UAV Sound Source Localization

Computational Neuro Engineering Project Laboratory
FINAL REPORT
handed in
by

Peter Hausamann

born on May 4th, 1990
residing in:
Kreillerstraße 71
81673 München

Institute of
AUTOMATIC CONTROL ENGINEERING
Technical University of Munich

Univ.-Prof. Dr.-Ing./Univ. Tokio Martin Buss

Supervisor: M.Sc. Cristian Axenie
Beginning: October 18th, 2013
Submission: January 14th, 2014

Abstract

Locating sound sources in the environment is an important part of perception for many biological organisms. All vertebrates make use of two ears in order to detect and localize sounds. Implementing a similar approach on a robot, in this case a quadrotor drone, makes it possible for the robot to perform localization tasks and act accordingly. An important challenge in the case of quadrotors is the inevitable operation noise during flight. This project describes a basic platform for stereo sound acquisition with signal processing performed off-board. A pair of microphones is mounted on a drone and transmits audio data via an FM radio link. This audio data is then recorded and processed on a remote computer. While the basic platform could be set up, many challenges regarding hardware and software have been encountered. This work should therefore contribute to developing a robust sound source localization system on a quadrotor.

Contents

Chapter 1

Introduction

1.1 Motivation of Sound Source Localization

A lot of developed animals make use of binaural hearing[1] to locate sound sources. This is important because some kinds of sounds represent dangers or similar events or objects of interest. Determining the location of the perceived sound is crucial in order to choose an appropriate behaviour and coordinate directed actions like flight or attack.

Implementing similar capabilities on a robot, in our case a UAV (unmanned aerial vehicle, drone), can be useful for various reasons. One possible application would be a scenario where a test person "calls up" the drone. The UAV would be able to determine a cue sound's location and fly towards it. For this purpose, a stereo microphone setup and a biologically inspired signal processing scheme is necessary. This allows for an automatic classification, cue selection and subsequent localization of the sound.

1.2 Objectives

The goal of this project is to set up the basic platform for a sound source localization application, including:

- Mounting a stereo microphone pair on the drone and evaluating the microphones' characteristics. The microphones send out the picked up audio signal via an FM radio link.
- Setting up the interface for picking up the transmitted radio signal and recording it into a computer. Capturing remote control data sent to the drone and ensuring its synchronicity with the audio recording.

[1] Hearing with two ears

- Preparing a basic signal processing setup for later use for the localization. The signal processing scheme should be inspired by biological systems such as the human hearing.

During the course of this laboratory it has become evident that the provided hardware is not suitable for the intended purpose of sound localization, especially in regard to the drone's high operation noise. This report should therefore be seen as a conceptual study for a possible sound localization setup as well as a guideline for future work.

Chapter 2

Main Part

2.1 Hardware Setup

2.1.1 Drone

The drone in use is based on the PX4 PIXHAWK MAV (micro air vehicle) developed by researchers at ETH Zürich [MTH+], i.e. it uses the PX4FMU flight controller and the PX4IOAR hardware adapter. The electronics can independently control four servo-driven rotors.

The UAV can be controlled via a WiFi link using a dedicated Linux application. A joystick connected to the Linux computer is used for remote control. The drone's electronics translate the joystick commands (roll, pitch, yaw and thrust) into control data for the servos. A schematic of the remote control link is shown in figure 2.1.

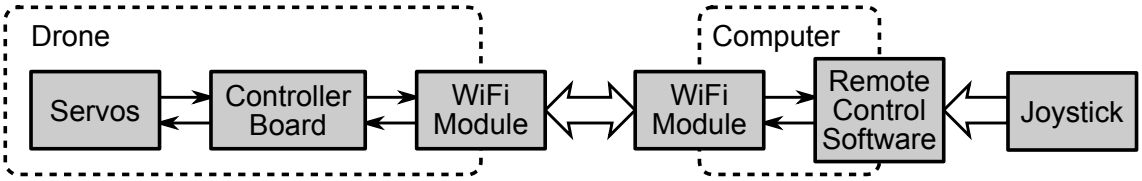

Figure 2.1: Functional diagram of the drone remote control

2.1.2 Microphones

Principle

The microphone supplied for this project is a FM radio spy microphone. Its dimensions are approximately $34 \times 15 \times 10$ mm excluding the power supply and antenna cables. It consists of a sound transducer, an amplifier circuit and a FM radio transmitter (see figure 2.2).

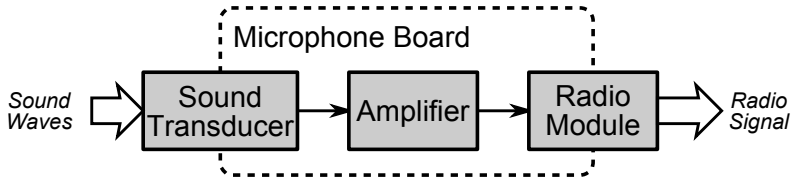

Figure 2.2: Functional diagram of the microphone module

A picture of one of the used microphones can be seen in figure 2.3. The pin visible in the top right corner is not part of the hardware, it has been soldered on the board as a ground pin for oscilloscope measurements.

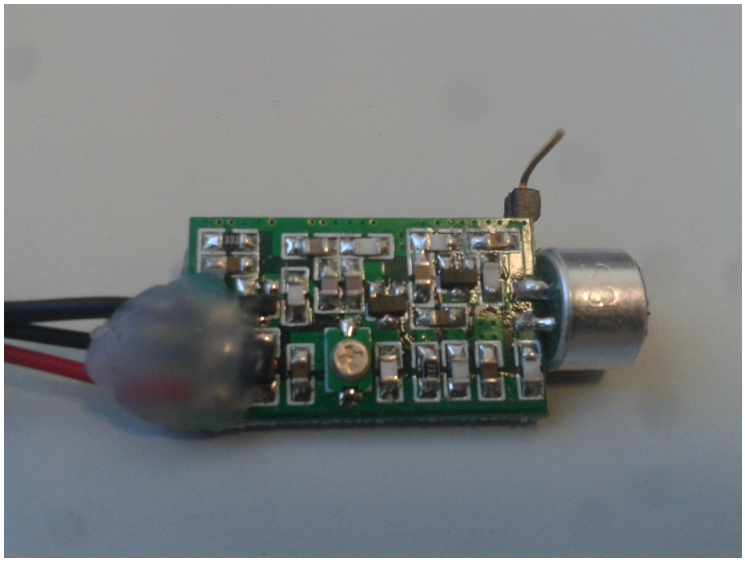

Figure 2.3: The microphone in use

Performance Measure

Frequency Response The measurement was performed by playing back a logarithmic frequency sweep from 20 Hz to 20 kHz with 20 seconds duration. A Yamaha HS80M speaker at 1 meter distance from the microphones was used for reproduction. The speaker has an approximately linear frequency response between 80 Hz and 20 kHz [Yam05, p. 67]. The reference level was not measured for lack of proper equipment. Figure 2.4 shows the mean frequency response of eight measurements.

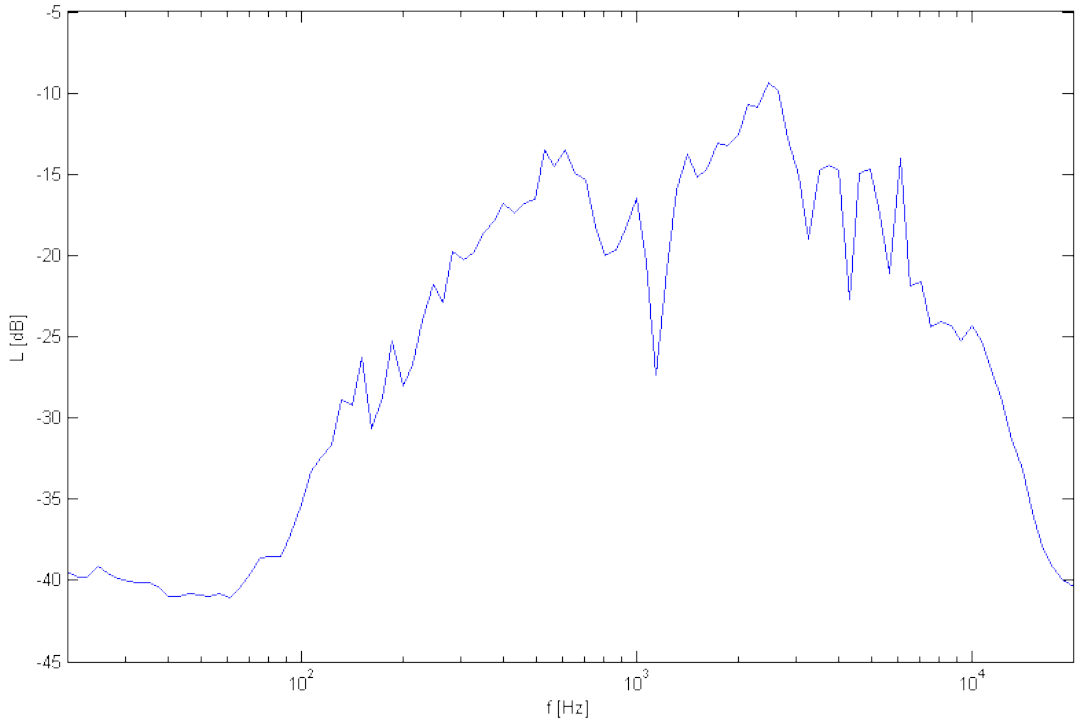

Figure 2.4: Frequency response of the microphones

The fact that the frequency response in the "pass band" has notches of magnitudes up to 15 dB shows that the used microphones are not appropriate for the intended purpose.

Directivity This measurement was performed by recording a sine wave signal of 880 Hz (A2) with a duration of 20 seconds from eight different directions. The speaker was also placed at 1 meter distance. The results are shown in figure 2.5.

The level of the picked up sound does not vary greatly with the sound's incidence angle, the microphone's directivity is therefore *omnidirectional*. This type of behaviour is also not favourable for this application as sound signals are not attenuated depending on their direction. While it is not an inherent drawback for stereophony and sound localization (see section 2.1.2), a more directed characteristic would be beneficial for suppressing operation noise from the rear rotors.

Dynamic Range A systematic measurement of the microphone module's SNR / dynamic range was not done because proper equipment was not available. However, it has become evident that the microphones are extremely sensitive because of their intended purpose as spy microphones. Because of this, they distort heavily even at low input levels (e.g. a person talking at normal conversation level into the micro-

Figure 2.5: Directivity of the microphones

phone at 20 cm distance). This is another fact that disqualifies the microphones from being used in a high noise environment as is present in this case.

Hardware Mount

The microphones are mounted in front of the drone, sticking out at a 45 degree angle each. This is inspired by the so called time-of-arrival stereophony (or A-B stereophony) principle [GÖ8, p. 302]. This mounting scheme is not exactly according to the A-B stereophony standard, where the microphones are supposed to be mounted in parallel. However, as shown in section 2.1.2, since the microphones possess an omnidirectional directivity the direction of the microphones themselves is irrelevant.

The horizontal separation of the microphones is about 60 mm. A sketch of the mounting board is shown in figure 2.6.

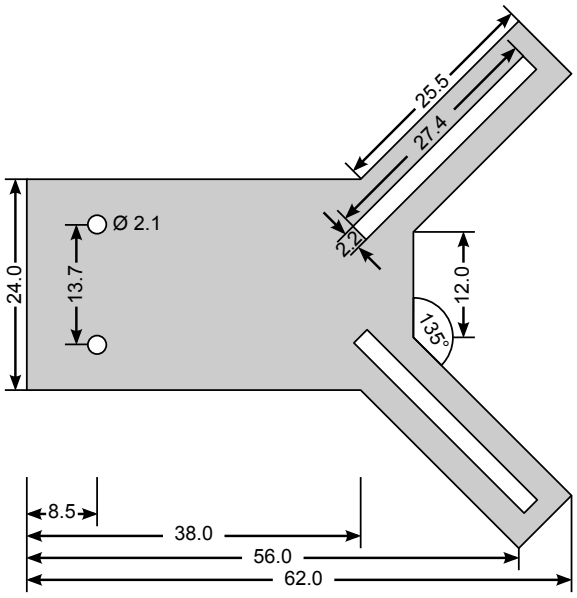

Figure 2.6: Schematic of the mounting device

The mounting device is cut out of a 1 mm thick PVC board with a laser cutter. The board is attached to the drone's hardware adapter with two M2 screws. Figure 2.7 shows the microphone pair mounted on the drone.

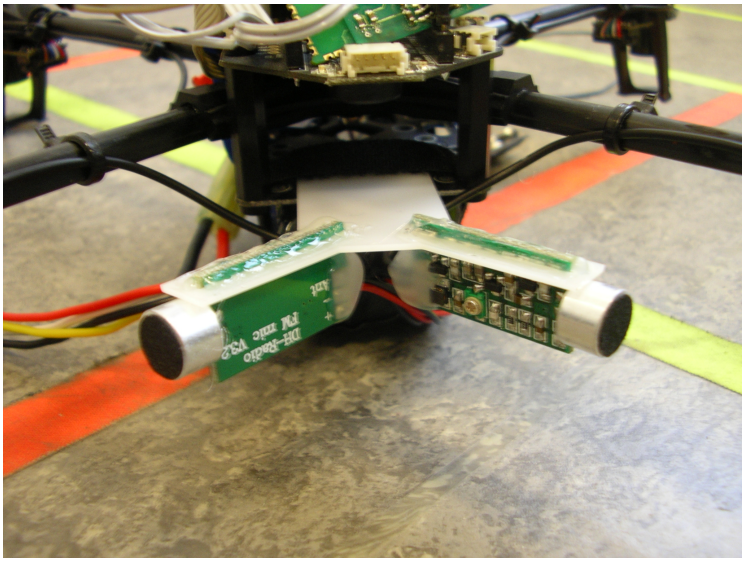

Figure 2.7: Mounting of the microphones on the drone

2.1.3 Radio Transmission

The microphones send out audio data via an FM radio antenna cable. The cables have a length of approximately 1.5 meters, corresponding to half a wavelength of a 100 MHz radio wave. The sending frequency can be modified with a potentiometer in a small range around 100 MHz but is also highly susceptible to antenna position, foreign objects, temperature and other factors.

The antenna cables were mounted in an X-shape along the arms of the drone (see figure 2.8). While this may not be the optimal configuration with regard to signal transmission, it is the most practical solution seeing as the cables are very long.

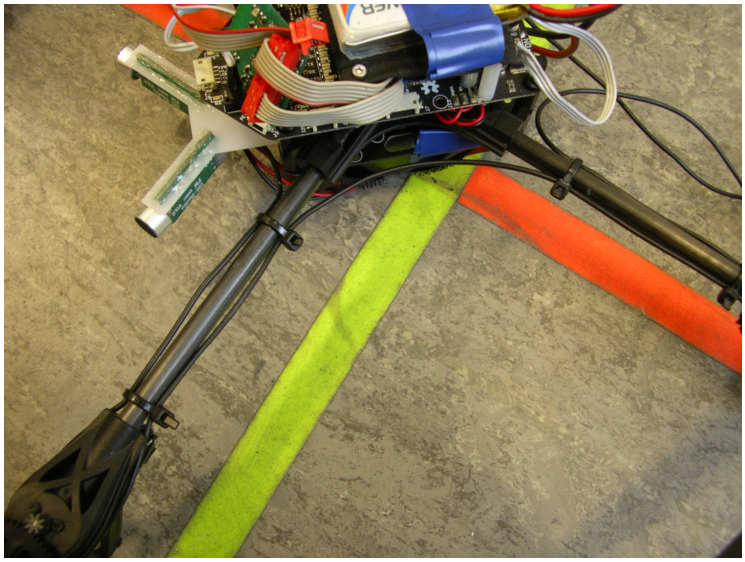

Figure 2.8: Mounting of the radio wires on the drone

Two hand-held consumer radios are used as receivers for the transmitted microphone signals. It should be noted that because of the high susceptibility of the signal strength to environment factors, the output volume of the sound signal from both radio receivers cannot be assumed as constant over time. This is one of the reasons level-based stereophony (see sections 2.1.2 and 2.2.3) is not usable in this setup.

2.2 Signal Acquisition and Processing

2.2.1 Hardware/Software Interface

All signal processing is done off-board. The audio data from the radio receiver is recorded via an audio interface into the signal processing software. For this project all signal processing has been done “off-line” with MATLAB for rapid prototyping purposes. In the future it would be beneficial to develop a standalone application

(C++/Python) that has the capability to process "live" audio streams. A signal flow chart of the hardware/software interface can be seen in figure 2.9.

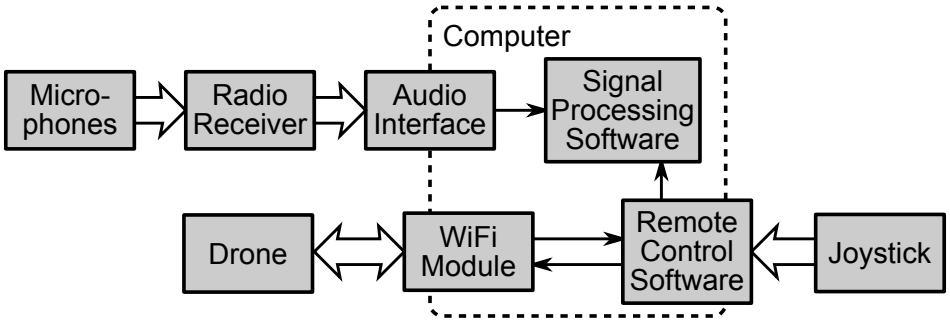

Figure 2.9: Functional diagram of the hardware/software interface

The implemented MATLAB-script allows the user to specify settings for the audio recording such as sampling frequency, number of channels and duration. Afterwards, it launches the drone remote control software while simultaneously recording audio data according to the specified settings. The user can then control the drone with a connected joystick. After the recording is complete, the remote control software is terminated and the logged data is being parsed. All collected data is then saved to a cell array in a `.mat`-file for later processing.

2.2.2 Synchronization

The code of the remote control application has been modified so as to log the sent out joystick data. The application logs roll, pitch, yaw and throttle together with a UNIX timestamp. This ensures synchronicity with the audio stream and can be used in the future to implement a signal adapter filtering scheme in order to filter out rotor noise. For this purpose it would be especially interesting to not just monitor the joystick data but rather the PWM signals sent to the rotor servos by the drone's own controller board. This would however need to be implemented in the UAV's firmware.

2.2.3 Sound Source Localization

Theory of Sound Source Localization

The human hearing system can determine the direction of an incoming sound with high precision in the horizontal plane (azimuth plane). For this purpose it uses two measures for determining the azimuth angle ϕ: interaural time differences (ITD) and interaural level differences (ILD). These correspond to the already mentioned concepts of time-of-arrival and intensity stereophony (section 2.1.2). The localization in the vertical (meridian) plane is far less precise and mostly utilizes the so called HRTF (head related transfer function) [Bla83].

Interaural Time Differences (ITD) This measure takes advantage of the fact that a sound has to travel a slightly longer distance to one of the ears if its source is located at an azimuth angle $\phi \neq 0$. This results in a time delay between the picked up sound signals. Figure 2.10 shows the dependency of the delay time on the sound's incidence angle. Note that the distance of the sound source has to be significantly greater than the distance between the ears (or equivalent sound transducers) d in order for this approximation to be correct.

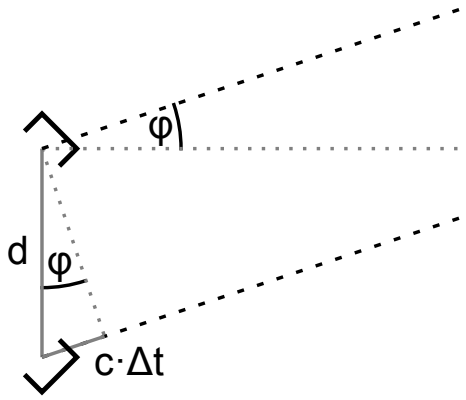

Figure 2.10: Time delay between sound signals depending on sound direction

When the time delay can be determined, the incidence angle ϕ (see figure 2.1) can be calculated like this:

$$\phi = \sin^{-1}\left(\frac{c \cdot \Delta t}{d}\right) \tag{2.1}$$

The time delay reaches its maximum $\Delta t_{\max}$ when the sound source is located at $\pm 90°$ azimuth. Frequencies with wavelengths shorter than $\Delta t_{\max}$ (and thus higher than $f_{\max} = \frac{1}{\Delta t_{\max}}$) yield ambiguous results and can not be located precisely. For the human hearing, the limit frequency $f_{\max}$ lies around 1.6 kHz [GÖ8].

Interaural Level Differences (ILD) This measure results from the fact that the human head diffracts sound waves of certain wavelengths. This causes a higher sound pressure on the side where the sound is coming from and thus a level difference between the ears. This phenomenon has a lower limit frequency determined by the nature of sound refraction which lies around 2 kHz.

Application

As already mentioned, the current setup allows only for ITDs to be evaluated. The time delay between the two microphones can be determined by cross correlating the

sound signals. The cross correlation of two discrete signals x and y is calculated as follows:

$$R_{xy}(k) = \frac{1}{N - |k|} \sum_{n=-K}^{K} x_{n+k} y_n^* \tag{2.2}$$

The parameter k represents a variable offset between the two signals. $K = \frac{F_S \cdot d}{c}$ is the maximal offset which corresponds to the maximum possible time delay (refer to section 2.2.3). With $F_S = 44.1$ kHz and $d \approx 6$ cm, we obtain $K = 8$. Note that the cross correlation is normalized to the length of the overlap of the two signals, otherwise the correlation at $k = 0$ would be favoured.

Several test recordings were made in a noise free environment in order to evaluate this approach. The first set of recordings consists of pulses of sine waves of different frequencies between 20 Hz and 20 kHz. The second set includes pulsed noises such as clapping and snapping. Before the correlation the signals were filtered with a lowpass filter with a 6 dB cutoff frequency of 5.71 kHz which corresponds to the maximum locatable frequency (see section 2.2.3). Table 2.1 shows the horizontal localization results for different sound incidence angles.

Type	Length [s]	Incidence angle [°]	Calculated angle [°]
Pulsed sine	16	45	51.05
	16	0	7.45
	16	-45	-22.89
Impulse	11.5	45	31.23
	16	0	-7.45
	14	-45	-15.03

Table 2.1: Calculated directions for sounds from different incidence angles

It is obvious that these results are very poor. This is amongst others due to the poor signal quality received from the microphones. Another factor is the rather small horizontal separation of the microphones which should have been considered beforehand.

Chapter 3

Summary

The primary goal of this laboratory, setting up a basic platform for a sound localization system on a drone, could unfortunately not be accomplished. The reason for this was the poorly chosen microphone hardware. This fact had unfortunately not become evident until an advanced stage of the project. A basic hardware mount and signal acquisition interface has been set up and a low level localization task could be performed, although very poorly.

However, a lot of insight on the possibilities and limitations of sound localization systems with remote signal processing has been gained. Firstly, especially regarding the high operation noise of the drone, microphones have to be chosen carefully and must be able to withstand high sound pressure levels with minimal distortion. Furthermore, the FM radio link has proven to be unfit for the purpose due to the high susceptibility of the signal strength to environment factors. Since the drone's WiFi link does not yet support the high bit rate needed for audio transmission, a different transmission standard (e.g. Bluetooth) could be taken into consideration. One may also consider implementing the signal processing on the drone itself, seeing as it has been designed for highly complex computer vision tasks [MTH+].

Regarding the signal processing for localization, the focus on bio-inspired techniques, especially related to the human auditory system, should be much more prominent. The dimensions of the drone allow for a setup that could very similar to the human head. The microphones could be mounted on both sides of the drone and special casing could be developed that imitates the human HRTF.

Many more ideas can be derived from the physiology of the human auditory system and psychoacoustics. The basilar membrane in the human cochlear, for example, acts like a bandpass filter bank [ZF90]. A biologically inspired signal processing scheme should take this fact into account and apply a similar filtering scheme before the implementation of the localization system.

Another important aspect, especially in low SNR scenarios, is so called cue selection. This topic deals with the capability of distinguishing multiple sound sources in a reverberant environment [FM04]. A robust cue selection algorithm is crucial in order for the drone to determine the sound source it is supposed to locate, especially considering its own rotor noise.

Finally, it would be beneficial to apply a signal adapted filtering scheme in order to filter out the drone's rotor noise. For this purpose, the PWM data from the UAV's servos could be used. A noise model depending on each rotor's speed could be estimated and appropriate notch filters applied to the sound signals. This would make a robust detection and localization of sound sources, in spite of the noisy environment, possible.

List of Figures

Bibliography

[Axe14] Axenie, C., Conradt, J. (2014) Cortically inspired sensor fusion network for mobile robot egomotion estimation. Robotics and Autonomous Systems, Special Issue on "Emerging Spatial Competences: From Machine Perception to Sensorimotor Intelligence" (obtained pre-print).

[Bla83] Jens Blauert. Spatial Hearing. The Psychophysics of Human Sound Localization. The MIT Press, USA-Cambridge MA, 1983.

[Con02] Conradt, J., Simon, P., Pescatore, M., and Verschure, PFMJ. (2002). Saliency Maps Operating on Stereo Images Detect Landmarks and their Distance, International Conference on Artificial Neural Networks (ICANN2002), p. 795-800, Madrid, Spain.

[Den13] Denk C., Llobet-Blandino F., Galluppi F., Plana LA., Furber S., and Conradt, J. (2013) Real-Time Interface Board for Closed-Loop Robotic Tasks on the SpiNNaker Neural Computing System, International Conf. on Artificial Neural Networks (ICANN), p. 467-74, Sofia, Bulgaria.

[FM04] Christof Faller and Juha Merimaa. Source localization in complex listening situations: Selection of binaural cues based on interaural coherence. J. Acoust. Soc. Am, 116:3075-3089, 2004.

[Gö08] Thomas Görne. Tontechnik. Hanser, 2°d edition, 2008.

[Hof13] Hoffmann R., Weikersdorfer D., and Conradt J. (2013) Autonomous Indoor Indoor Exploration with an Event-Based Visual SLAM System, European Conference on Mobile Robots, p. 38-43, Barcelona, Spain.

[MTH+] Lorenz Meier, Petri Tanskanen, Lionel Heng, Gim Hee Lee, Friedrich Fraundorfer, and Marc Pollefeys. Pixhawk: A micro aerial vehicle de-sign for autonomous flight using onboard computer vision. Autonomous Robots, pages 1-19. 10.1007/sl0514-012-9281-4.

[Wei12] Weikersdorfer D., Conradt J. (2012), Event-based Particle Filtering for Robot Self-Localization, Proceedings of the IEEE International Conference on Robotics and Biomimetics (IEEE-ROBIO), pages: 866-870, Guangzhou, China.

[Wei13] Weikersdorfer D., Hoffmann R., and Conradt J. (2013) Simultaneous Localization and Mapping for event-based Vision Systems, International Conference on Computer Vision Systems (ICVS), p. 133-142, St. Petersburg, Russia.

[Wei14] Weikersdorfer D., Adrian DB., Cremers D., and Conradt J. (2014) Event-based 3D SLAM with a depth-augmented dynamic vision sensor, IEEE ICRA 2014, Int. Conf. on Robotics and Automation, Hong-Kong, China, June 2014 (obtained pre-print).

[Yam05] Yamaha Corporation. HS Series Owner's Manual, 2005.

[ZF90] Eberhard Zwicker and Hugo Fastl. Psychoacoustics: Facts and Models. Springer series in information sciences. Springer-Verlag, 1990.