpensive but still requires an amplifier and its assembly is complicated. In addition, commercially available housings are of large diameter influencing signal quality negatively due to echoes, reflections and diffraction. Larger diameter microphones based on EMFi [24] or electrostatic foils [20] are highly sensitive but have a strongly focusing directivity pattern. Microphone arrays ideally demand omnidirectional microphones since focused directivity leads to a minor field of view. Furthermore, directional microphones result in varying signals at each microphones which leads to signal processing issues (e.g. incorrect results of cross correlation).

To overcome these negative effects and to reduce costs we designed new ultrasonic microphones based on the silicon microphone SPU0410 (costs < 1 EUR) introduced by Knowles in 2014. Our microphones feature a minor echoic footprint due to a small diameter of 8 mm. Minimized microphone dimensions reduce the echoic surface in the setup. Echoes lead to different signals on each microphone which cause issues in signal processing. Our small housing not only provides space for the MEMS microphone but also includes the pre-amplifier. To ease the use of the microphones with other hardware we equipped the microphone with a standard 3.5 mm audio jack which makes this sensor also compatible with the audio input of standard sound cards on PCs and Latops.

B. Acoustic Signal Processing Hardware

Hardware for digitizing and processing analog microphone signals produces the majority of costs for microphone arrays. Currently used by researchers are for example multichannel USGs (UltraSoundGate) by Avisoft Bioacoustics [1] or general purpose data acquisition (DAQ) cards such as the NI-6356 by National Instruments. DAQ cards have the advantage of lower cost ($\sim$ 4000 EUR) but require large amount of programming and electronic engineering. Multichannel USGs are designed to record echolocation calls of bats and thus easy to use. But they entail costs of up to 20000 EUR which restricts their usage to large fund projects. Additionally, commercially available platforms only cover the recording hardware which leaves the mechanical setup to still be designed. All current solutions for microphone arrays require laptops or PCs for controlling the DAQ card and saving recorded data.

We present a novel acoustic recording device for up to eight microphones which combines the ease of use advantage of USGs with very low costs of only 500 EUR. To achieve this goal, we had to choose a different approach as commercially available products. Our design criteria were

- ease of use
- usage without laptop
- automatic recording of echolocation calls
- up to date access with any mobile devices
- configurable parameters
- battery powered
- synchronized sampling of eight microphones at 1 MS/s per channel at 16 Bit.

High speed synchronous recordings at 8 x 1 MS/s and 16 Bit without laptop, while still being configurable and deployable in the field demand for new approaches in hardware as well as software development. To reach the necessary bandwidth and synchronicity an FPGA (field programmable Gate Array) is the best practice. Disadvantages of FPGAs are limitations in reconfiguration and communication during run-time. Reconfiguring parameters in a hardware programmed FPGA is in general not a straight forward task where in the worst case the complete FPGA needs to be reinitialized. Also FPGAs

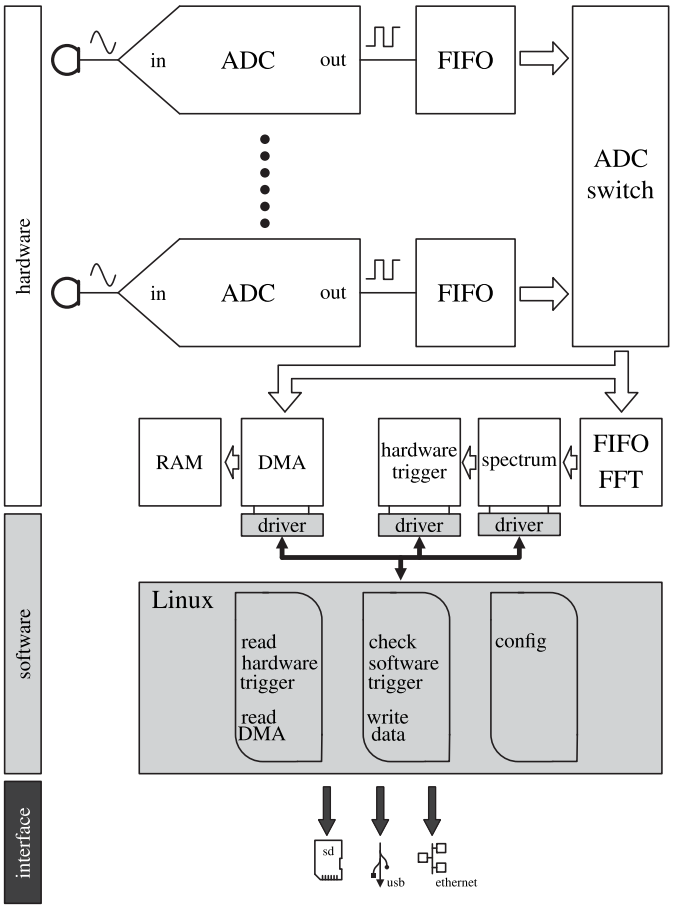

Fig. 2. Overview of the setup with the ADC board and the task splitting on the Zynq-7000 All Programmable SoC.

are limited in interface possibilities as high level interfaces like Ethernet are not natively implemented. The Zynq®-7000 All Programmable SoCs (System on a Chip) combine Programmable Logic (PL) with a dual-core ARM® Cortex™-A9 MPCore Processing System (PS), fast DDR3 Memory and high-end interfaces like Gigabit Ethernet. Utilizing this SoC enabled us to split task between PL and PS and use the advantages of both systems. Time-critical and real-time functions are implemented directly on hardware (PL) whereas the PS executes all non time-critical and dynamic tasks (see Fig. 2). Final components to complete the acoustic recording setup is the ADC board which incorporates eight amplifiers, eight ADCs of the type AD7980 by *Analog Devices* and the power supply for the eight microphones.

C. Stereo Infrared Cameras

To visually detect and locate bats with respect to low costs a new approach is introduced based on the highly promoted and supported single board computer (SBC) Raspberry Pi. Adding two infrared sensitive cameras, sensors and battery power supply transforms the SBC to a measurement unit for the field. The camera boards are based on 5MP 1/4" CMOS sensor OmniVision OV5647 and offer a CS mount for various lenses. The manual zoom lens T3Z2910CS-IR completed our low-cost infrared camera system for night recordings.

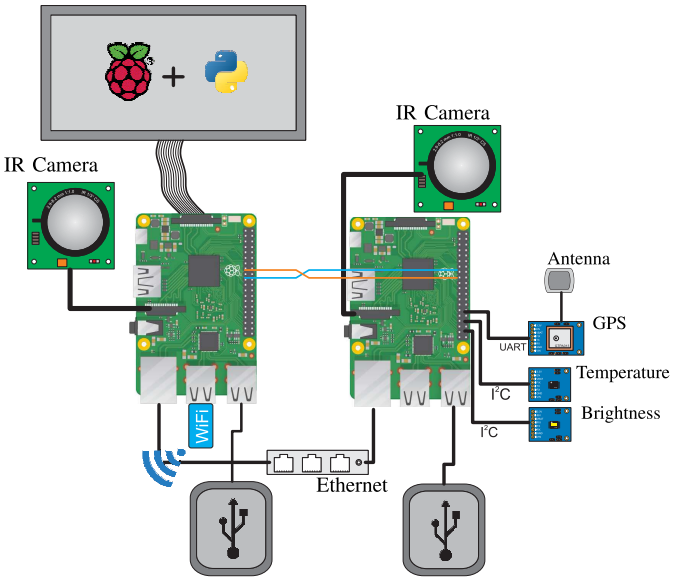

Fig. 3. Visual night recording setup based on two Raspberry Pis and several sensors.

Fig. 3 illustrates the final stereo-infrared system to record flight paths at night. It is equipped with a GPS module to supply the vision and acoustic system with the current time and location. Adjusting the system in the field is achieved with a touch sensitive display since camera parameters have to be adapted to the current situation.

D. Algorithms

1) Acoustic Localization: We implemented two different approaches to localize an acoustic source in space. We compare the straightforward multilateration based on the time difference of arrival (TDOA) with a novel approach which uses the Smith-Abel estimate [22] to determine the position $\mathbf{x}_s$ in space.

Multilateration: is based on the spherical relation

$$c\tau_{1,2} = c\tau_1 - c\tau_2 = R_1 - R_2 = r_{1,2} \tag{1}$$

with the speed of sound c, the time difference of arrival $\tau_{1,2}$ of the time of flights τ_1 and τ_2. The distance R_m between the source $\mathbf{x}_s = (x_s, y_s, z_s)$ and the m-th microphone (x_m, y_m, z_m) can also be expressed with the spherical relation

$$R_m = \sqrt{(x_s - x_m)^2 + (y_s - y_m)^2 + (z_s - z_m)^2}\,. \tag{2}$$

Combined with the origin microphone $x_1 = 0$, $y_1 = 0$, $z_1 = 0$ we end up with

$$\begin{pmatrix} x_2 & y_2 & z_2 & c\tau_{1,2} \\ x_3 & y_3 & z_3 & c\tau_{1,3} \\ x_4 & y_4 & z_4 & c\tau_{1,4} \\ x_5 & y_5 & z_5 & c\tau_{1,5} \\ x_6 & y_6 & z_6 & c\tau_{1,6} \\ x_7 & y_7 & z_7 & c\tau_{1,7} \\ x_8 & y_8 & z_8 & c\tau_{1,8} \end{pmatrix} \begin{pmatrix} x_s \\ y_s \\ z_s \\ R_1 \end{pmatrix} = \frac{1}{2} \begin{pmatrix} x_2^2+y_2^2+z_2^2 - c^2\tau_{1,2}^2 \\ x_3^2+y_3^2+z_3^2 - c^2\tau_{1,3}^2 \\ x_4^2+y_4^2+z_4^2 - c^2\tau_{1,4}^2 \\ x_5^2+y_5^2+z_5^2 - c^2\tau_{1,5}^2 \\ x_6^2+y_6^2+z_6^2 - c^2\tau_{1,6}^2 \\ x_7^2+y_7^2+z_7^2 - c^2\tau_{1,7}^2 \\ x_8^2+y_8^2+z_8^2 - c^2\tau_{1,8}^2 \end{pmatrix}. \tag{3}$$

This can be rewritten as

$$\mathbf{M}\mathbf{x}_s - \mathbf{b} = 0 \tag{4}$$

and solved with standard algorithms for linear equations. Matrices are denoted upper case letters whereas vectors are represented through lower case letters.

Smith-Abel estimation: Acoustic localization by multilateration demands accurate knowledge of time differences and speed of sound. The method introduced by Smith and Abel [22] is robust towards errors in TDOAs or speed of sound. An acoustic source can be determined by the formal least-squares solution [22]

$$\mathbf{x}_s = \frac{1}{2}\mathbf{S}_W^*(\delta - 2\mathbf{R}_s\mathbf{d}) \tag{5}$$

where

$$\mathbf{S}_W^* = \left(\mathbf{S}^T\mathbf{W}\mathbf{S}\right)^{-1}\mathbf{S}^T\mathbf{W} \tag{6}$$

minimizes the equation error with weighting $\mathbf{W}$. Using the spherical interpolation method $\mathbf{R}_s$ can be solved by

$$\mathbf{R_s} = \frac{\mathbf{d}^T\mathbf{P}_S^{\perp}\mathbf{W}\mathbf{P}_S^{\perp}\delta}{2\mathbf{d}^T\mathbf{P}_S^{\perp}\mathbf{V}\mathbf{P}_S^{\perp}\mathbf{d}} \tag{7}$$

$$\tag{8}$$

with

$$\mathbf{P}_S \mathrel{\widehat{=}} \mathbf{S}\mathbf{S}_W^* = \mathbf{S}(\mathbf{S}^T\mathbf{W}\mathbf{S})^{-1}\mathbf{S}^T\mathbf{W} \tag{9}$$

$$\mathbf{P}_S^{\perp} \mathrel{\widehat{=}} \mathbf{I} - \mathbf{P}_S \tag{10}$$

and

$$\mathbf{d} = \begin{bmatrix} R_2 - R_1 & R_3 - R_1 & \cdots & R_m - R_1 \end{bmatrix}^T \tag{11}$$

$$\mathbf{S} = \begin{bmatrix} x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \\ \vdots & \vdots & \vdots \\ x_m & y_m & z_m \end{bmatrix},\ \delta = \begin{bmatrix} R_2^2 - \mathbf{d}_1^2 \\ R_3^2 - \mathbf{d}_2^2 \\ \vdots \\ R_m^2 - \mathbf{d}_{m-1}^2 \end{bmatrix} \tag{12}$$

Time difference of arrival estimation: Regardless of the applied localization algorithm the estimation of the time differences of arrival determines the accuracy of the results. Thus, strong focus on the time differences is necessary to obtain accurate localizations. Studies revealed multiple algorithms to determine time differences of signals. The applicability of known algorithms like the straightforward generalized cross-correlation (GCC) or the average square difference function (ASDF) strongly depends on the type of signal, signal quality and type of disturbances (echoes or noise). To minimize echoic overlays the array is coated in absorbing foam (Basotect®covering aluminum beams) and it has to be placed in distance to reflecting surfaces. Signal quality can only be influenced to a certain degree since it mostly depends on the echolocation call. Nevertheless, we maximized the sensitivity of the microphones in the ultrasonic range and at the same time filtered frequencies lower than 16 kHz to achieve a good signal to noise ratio (SNR). The time difference of arrival can be determined with the generalized correlation method

$$\tau_{i,j} = \arg\max_{\tau}\left\{\mathrm{R}_{h_i,h_j}^{\mathrm{GCC}}(\tau)\right\} \tag{13}$$

TABLE I
TIME DIFFERENCES OF ARRIVAL MATRIX

	τ_1	τ_2	τ_3	τ_4	τ_m
τ_1	0	$\tau_{1,2}$	$\tau_{1,3}$	$\tau_{1,4}$	$\tau_{1,m}$
τ_2	-$\tau_{1,2}$	0	$\tau_{1,2}$-$\tau_{1,3}$	$\tau_{1,2}$-$\tau_{1,4}$	$\tau_{1,2}$-$\tau_{1,m}$
τ_3	-$\tau_{1,3}$	$\tau_{1,3}$-$\tau_{1,2}$	0	$\tau_{1,3}$-$\tau_{1,4}$	$\tau_{1,3}$-$\tau_{1,m}$
τ_4	-$\tau_{1,4}$	$\tau_{1,4}$-$\tau_{1,2}$	$\tau_{1,4}$-$\tau_{1,3}$	0	$\tau_{1,4}$-$\tau_{1,m}$
τ_m	-$\tau_{1,m}$	$\tau_{1,m}$-$\tau_{1,2}$	$\tau_{1,m}$-$\tau_{1,3}$	$\tau_{1,m}$-$\tau_{1,4}$	0

with

$$\mathrm{R}_{h_1,h_2}^{\mathrm{GCC}} = \mathcal{F}^{-1}\left\{\mathcal{W}(\omega)\mathrm{G}_{h_1,h_2}(\omega)\right\} \tag{14}$$

where $\mathrm{G}_{h_1,h_2}(\omega)$ denotes the onesided cross spectral density, $\mathcal{W}(\omega)$ an optional weighting function and $\mathcal{F}^{-1}$ the inverse fourier transform. Determining time differences with the average squared difference function is similar to the cross correlation with

$$\tau_{i,j} = \arg\min_{\tau}\left\{\mathrm{R}_{h_i,h_j}^{\mathrm{ASDF}}(\tau)\right\} \tag{15}$$

with

$$\mathrm{R}_{h_i,h_j}^{\mathrm{ASDF}}(\tau) = \int_{-T/2}^{T/2}\left(h_i(t) - h_j(t-\tau)\right)^2 dt\,. \tag{16}$$

Both methods can be normalized with respects to various weighting terms [7]. Other researches showed that the GCC based TDE is robust towards low SNR whereas ASDF based methods feature high accuracy.

The multilateration solution (see eq. 3) demands time differences $\tau_{1,m}$ for $m = 2...8$, whereas our system of microphones offers

$$M = \frac{m(m-1)}{2} \tag{17}$$

independent time difference information, where m is the number of microphones included in the array. Having a setup with 8 microphones leads to 28 time differences (see Tab. I). To utilize all 28 information from our system we propose a weighted average of the extracted TDOAs. Substracting the first column from each column of Tab. I results in multiple information of $\tau_{1,m}$ for $m = 2...8$. Each extracted TDOA can now be weighted before averaged. Weighting of the time differences can be based on signal to noise ratio or correlation coefficients which depend on the issues in the field.

2) Visual Localization:

Two-dimensional Localization: The initial problem to be solved is the detection of a small and fast moving object in a set of images of a single camera. Due to lighting and size of the bat, the low contrast between the object and the background clutter challenges the automated object detection. We introduced an object detection based on the moving variance V of a single value of a pixel $\hat{\mathbf{x}}_i(k,l)$ over a defined number of images N

$$\mathbf{V}(\hat{\mathbf{x}}_i(k,l)) = \frac{1}{N-1}\sum_{i}^{j}\left|\hat{x}_i(k,l) - \mu\right|^2 \tag{18}$$

with

$$\mu = \frac{1}{N}\sum_{i}^{j}\hat{x}_i(k,l) \tag{19}$$

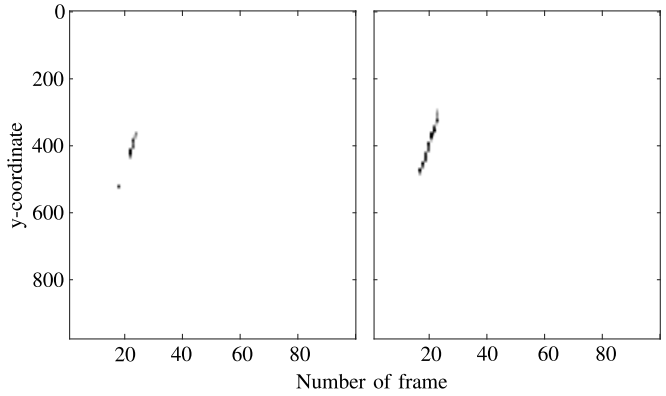

Fig. 4. Images of a partial flightpath recoded based on the y-coordinates in the images of the detection (y-axis) and current frame of the detected objects for both cameras (x-axis).

and

$$i = 1...\text{n-N} \quad j = i + N\,. \tag{20}$$

The overall amount of recorded images n and the window length of $N = 20$ returns $n - N$ variance images $V_{j,i}(k,l)$. Subtracting two following variance images

$$\mathbf{D}_{j,i}(k,l) = \mathbf{V}_{j,i}(k,l) - \mathbf{V}_{j-1,i-1}(k,l) \tag{21}$$

eliminates regular clutter and movement in the images and the moving object can be extracted using a threshold

$$\mathbf{x}_{j,i} = \mathbf{D}_{j,i}(k,l) > \text{thresh} \tag{22}$$

where $\mathbf{x}_{j,i}$ represents the 2D location of a detected object and is further used for the localization.

Three-dimensional localization: Before computing the 3D location of an object, we have to match detections from both cameras to find corresponding points for the triangulation. The majority of researches focuses on a defined manual measurement setup and, thus, they extract and match detections manually. Since our aim is to automate three-dimensional localization of bats, we implemented an automated matching algorithm for single camera detections. The basis of our analysis is that after correcting the single images for lens distortion and skew, corresponding detection share two parameters: y-coordinate and frame. To find correspondences we recoded detections of 100 frames into an image based on the y-coordinates of the detections together with their frame (see Fig. 4). These two y over time images (see Fig. 4) were afterwards correlated (Fig. 5). If the 100 evaluated images feature a corresponding flightpath in the left and right camera, we will see a maximum of the correlation in the center (Fig. 5).

The object detection followed by the matching algorithm returns corresponding object points $\mathbf{x}$ and $\mathbf{x}'$ from the two views. Due to reprojection errors and skew the points will not satisfy $\mathbf{x} = \mathbf{PX}$, $\mathbf{x}' = \mathbf{P'X}$ respectively. Thus, a triangulation method Λ

$$\mathbf{X} = \Lambda(\mathbf{x}, \mathbf{x}', \mathbf{P}, \mathbf{P'}) \tag{23}$$

is used according to [10] to compute the 3D coordinates of $\mathbf{X}$ from two corresponding points $\mathbf{x}, \mathbf{x}'$ and their camera matrices $\mathbf{P}, \mathbf{P'}$.

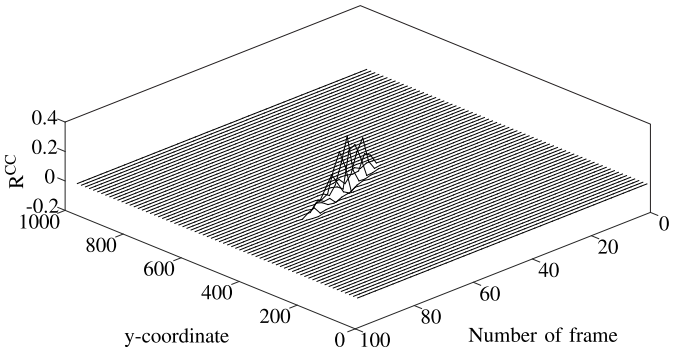

Fig. 5. Correlation result of the recoded images from Fig. 4; maximum in the correlation indicate corresponding detections.

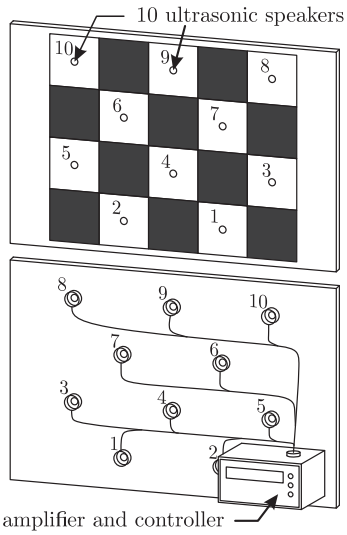

Fig. 6. Visual-acoustical checkerboard pattern; 10 ultrasonic speakers for acoustic localization combined with checkerboard pattern for automated visual localization.

E. Calibration

Previous chapters introduced the two localization systems in a separated manner whereas the following describes the matching process of the acoustical and visual localization. The two systems are independent of each other which implies that each system has its own coordinate system. In order to fuse the two systems the six degrees of freedom between the coordinate systems need to be determined. These are the translations t_x, t_y and t_z in x, y and z direction and the rotations α, β and γ around the x, y and z axis. The process of determining the necessary translations and rotations is called calibration. Inspired by classical camera calibration we merged a visual checkerboard pattern with a acoustic checkerboard pattern. It consists of a black and white checkerboard pattern, with one speaker placed in the middle of each white field (Fig. 6). The visual localization system detects the junction between the black and white fields, whereas the microphone array localizes the speakers. This novel visual-acoustical checkerboard enables high accuracy automated calibration.

- Identification of the translations in x, y and z-direction: The array was mechanically designed and manufactured with high precision, thus, the positions of the

microphones are known. The two cameras are situated in the plane behind the microphones which enables the visual localization of the center microphone since it is visible from both perspectives. Since the distance between the cameras and the central microphone is only ~ 450 mm we can assume that the resulting coordinates resemble dominantly the three translation unknowns.

- Identification of the rotations around x, y and z-axis: The determination of the three rotation angles requires multiple measurements for each angle to ensure high accuracy. The calibration is for all three angles identical which is why we describe exemplarily the determination of the rotation around the x axis. A set of measurements along the z-axis (depth) in different distances z is used to optimize the rotation parameter.

The rotation transformation

$$\begin{pmatrix} x_v \\ y_v \\ z_v \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix} \begin{pmatrix} x_a \\ y_a \\ z_a \end{pmatrix} \tag{24}$$

can be reduced to

$$y_v = \begin{pmatrix} \cos\alpha & \sin\alpha \end{pmatrix} \begin{pmatrix} y_a \\ z_a \end{pmatrix} \tag{25}$$

where indice v resembles the visual localized coordinates and a, respectively, the acoustic coordinates. Using multiple measurements i we can solve the system of equations

$$\begin{pmatrix} y_{v_1} & \cdots & y_{v_i} \end{pmatrix} = \begin{pmatrix} \cos\alpha & \sin\alpha \end{pmatrix} \begin{pmatrix} y_{a_1} & \cdots & y_{a_i} \\ z_{a_1} & \cdots & z_{a_i} \end{pmatrix} \tag{26}$$

using least mean squares.

III. Results

The introduced system features numerous novel developments starting from single sensor up to the recording hardware as well as the post processing. The following section presents results of the key components of this system.

A. Microphones

To maximize the detection range and facilitate the time delay estimation and, therewith, the localization the microphones are designed with small dimensions and high ultrasonic sensitivity. Knowles designed their SPU0410LR microphones with focus on ultrasonic performance. Thus, the frequency response is relatively flat up to 60 kHz (Fig. 7) compared to commonly used hearing range microphones. The frequency response was measured for the final microphone including its holder. In addition to the flat response, the directivity pattern is highly omnidirectional up to frequency of 70 kHz. At higher frequencies we can see the directivity changing from a single main lobe at an incident angle of 90° to two symmetrical lobes at $\pm 40°$ (Fig. 8). With a 3 dB (6 dB) drop we reach an opening angle of 90 (130) degrees at a frequency of 40 kHz

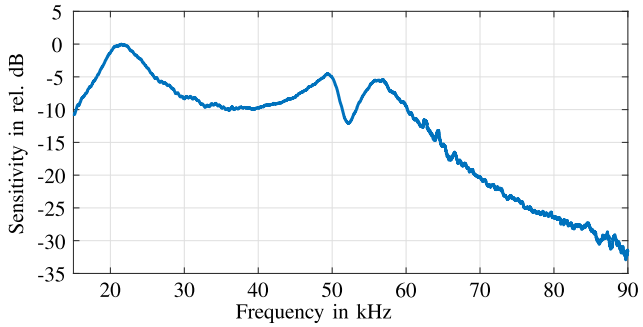

Fig. 7. Frequency response of the developed low-profile ultrasonic microphones; in reference to 1/4 inch free-field microphone by Brüel & Kjær and normalized with reference to the maximum sensitivity.

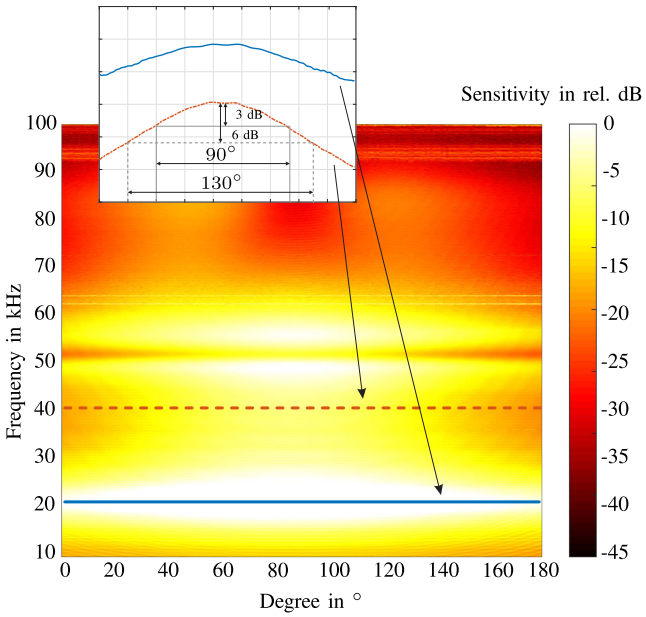

Fig. 8. Directivity pattern of the microphone for different frequencies; in reference to 1/4 inch free-field microphone by Brüel & Kjær; normalized with reference to the maximum sensitivity.

B. Localization Range

An important aspect of any recording system is its detection range or volume. The distance in which an echolocation call can be detected and localized is dependent on various factors such as

- trigger parameter
- SNR, directivity pattern and frequency response of microphones
- SPL and frequency of echolocation call
- temperature and humidity.

Since there are numerous influencing factors it is not possible to determine a generally valid localization range. The range has to be determine for the specific surroundings and therefore we present method of estimating the detection range and volume. Assuming a temperature T of 20°, relative humidity φ of 70 % and a frequency f of 20 kHz we can determine the damping $a(f, T, \varphi)$ of the ultrasound in air according to Bass *et al.* [2], [3]. Together with the resulting damping of

TABLE II

DETECTION RATE USING 1000 RANDOMLY CHOSEN RECORDINGS OF COBS BACKORDER OF WHICH 380 WERE BAT CALLS AND 620 WERE NOISE FILES; CLASSIFICATION BASED ON MANUAL EVALUATION

	Reference	True positive	Ratio in %
Bat calls	380	342	90.0
Noise	620	592	95.5
Overall	1000	934	93.4

the echolocation call and its presumed sound pressure level of 120 dB at a distance of 10 cm, we can iteratively solve the detection range d_r

$$s(f) - 20\log_{10}\left(\frac{d_r}{0.1\,\mathrm{m}}\right) - (a(f,T,\varphi)\cdot d_\mathrm{r}) + P_\mathrm{trigger} = 120\,\mathrm{dB} \quad (27)$$

with

$$s(f) = P_\mathrm{m}(f) - P_\mathrm{m}(f_\mathrm{ref}) \quad (28)$$

where P_trigger is the trigger sound pressure level, $P_\mathrm{m}(f)$ the microphone sensitivity in dB at the frequency f and $P_\mathrm{m}(f_\mathrm{ref})$ the sensitivity in dB at the calibration reference frequency f_ref. The signal-to-nose ratio limits the minimal trigger level P_trigger. To ensure sufficient detection rate we defined the minimal SNR to 6 dB which led to a maximum detection range of 50 m at the frequency of 20 kHz, a temperature of 20 °C and a relative humidity of 70 %. Echolocation calls of higher frequency or lower initial sound pressure level will result in lower detection range.

C. Automated Recording

A key feature of this system is its capability to compute a real-time FFT on all eight channels (capable of up to 20 channels). Using this feature it was possible to develop a highly sophisticated bat call trigger algorithm. Our novel trigger is based both on time and frequency domain parameters and is set to record all microphones if only one microphone detects a bat echolocation call. We evaluated our trigger by loading 1000 recordings of bat calls and noise made with Backorder (cobs GmbH, Nuremberg, Germany) directly into the memory and running the trigger computation on each file. All 1000 files were manually evaluated which resulted in 620 noise recordings (false positive) and 380 bat calls (true positive). Thus, for these 1000 recordings, Backorders triggered positively at a rate of 38 % whereas our novel trigger featured a detection rate of 93.4 % (see Tab. II).

D. Localization Accuracy

Figure 9 shows localization results for one measurement of the visual-acoustical checkerboard pattern with the acoustic position of the 10 speakers ♦ and the corresponding visual position ● of the speakers within the pattern. Due to various dependencies, the two independent localization methods lead slightly differing results. Accuracy of speed of sound, estimation of time differences of arrival, mechanical errors in the geometry are all reasons for the inconsistency. To quantify localization errors, we performed several measurements

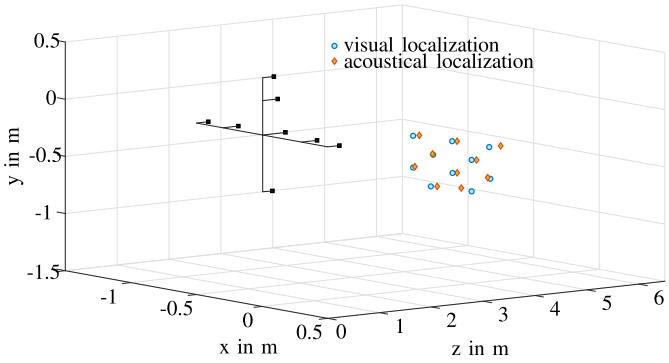

Fig. 9. Visual ● and acoustical ♦ located checkerboard pattern with ten ultrasonic speakers.

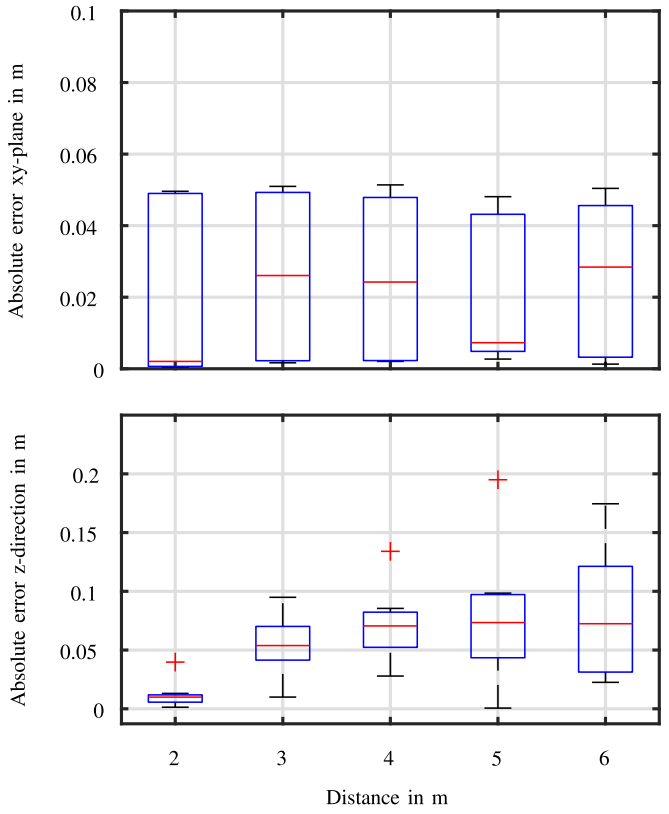

Fig. 10. Absolute localization error over ten speakers along the z-axis for multiple measurements with increasing distance; top: 2D-error in the xy-plane with a maximum of 5 mm and variation constant over distance; bottom: error in z-direction with its maximum at 6 m distance and strong dependency on distance.

in increasing distances. Figure 10 illustrates absolute errors over multiple measurements for each distance. As illustrated, the accuracy within the xy-plane seems to not strongly dependent on distance. On the contrary, the error in the z-coordinate is increasing with distance. We achieved within a room of 6 **x** 6 **x** 3 m a localization accuracy of ±5 cm in the xy-plane and ±12 cm (75 percentile) in the z-coordinate.

The results also depend on the implemented localization algorithm. Figure 11 compares the straightforward multilateration with the Smith-Abel Algorithm. The first 25 measurements were recorded in a non-echoic environment whereas

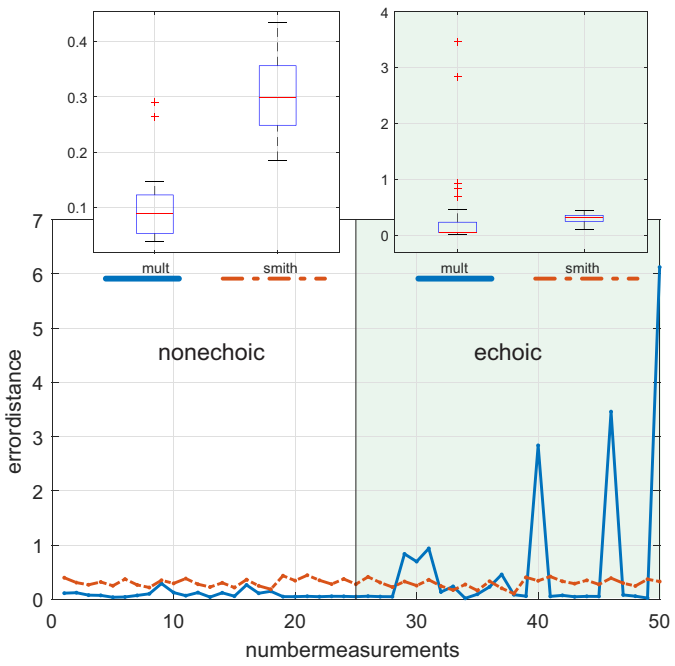

Fig. 11. Absolute localization error over multiple measurements; first 25 measurements with non-echoic surroundings and good SNR; second 25 measurements with reflecting surfaces producing echo overlays; linear solution of the multilateration produces overall better results but solving with the Smith-Abel-Algorithm tends to be tolerant against outliers.

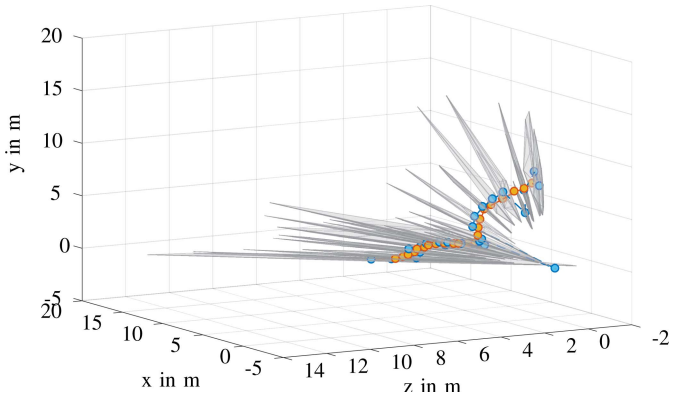

Fig. 12. Acoustically tracked flight path of pipistrellus pipistrellus; grey volumes mark the error range due to tdoa measurements; ●: unfiltered flightpath; ●: filtered with tdoa error and correlation values.

the second 25 measurements included echoic overlays. The exact solution of the multilateration shows superior results over all measurements but produces outliers in echoic conditions. Solving the system according to Smith and Abel [22] results in a mean error of approximately 0.3 m over all measurements. This error and its distribution was constant over multiple measurement independent on the existence of environment.

Figure 12 shows a recorded three dimensional flight path. Knowledge on the possible errors is vital for quality results. The gray volumes represent the possible error range for each localized echolocation call due to randomized error in the determined TDOAs of $\Delta\tau_{i,j} = 0 \cdots \pm 20\,\mu$ s.

IV. Conclusion

We introduced a novel multi-sensor array to track flight paths of bats in three-dimensional space. The combination of acoustic localization of each echolocation call and visual 3D tracking of the flight path allows the fusion of different sensor information. Localizing echolocation calls limits the information of the whereabouts on the current position during the call. Visual tracking of the flight path restricted to the field of view of the stereo camera system. Fusing both systems we receive information on the bats behavior outside the field of view of the cameras and more detailed information within the overlapping region of both systems.

The goal of developing this array to track bats in space was to overcome the limitations of currently used array technologies. The greatest disadvantage of microphone and camera arrays in bat research is their costs. Current research arrays used for bat research [6], [16], [25] produce high quality recordings but are based on expensive condenser microphones (mostly *G.R.A.S* or *B&K*) and use costly data acquisition devices. We focused on cost reduction to allow the usage of array technology for low budget researches in the future. The complete system is below 1000 EUR and, therewith, even cheaper than most single channel bat recorders. To achieve such low costs for high speed recordings we developed a novel real-time acoustic processing and recording hardware together with low cost microphones. Our hardware is able to process eight acoustic signals simultaneously in real-time in time and frequency domain and record echolocation calls at 1 MS/s/CH with 16 Bit whereas the widespread Avisoft recording hardware is limited to 750 kHz. Another limitation of current arrays is their usability and applicability in the field. Most studies with arrays are conducted in laboratories [5], [14], [17] with defined conditions and power supply. The introduced system can be used in the lab as well as in the field where it can run approximately 50 h on battery. In addition, the combined system allows automated recordings similar to Avisoft based arrays, but without the necessity of laptops. One great advantage of the Avisoft system is the live display of spectrogram which is still missing in our recording hardware. Furthermore, systems based on commercially available parts such as Avisofts USGs are more robust and tested in handling compared to our novel development. Our mechanical setup is designed to allow untrained researchers to assemble and disassemble the array in the field.

The array was evaluated and calibrated in the lab. We developed an acoustic-visual checkerboard to enable easy calibration as well as error estimation of the stereo cameras and the microphone array. After the calibration the array achieved a precision of $\pm$ 10 cm in z-direction and sub centimeter in the xy-plane. Being able to track with high precision is not the only feature of the introduced system. We can automatically log environmental parameters such as temperature, humidity and illumination. To georeference all automated recorded datasets, the system is equipped with a GPS module, a compass module and a three axis accelerometer.

Enhancing microphone arrays as well as stereo-camera systems towards usability and automated field applicability

will promote more research projects to make use of these systems. Ideally their employment will be as straight forward as current single channels bat recorders. Acoustic monitoring of bat activity with single channel recorders is a powerful tool for long-term assessments but adding a location to each echolocation call could be beneficial for many studies. For example studies of flight and approach behavior of bats around wind turbines would greatly benefit from knowledge of the location. We were able to deploy the stereo-infrared system on two behavioral studies at small wind turbines [11]. With our continuing effort to improve the localization systems we aim towards lowering hurdles to employ localization systems for various research studies.

REFERENCES

[1] *Avisoft Bioacoustics*, Avisoft Bioacoustics e.K., Berlin, Germany, 2018.

[2] H. E. Bass, L. C. Sutherland, A. J. Zuckerwar, D. T. Blackstock, and D. M. Hester, "Atmospheric absorption of sound: Further developments," *J. Acoust. Soc. Amer.*, vol. 97, no. 1, pp. 680–683, Sep. 1994.

[3] H. E. Bass, L. C. Sutherland, A. J. Zuckerwar, D. T. Blackstock, and D. M. Hester, "Erratum: Atmospheric absorption of sound: Further developments," *J. Acoust. Soc. Amer.*, vol. 99, no. 2, p. 1259, Feb. 1996.

[4] O. Behr *et al.*, "Mitigating bat mortality with turbine-specific curtailment algorithms: A model based approach," in *Wind Energy and Wildlife Interactions*, J. Köppel, Ed. Cham, Switzerland: Springer, 2017, pp. 135–160.

[5] A. Boonman, Y. Bar-On, N. Cvikel, and Y. Yovel, "It's not black or white—On the range of vision and echolocation in echolocating bats," *Frontiers Physiol.*, vol. 4, p. 248, Sep. 2013.

[6] S. Brinkløv, L. Jakobsen, J. M. Ratcliffe, E. K. V. Kalko, and A. Surlykke, "Echolocation call intensity and directionality in flying short-tailed fruit bats, Carollia perspicillata (Phyllostomidae)," *J. Acoust. Soc. Amer.*, vol. 129, no. 1, pp. 427–435, Jan. 2011.

[7] L. Chen, Y. Liu, F. Kong, and N. He, "Acoustic source localization based on generalized cross-correlation time-delay estimation," *Procedia Eng.*, vol. 15, pp. 4912–4919, 2011.

[8] E. Fujioka, S. Mantani, S. Hiryu, H. Riquimaroux, and Y. Watanabe, "Echolocation and flight strategy of Japanese house bats during natural foraging, revealed by a microphone array system," *J. Acoust. Soc. Amer.*, vol. 129, no. 2, pp. 1081–1088, Feb. 2011.

[9] J. E. Gaudette, L. N. Kloepper, M. Warnecke, and J. A. Simmons, "High resolution acoustic measurement system and beam pattern reconstruction method for bat echolocation emissions," *J. Acoust. Soc. Amer.*, vol. 135, no. 1, pp. 513–520, Jan. 2014.

[10] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, vol. 2. New York, NY, USA: Cambridge Univ. Press, 2003.

[11] S. A. Hartmann *et al.*, "Bats and small wind turbines in Northern Germany: 3D analysis of flightpaths to determine collision risk," *PLoS ONE*, to be published.

[12] S. Hiryu, M. E. Bates, J. A. Simmons, and H. Riquimaroux, "FM echolocating bats shift frequencies to avoid broadcast–echo ambiguity in clutter," *Proc. Nat. Acad. Sci. USA*, vol. 107, no. 15, pp. 7048–7053, Apr. 2010.

[13] S. Hiryu, T. Hagino, E. Fujioka, H. Riquimaroux, and Y. Watanabe, "Adaptive echolocation sounds of insectivorous bats, pipistrellus abramus, during foraging flights in the field," *J. Acoust. Soc. Amer.*, vol. 124, no. 2, pp. EL51–EL56, Aug. 2008.

[14] M. W. Holderied, C. Korine, M. B. Fenton, S. Parsons, S. Robson, and G. Jones, "Echolocation call intensity in the aerial hawking bat Eptesicus bottae (Vespertilionidae) studied using stereo videogrammetry," *J. Exp. Biol.*, vol. 208, no. 7, pp. 1321–1327, Apr. 2005.

[15] M. W. Holderied, G. Jones, and O. V. Helversen, "Flight and echolocation behaviour of whiskered bats commuting along a hedgerow: Range-dependent sonar signal design, Doppler tolerance and evidence for 'acoustic focussing,'" *J. Exp. Biol.*, vol. 209, no. 10, pp. 1816–1826, May 2006.

[16] L. Jakobsen and A. Surlykke, "Vespertilionid bats control the width of their biosonar sound beam dynamically during prey pursuit," *Proc. Nat. Acad. Sci. USA*, vol. 107, no. 31, pp. 13930–13935, Aug. 2010.

[17] J. C. Koblitz, P. Stilz, and H.-U. Schnitzler, "Source levels of echolocation signals vary in correlation with wingbeat cycle in landing big brown bats (Eptesicus fuscus)," *J. Exp. Biol.*, vol. 213, no. 19, pp. 3263–3268, Oct. 2010.

[18] J. C. Koblitz *et al.*, "Arrayvolution—An overview of array systems to study bats and toothed whales," *J. Acoust. Soc. Amer.*, vol. 136, no. 4, p. 2091, Oct. 2014.

[19] M. G. C. MacSwiney, F. M. Clarke, and P. A. Racey, "What you see is not what you get: The role of ultrasonic detectors in increasing inventory completeness in Neotropical bat assemblages," *J. Appl. Ecol.*, vol. 45, no. 5, pp. 1364–1371, Oct. 2008.

[20] H. Peremans, R. Mueller, J. M. Carmena, and J. C. T. Hallam, "Biomimetic platform to study perception in bats," *Proc. SPIE*, vol. 4196, pp. 168–180, Oct. 2000.

[21] J. A. Simmons and J. E. Gaudette, "Observing the invisible: Using microphone arrays to study bat echolocation," *Acoust. Today*, vol. 10, no. 3, pp. 16–24, 2014.

[22] J. O. Smith and J. S. Abel, "Closed-form least-squares source location estimation from range-difference measurements," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, no. 12, pp. 1661–1669, Dec. 1987.

[23] J. Steckel and H. Peremans, "A novel biomimetic sonarhead using beamforming technology to mimic bat echolocation," *IEEE Trans. Ultrason., Ferroelect., Freq. Control*, vol. 59, no. 7, pp. 1369–1377, Jul. 2012.

[24] A. Streicher, R. Müller, H. Peremans, and R. Lerch, "Broadband ultrasonic transducer for a artificial bat head," in *Proc. IEEE Symp. Ultrason.*, Oct. 2003, pp. 1364–1367.

[25] A. Surlykke and E. K. V. Kalko, "Echolocating bats cry out loud to detect their prey," *PLoS ONE*, vol. 3, no. 4, Apr. 2008, Art. no. e2036.

Klaus Hochradel, photograph and biography not available at the time of publication.

Timm Häcker, photograph and biography not available at the time of publication.

Tino Hohler, photograph and biography not available at the time of publication.

Andreas Becher, photograph and biography not available at the time of publication.

Stefan Wildermann, photograph and biography not available at the time of publication.

Alexander Sutor, photograph and biography not available at the time of publication.