Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Electronic Systems

Kristian Øverli

Acquisition System with Recording of Audio and Images for Hymenoptera

Master's thesis in Electronics Systems Design and Innovation
Supervisor: Guillaume Dutilleux
July 2022

NTNU
Norwegian University of
Science and Technology

Kristian Øverli

Acquisition System with Recording of Audio and Images for Hymenoptera

Master's thesis in Electronics Systems Design and Innovation
Supervisor: Guillaume Dutilleux
July 2022

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Electronic Systems

NTNU
Norwegian University of
Science and Technology

Abstract

The main objective of this Master's thesis is investigating, designing, implementing and evaluate a data acquisition system for recording audio and capturing images. The system is designed for the insect order of Hymenoptera.

A prototype of the data acquisition system is built on Sony Spresense. The Master's thesis explores the capabilities of the Sony Spresense and the development of hardware and software to enable data acquisition.

The capabilities for four channel audio recording of the Sony Spresense were found to be of high quality and useful in an acquisition system. Memory issues arose with handling of both audio and images with the Arduino IDE. Therefore, we find the selected solution with the Arduino IDE to not be suitable for data acquisition in the envisioned system. Further work includes optimizing the hardware and explore Spresense SDK for data acquisition.

Sammendrag

Målet med denne masteroppgaven er å undersøke, designe, implementere og evaluere et system for å samle data i form av lydopptak og bilder. Systemet er designet for en gruppe av insekter, som inkluderer vepser, humler, geithams og bier.

En prototype av data anskaffelsessystemet er bygget på Sony Spresense. Masteroppgaven utforsker Sony Spresense's evner, med utviklingen av maskinvare og programvare for data ansamling.

Lydansamlingsevnene til Sony Spresense fra fire-kanals opptak var av høy kvalitet og brukbart i det resulterende anskaffelsessystemet. Minneproblemer oppstod under bearbeiding av både lyd og bilde med Arduino IDE. Den valgte løsningen med Arduino IDE ble funnet å ikke være stand til å anskaffe et system for vårt formål. Videre arbeid med systemet inkluderer optimalisering av maskinvare og se på Spresense SDK som mulig løsning for dataanskaffelse.

Acknowlengments

Thanks to Prof. Guillaume Dutilleux for guidance and conversations through out the thesis, the help was much appreciated and Tim Cato Netland for help with equipment and questions. Thanks to Elprolabben for producing the PCB prototype and research and development lab at NTNU for access to equipment and soldering station. Thanks to Peter Remøy Paulsen and Håvard Olai Kopperstad for helping with the 3D-printing of my design. Thanks to Stine Sletvold for proof reading, Torge Øverli for proof reading and improving the structure of the thesis and Andrea Worsøe for discussions about the Master's thesis.

Contents

Chapter 1

Introduction

Biodiversity is important for the planet's ecosystems and an important tool to map biodiversity is detecting insect populations and learning about their habitats. This is described by Hill et al. [1] and Miller et al. [2]. Another important aspect is that the degree of diversity has been worsened by the rapid decline in insect populations [3], [4].

Some of the pollinators we rely on are the Hymenoptera, a large order of insects including wasps, bees and bumblebee. Currently, a method for looking at the biodiversity by counting insects within a specific area includes capturing them in traps, killing the insects in the process. Recording sounds from insects, however, is a non intrusive measure for gathering data, and through my specialization project [5] evidence was found that it might be possible to use machine learning to distinguish the species by their flight sounds.

The flight sounds, or buzzing, of the insects produce a quite specific frequency. This fundamental frequency changed dependant on the size of the insect and the wing span. In figure 1.1, the spectrogram made from the flight sounds of a bee is shown. The dark orange line seen around 256 Hz on the y-axis showes the fundamental frequency of this bee and the orange line around 512Hz shows the first harmonic overtone. This fundamental frequency and harmonics can be extracted from the audio and be used to identify the insect.

Looking at recordings from different insects within the order Hymenoptera, we found spesific frequency ranges for each insect. In figure 1.2, 2454 samples of bees, bumblebees, wasps

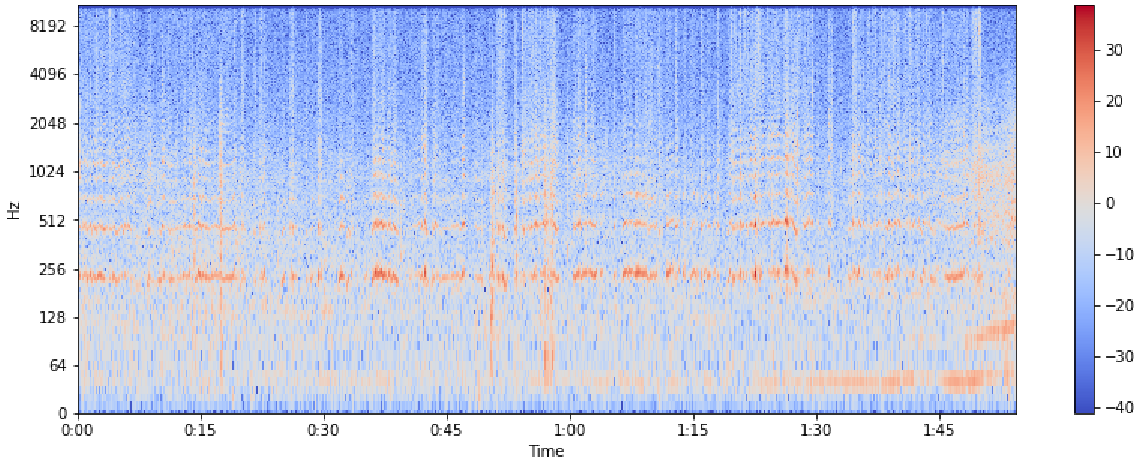


Figure 1.1: Spectrogram of recorded flight sound of a bee from [5].

and simulated hornet sounds were processed to find the fundamental frequency. This statistical data plot shows the frequency range of the insect sounds recorded. Finding that the fundamental frequency within the order Hymenoptera explored here resides between 70 Hz to 310 Hz.

Knowing the frequency range of the insects gives insight into which frequencies that are important when optimizing a system for acquiring sounds from the order Hymenoptera.

1.1 Background and motivation

Through the specialization project [5], at Norwegian University of Science and technology, NTNU, recordings bees, bumblebees and wasps were acquired manually by using a microphone in the field. The hornet sounds was acquired by augmenting the wasps flight sounds. An important aspect of data used in machine learning is that it is labeled and verified for the algorithm to train and test based on this information. Thus the data has to be of high quality securing both audio and visual confirmation.

Working with acquiring the sounds from insects within the order Hymenoptera, I found that the insects did not stay in the same area for long. Capturing images before starting the recording was not effective because the insects flew away before I could start recording. The achieved recordings were collected from the insect that were visually verified and followed

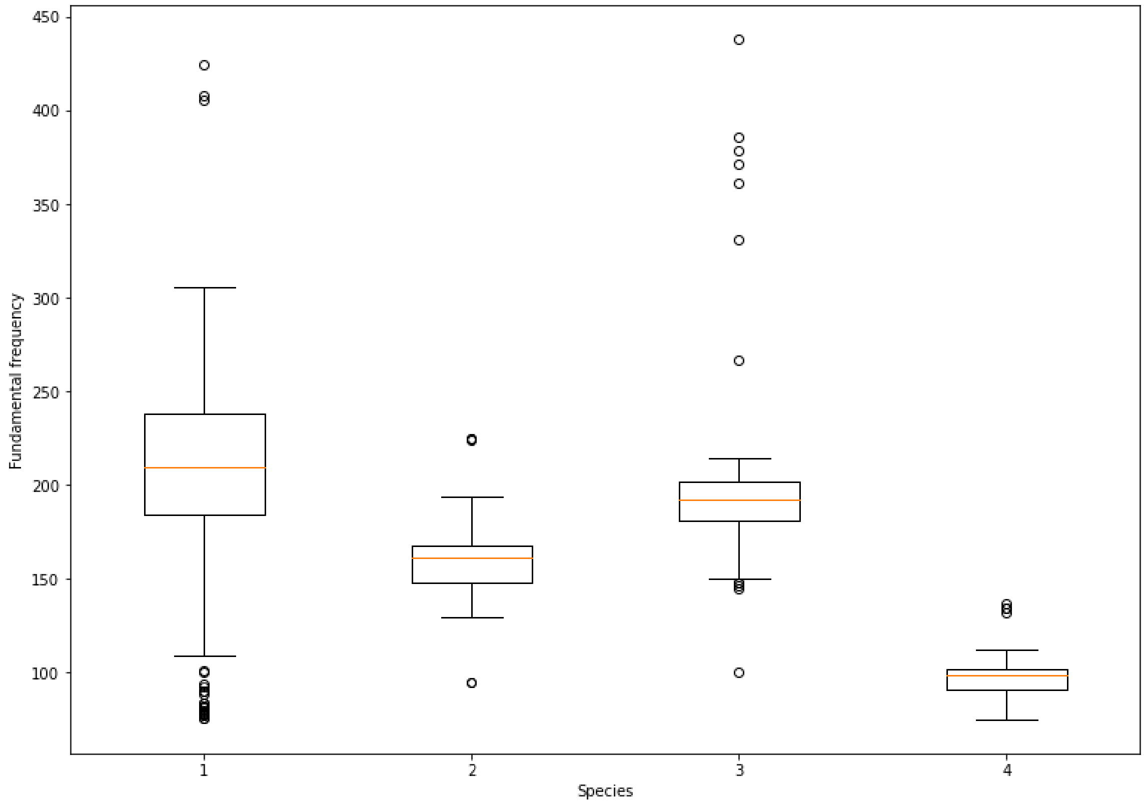


Figure 1.2: Frequency range of flight sounds for 1 (bees), 2 (bumblebees), 3 (wasps) and 4 (simulated hornet sounds) from [5]. The orange line showing the mean value, each square box show the standard deviation, whiskers are the 25- and 75 percentile and the circles show the outliers in the data set.

with the microphone. With this process, the data set was not diverse with regard to different species and locations. The conclusion was that the method of insect verification was promising, but with the caveat that the data was recorded in the same area and with the presence of the same background noises.

With the challenges explained above experienced in the specialization project, the object of the Master's thesis was derived. We need to design a system that is able to capture both sound and images. This is necessary to be able to verify the different species without special knowledge of each species and their families at the time of recording.

Three of the vital pollinators, referred to in the earlier section, play a huge part in the ecosystem and contributes positively to agriculture around the world. Following the results from the specialization project, [5], the K-nearest neighbors (KNN) algorithm resulted in the best accuracy with a weighted average of precision and recall to 98% (0.98) of the test data. This means that 98% of the test set was correctly classified. Using this as a start, the future possibilities in practical use of machine learning look promising. The expansion of machine learning and its potential applications have exploded [6]. The bottleneck is the amount of quality data available to train and test such a system. Creating a low cost recording device for the insects in the order Hymenoptera is vital for further work in cataloging the insects and number of individuals. The system will be designed for Hymentoptera, but is not limited to them as a target.

1.2 The missing acquisition system

The need for building a large database for use in machine learning is growing. Audio and images is ideal to use in machine learning. Understanding insects and animals through sound is widespread, [7], [8] and [9]. Equipment for recording audio, as well as for capturing images is available. A setup that can capture images and record audio when an insect is present and operate for a long time outside, is not known by the author.

There are different possibilities for simultaneous recording of audio and images. One possibility can be smartphones. Most people own a smartphone with a high quality camera and microphones that can record stereo audio and video at the same time. The quality of smartphone images is more than adequate for identifying and verifying the species for labeling the data. The microphones on the other hand do not produce audio of high quality. External microphones can be fitted to the smartphone, but the smartphone then uses a lot of power. Another possibility can be to use an ambisonic setup. This is a surround sound format for full-sphere audio microphone and a 360° camera. We will still need to trigger both devices at the same time and is currently expensive to acquire. It is interesting to find a low cost solution that solves this challenge.

When instigating possible solutions it was clear that the field of machine learning based on the flight sounds of insects is not fully explored. When researching the topic for the specialization project [5], I found the work presented below by other researchers.

The paper by Kawakita et al. [10] investigates the use of flight sounds to classify insects with machine learning. They look at three bee species and one species of hornets from Japan. This was done using the fundamental frequency and the mel-frequency cepstral coefficients, MFCC, as features and classify the insects from the background noise with a support vector machine, SVM. The promising results from this study gives a possible correlation to Norwegian species of bees, bumblebees, wasps and hornets. Another researcher, Parmezan et al. [11], explored changes in wing-beat frequency of bees and wasps using optical sensors to look at the flight sounds of Hymenoptera. And Gradisek et al. [12] looks at different bumblebee species to determine if using the acoustic features was enough to distinguish the 12 different species they looked at. They use audio features and the random forest algorithm as the best training setup.

To this date, and to the authors best knowledge, an acquisition system that is low cost, with high quality audio, camera and which is expandable does not exist. Creating this system should be further explored.

1.3 Aim of the project and research questions

The objective of this project is designing a data acquisition system for recording high quality multichannel audio and capturing images of the subject for identification. The system will be built as a low cost embedded system with the capabilities for improvement in the future. Building prototypes of hardware and testing the capabilities of the chosen microcontroller is the main intention. The research question for this Master's thesis then becomes:

Is it possible to create a low cost embedded acquisition system that records multichannel audio of high resolution, and with the capability of capturing images?

1.4 Outline of report

This thesis is divided into six sections including this introduction. The report starts with section 2, the specification and design decisions of the acquisition system. The selected technologies, components and systems are presented. In section 3, the implementation of the acquisition system is explained in detail. Each part of the system, from the hardware to the software design, are presented in subsections. In section 4, the evaluation of the recording system is described. Results from the tests of hardware and software are presented. The second to last section 5, is the discussion of capabilities and restrains of the implemented acquisition system. This section contains a necessary discussion regarding the results gathered and on the technology and the system as a whole. The section also contains information on further improvements, lessons learned and the direction for future exploration. The final section 6 is the conclusion with the final remarks of the project.

The report is written with the intended reader of a master student studying embedded software in mind. The report also assumes some background knowledge in acoustics and electronics.

Chapter 2

Specification of acquisition system

The project starts with the specification of the acquisition system and the selected approach. The following section explores the specification of building the acquisition system and will cover the system for both functional and non-functional requirements. This section will also look at the overall system with the technical design and overview of the product. Finally, each contributing part will be described in detail with the decisions made for the technical design.

2.1 Requirements

The requirements describe the necessary functionality the system must have to function as intended. The functionality demands listed under is mainly derived from experience gained while recording insects for the specialization project and from discussions with the supervisor of the project. The list also includes properties the system should have, but are not strictly necessary for function, for a possibility of expanding the project in the future.

The system **must**:

- be able to store 4-channel audio from embedded microphones and must be able to store images.
- have embedded low cost microphones with low noise and high sensitivity.

- have a microphone circuit for each the four microphones and must have a strong amplification of at least 20 dB static gain.
- camera for verification must be available to connect to the system.
- have a wired controller to start and stop the recordings.
- have 1st order ambisonic 360 ° audio recording capabilities.

The system **should**:

- have the option to run on a power supply for 7 days, without access to mains.
- have a possibility for a software controlled gain, to adjust the gain depending on the subject recorded.
- contain memory storage with capacity of 7 operating days.
- be energy efficient and have sleep modes for saving power and allow for scheduling.
- processor strong enough to enable direction detection.
- be able to listen to the audio during, before or after the system is recording to check the state of the audio.
- give the user a possibility to listen to the last recording, thereby confirming the quality.
- have processing capabilities for highpass- and lowpass filter, post processing and root mean square (RMS) processing for triggering purposes.

2.2 Technical design

Designing the acquisition system includes a collaboration between the hardware and the software used. The embedded system contains multiple parts to operate according to the specification.

To get a better overview of the total system, a high-level block diagram is shown in figure 2.1.

The envisioned system have a central processor controlling the system with input from 4 channel audio for recording, pre-amplifier stage to amplify the audio, analog-to-digital converter (ADC) to convert the analog audio signal to a digital signal, input from a camera for verification, input for buttons for controls, input and output data storage for saving images and photos, and program memory for the scripts.

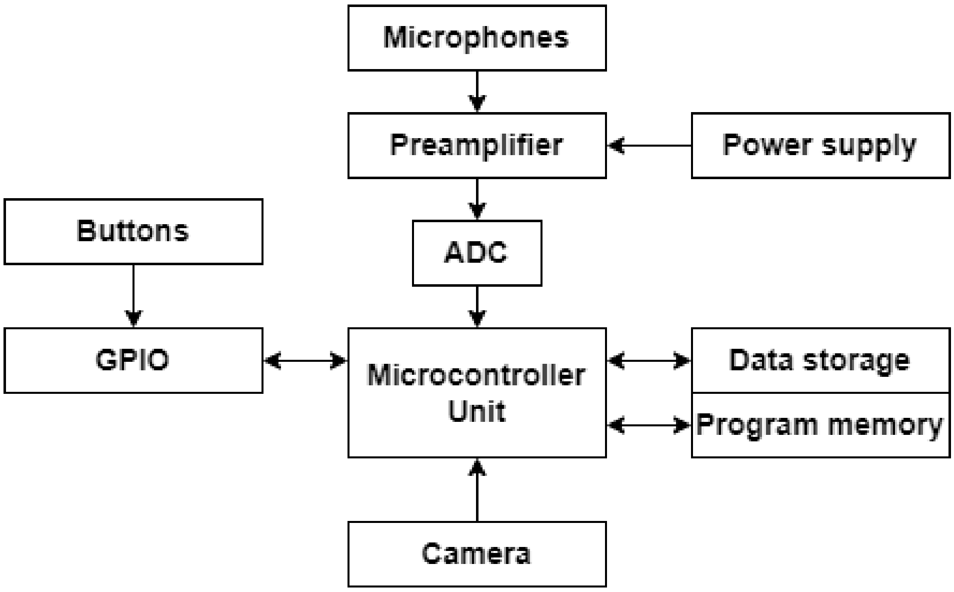


Figure 2.1: Block diagram for proposed system.

2.3 Design decisions

Following the requirements, evaluating and choosing the technologies and components that make up the design will be explored in this section.

2.3.1 Microcontroller

The microcontroller for the design is important for controlling the system, and its capacity and utilities can be engineered according to the needed system properties. When selecting a processing unit, the requirements listed above will be used to choose the right processor for the job. Working with the requirements of processing power, the system should have the capacity to operate the camera and take images, record 4 channel audio and save the acquired data to a data storage on device, SD card or external hard drive.

The system is intended to be placed outside for recording purposes, and for future endeavors, the system should have low energy consumption and be able to be scheduled with sleep modes and needs to be able to run off a portable power supply. In this feasibility study, the weather proofing of the system will not be considered. For running multiple functions at the same time, like the recording of audio and images at the same time, the system needs to have multi core operation. For acquiring the data, the system must have I/O for recording the data. The ports can be used to get the data and shields to extend the boards functionality or built in capabilities and extension boards. Shield or HAT (Hardware Attached in Top) adds functionality to a microcontroller with additional features not present on the device.

The desired solution is a low cost system that can be programmed to execute the operations laid out in the requirement section. The system will need to be easily programmable and have an extendable base for the rest of the system, an embedded system with possibilities. The support of a platform is important, both for quality documentation and examples, but also for the necessary forums and active users. Having many active users exploring the platform with projects results in finding issues and reporting problems or solutions.

Table 2.1: Overview of potential platforms.

Feature	Arduino Uno	Sony Spresense	Raspberry Pi 3B
Processor	AVR	ARM M4F	ARM A53
Bits	8	32	64
N of cores	1	6	4
Memory	2kB	1.5MB	1GB
Audio Inputs	Shield	4 analog, up to 8 digital	Stereo
Audio quality	Shield	High	Low
Camera	Shield	5MP	8MP
Support	High	Low	High
SD Card	Shield	On-Board(Extension)	On-Board
Operating power	50mW	100mW	2400mW
GNSS	Shield	GPS	Shield

For this project, three contenders were chosen. First, the Arduino Uno was evaluated. This is a well established microcontroller unit with a lot of support and projects. The Arduino is easy to use and has an array of shields to add functionality to the board. The Sony Spresense was introduced as an alternative with specifications that match the requirements of the task. The

Raspberry Pi 3B is an older version, but is capable with the processor capacity. The raspberry Pi has, like the Arduino, been around for many years.

In table 2.1, some selected features deemed important from each board are presented. The overview is collected from [13] section 3, and selected with the features in mind.

The first aspect to consider is the processor and cores. The Arduino only has one core, the Spresense has a stronger processor with six cores and Rasberry has a more powerful processor with four cores. In regards to audio captured, high quality audio is key for future use in datasets for machine learning. Knowing that the sample rate needs to be high for the high quality audio, it is not possible to take images and record audio with timing intervals on one core and therefore, the Arduino is not a good alternative. For the audio on the two other contenders, the Spresense has a built in 4 channel analog high quality recording, while the Raspberry has low quality stereo. Both systems have camera extensions for ease of use and options with SD-card slots for data storage. The power consumption while operating normally of the Raspberry is 2400 mW, that is 24 times the power consumption of the Spresense. The Spresense was launched three and half years ago, but can't compete with the support for the more established platform of Raspberry.

The Sony Spresense does not have the same support as the Raspberry community and is a down side of the system. This might result in a slower development and feedback through forums. Less information and earlier projects to work from is not an advantage and the platform is new to the author of this report. The support for developing with the Arduino IDE and a more advanced option with Sony's Spresense SDK for low level programming, gives the user options for programming the device.

When looking at the alternatives, the Sony Spresense showed the most promise for the requirements set for operation above in this section and was chosen for the implementation. The convincing arguments for the Spresense was the superior audio features, extension boards and powerful processor.

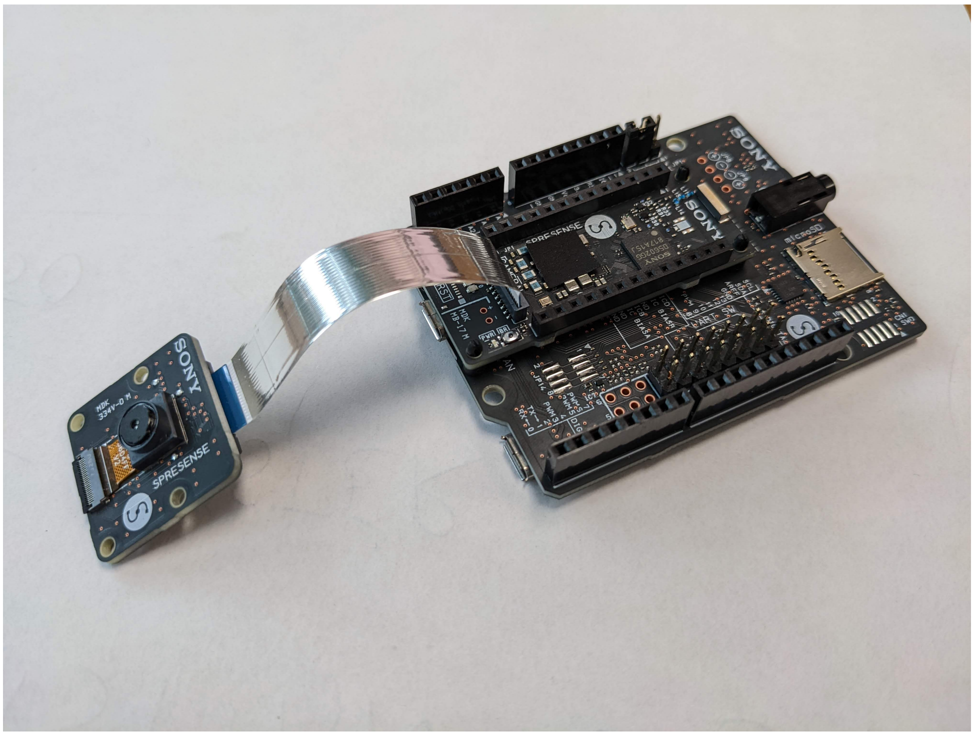


Figure 2.2: Sony Spresense with extension and camera board.

2.3.2 Data acquisition on Spresense

The data acquisition is an important part of the system and in this subsection the hardware on the Spresense, extension- and camera board are explored. The technology, capabilities and function will be explored below.

Audio

The Spresense has four high resolution analog inputs on the extension board. They are placed on the extension board with three pins per channel; the microphone input, the 2 V bias and ground. There is one bias for mic A and B and one bias for mic C and D. It is recommended to use the ground of the pins in JP10 for low noise operation.

The Spresense uses the CXD5247GF, Spresense's power supply system management and audio integrated circuit (IC), with a built in audio block called CXD5602 with support for 4 channels, an AD converter, amplification and a SNR 90 dB.

The system makes it possible to record in different formats, both mp3 and WAV are suppor-

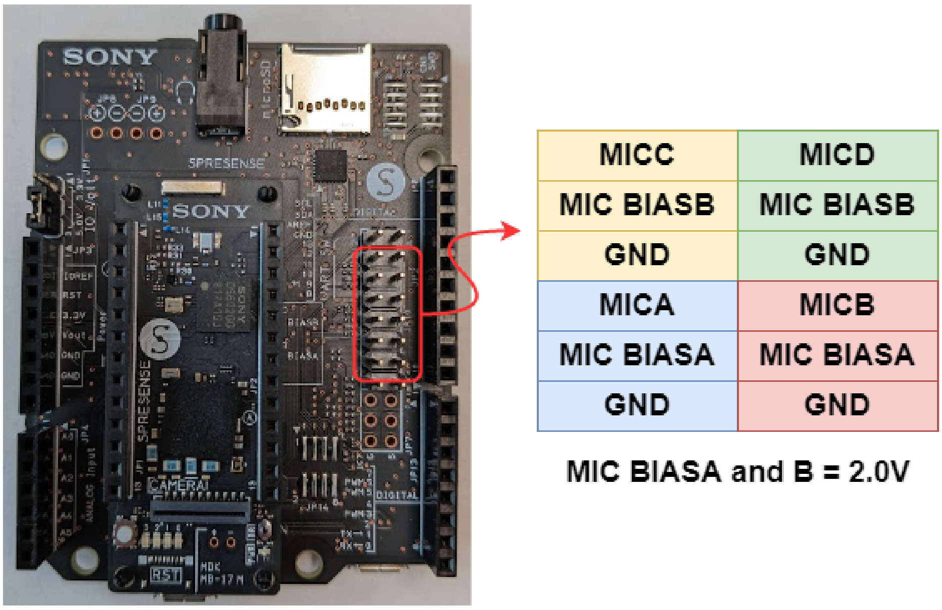


Figure 2.3: Microphone input on the Sony Spresense extension board, inspired by [13], figure 6.

ted. Recording can be done with sampling rates up to 192 kHz, 48 kHz and a 16 kHz option, with 16 or 24 audio bit depth. It is set up for 1, 2 or 4 channel recording in Arduino.

On the input of the audio pins, pull-up resistors are in place to receive input signals, [14], with a range of 0-1.8 V on the input before clipping. Matching the input signal from the pre-amplifier to force the voltage level to reside around 0.9 V with an amplitude less than 0.9 V to eliminate clipping.

Camera

The camera board is a separate extension that can easily be attached by the ribbon cable provided in the package. The camera board is connected to the main board through the dedicated camera slot. The camera board is shown in figure 2.4.

The camera can record images at 5 MP. The camera can record JPEG images and with the formats Y/C, RGB and RAW. For video, the camera can record 1080p with 30 frames, video recording in JPEG format. The lens is a normal 78 ±3 degree field of view with an aperture of 2.0 ± 5% and fixed focus and depth of field from 77.5 cm to ∞.

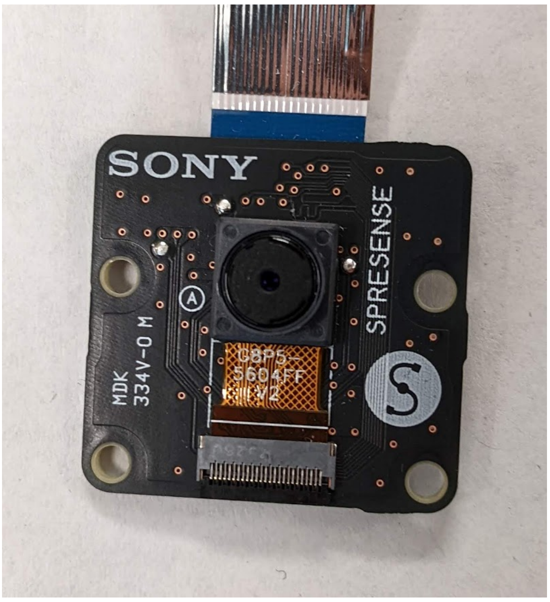


Figure 2.4: Sony Spresense with extension and camera board.

2.3.3 Amplification circuit and microphones

Amplification is a crucial step in recording insects with low sound pressure emitted from their flight sounds. To be able to create a circuit to enhance the signal from inexpensive embedded microphones. The circuit must contain circuitry to power the microphone, to filter and amplify the signal coming from the microphones. The type of analog microphone technology working with this system will be explored next.

A single microphone operating omnidirectional can record 360° at the low frequencies produced by the flight sounds. But to achieve directionality, an advanced microphone setup will be needed and placed in an arrangement that makes it possible to record sound in 360°. Using multiple microphones, the arrangement of the microphones can be further used for focusing on the insect as the source and filter out the background noise.

There are three common microphones are : MEMS, Electret and Condenser. The two specially considered for the project are MEMS and Electret. Larger condenser microphones might deliver better sound, but will be more expensive and therefore not right for this task. The MEMS technology is more used than ever due to its size and use in smart applications. Both Electret and MEMS microphones are available with weather proof systems, but are more expensive than the standard counterparts. The main reason for selecting Electret over

MEMS is the natural dust resistance with the mesh cover and larger size makes for faster prototyping and soldering by hand.

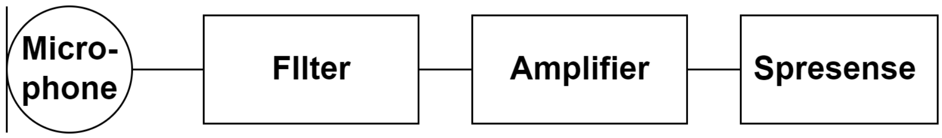


Figure 2.5: Circuit design for pre-amplifier.

There are many different ways to amplify a signal. For large amplification an inverting amplifier is developed using an operational amplifier with a negative feedback. Inverting the signal does not change the signal, only inverting the amplitudes and amplifying at the same time. The resistors used in the design define the size of the amplification, and the resistors together with capacitors are used for the highpass and lowpass filters. For this design, the amplifier used is the NE5532AP, a dual low noise operation amplifier for use in audio.

The circuit will need to be produced on a PCB to eliminate noise that can be generated when testing on a breadboard. The prototype of the circuit would need to be produced and tested to see the potential of such a system. Using the Electronics and Prototype Laboratories as a convenient place to start for production of the PCB. Using the local alternative, the production time was less than a week and no time spent waiting for delivery. When looking for a program to design the PCB, the first program evaluated was Altium Designer with a powerful set of features. The other alternative for designing was KiCad EDA, an open-source electronics design tool. Finding KiCad an easier program to learn and design the intended design, it was chosen as the PCB and schematic program for the Master thesis. KiCad will be used to generate a bill of materials and was used for exporting Gerber files for printing the PCB.

2.3.4 Power supply

The source of power is important to consider in order to reduce the amount of noise introduced into the recordings. Powering the microphones and amplifier with a single power source means that the power source needs to be filtered for noise. The power source also needs to be convenient and portable. In this project the choice for powering the circuit resulted in a 9 V battery which is affordable, has enough power to run the circuit and is easily available.

The power circuit is designed with a linear voltage regulator to obtain the 5 V needed for the circuit. Even with the loss to heat, the linear voltage regulators produce less noise than switching regulators. The linear voltage regulator used in this work is the L7805AC. The battery is connected with two capacitors, the first to keep the charge from the battery and the second to reduce noise between the battery and the regulator. On the 5 V output on the regulator, another capacitor is placed between 5 V and ground to reduce noise from the regulator. The schematic is shown in figure 2.6.

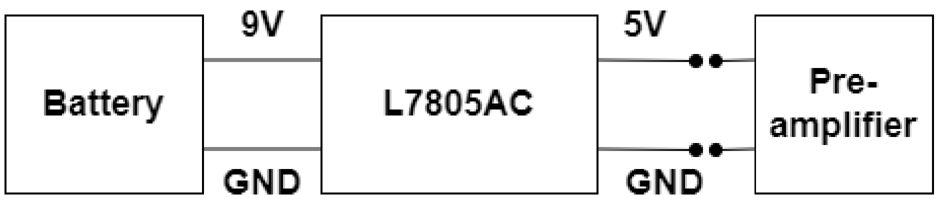


Figure 2.6: Power supply circuit functionality.

2.3.5 Button design

The button layout is present mainly to control the system remotely and test the deployment of the system. The buttons are intended to start and stop data acquisition, using two push buttons not mounted on the Sony Spresense to be able to control the system from a suitable distance. To make sure the push of a button is not affected by multiple accidental pushes, a debounce RC circuit or a software implementation is needed to eliminate oscillations when pushing the button. A RC circuit was chosen to solve this problem and the RC circuit will charge the capacitor over a short amount of time, reducing the oscillations. A potential button design as a block diagram shown in figure 2.7.

The buttons are connected to the Spresense using the digital I/O pins on the extension board and powering the circuit with the Sony Spresense, 5 V from the I/O pins and ground. Two ways to implement the button signal were explored, interrupt and polling the input signal.

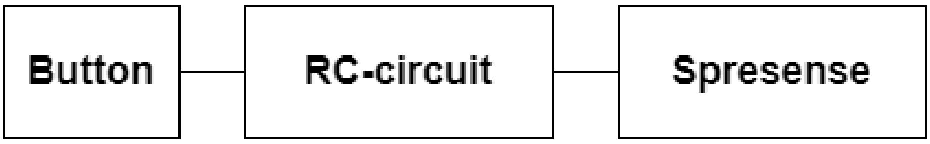


Figure 2.7: Button with RC-circuit for debounce.

Both solutions will work for this task, polling is less effective, but easier to implement while interrupt is a more convenient and robust solution for future work with scheduling and sleep. This design is implemented with polling for simplicity.

2.3.6 Microphone array

The locations and directions of the microphones are important aspects of the system. Knowing that the insects move constantly, a system for recording 360° is a useful feature. Arranging the microphones in a 1st order ambisonic setup is possible with four microphones to obtain coverage for audio from all directions. The ambisonics is used in VR-games and record audio in multiple directions and use this data to detect the direction of the audio. Using this to identify the direction of the sound, might be used for detecting the direction of the insects and possibly enhance the signal by tracking the insects in post production.

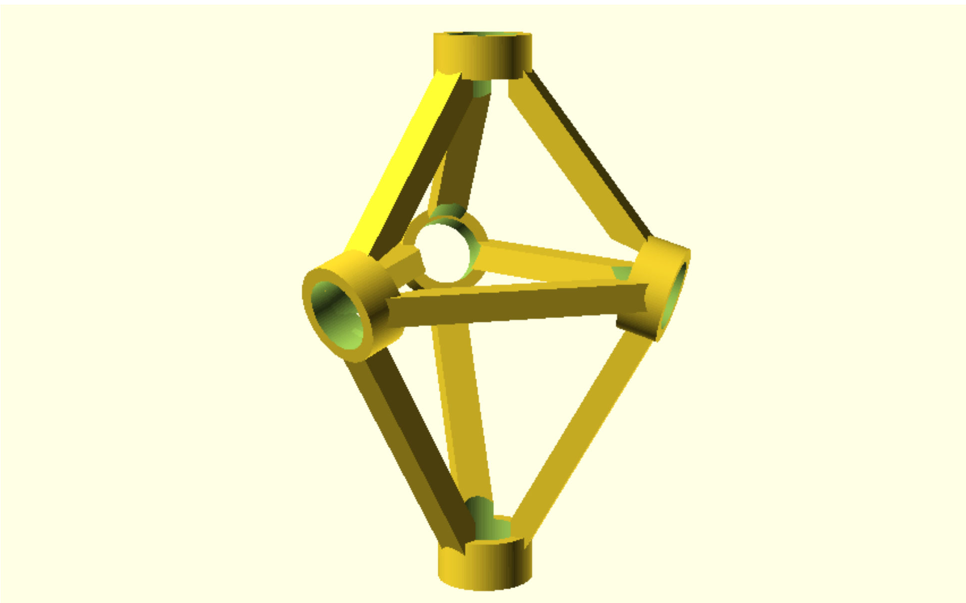


Figure 2.8: Possible ambisonic 3D design.

The 1st order ambisonic setup is shaped like a tetrahedron, with the microphones placed on the outer corners, at the top and three around the sides. The circle at the bottom of the design is used to feed the cables from the microphones to the pre-amplification circuit. Resulting in four microphones aligned in each direction. This is important for the conversion from A-format to B-format, [15], converting the recordings from the microphones to audio in A-format to a 360° sound field in B-format in post production to get the directional data. The design of the microphone holder was 3D printed, which is a fast and convenient method for creating odd shapes.

Designing the ambisonic mount was done in OpenSCAD, a script based solid 3D design tool 2.8. The reason for choosing a scripted tool over the traditional moulding and shapes was the ease of use, quick introduction and approachable for someone with programming experience. The design is as simple as possible to remove unwanted difficulties in designing and printing. The design will mainly use common shapes in the design process, cylinders for the mounting points and a wire mount, and rectangles for the beams holding the structure together.

2.4 Future proofing the system

The **should** requirements earlier in this section are presented based on the intended function of a complete recording system. The entire system of functionality was not possible to develop in the time allotted to this Master's thesis, but are included to ensure that the tests and developments of this project can be expanded upon so that the system can handle additional features for improving the design at a later stage.

Chapter 3

Implementation of the acquisition system

The implementation section follows the design decisions made based on the specifications described in the previous chapter and contains information on the implementation of the recording system. The section covers the setup and configuration of the Sony Spresense used with the Arduino IDE for prototyping, package- and library installations. The section also covers the design process for designing the hardware, including the design of the amplification circuit, power circuit and button layout. Also the design and 3D-modeling of the ambisonic setup are covered in the section. The final part of the section is dedicated to software design. Looking at the use of the Arduino IDE for memory allocation, code and scripts implemented to test and run the recording system.

3.1 Setup and Configuration

The Sony Spresense board is created with a modular design, and the module used from the Spresense lineup consists of the following boards:

- Main board
- Extension board
- Camera board

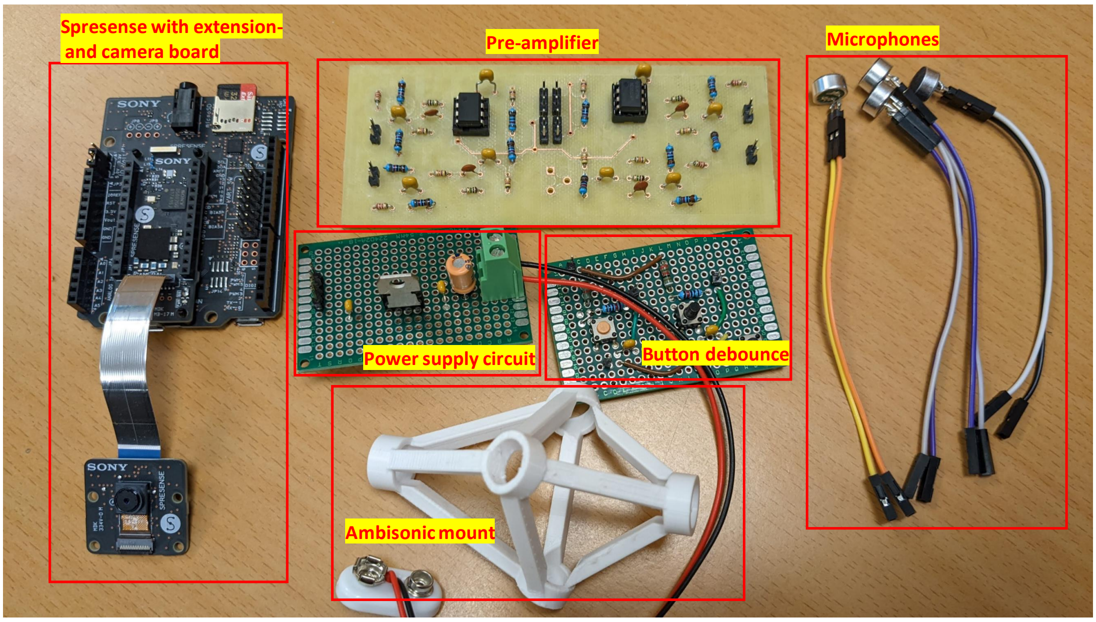


Figure 3.1: A complete acquisition system with all of the parts. The red boxes name each part of the system implemented in this section.

Regarding the specifications, it is desired to use the Main board for the computations, the extension board for access to the audio inputs, GPIO and SD-card slot, and the camera board for capturing images for verification.

The main board is connected to the extension board with a Board-to-Board connector, B-2-B, and attached with the mini-spacers, AS-2003, that comes with the board. The camera board is attached with the camera connector on the main board. The next part of the setup was placing a shading seal on the IC4 on Spresense main board, with Shurtape, CP-743, provided in the packaging, for damage protection in direct light.

3.1.1 Spresense Arduino Library

The setup and installation for the Spresense on Arduino follows the Spresense Arduino Library Getting Started Guide [16], for more information provided by Sony Spresense. In short, the following steps were followed before setting up the environment.

- Download and install Arduino IDE and verified supported operating system.
- Download and install necessary driver for the USB-to-Serial.

- In the Arduino IDE, install SprArdLib using board manager and install Spresense Refrence Board.
- Spresense attached with USB to the computer, the Arduino IDE is used to find the port of the Spresense and install the bootloader on the main board.

With the steps above completed, the board was ready to use with the installed packages.

The Spresense Arduino library contains board packages for the pre-built Spresense SDK with the Spresense library wrapper for Arduino. The package includes standard Arduino libraries, specific libraries for Spresense, example projects for using the camera and audio recording to mention some of the included support. The board package installed also includes tools for binary generation and a flash writer. The NuttX Kernel, Spresense BSP and ARM GCC for building the toolchain for the Arduino sketches, are also included.

When using the audio functionality on the Spresense DSPs (digital signal processing) for audio encoding and decoding was installed to the SD-card to process the audio in and out from the Spresense. To record in the "lossless" audio format WAV the dedicated WAV decoder was installed to the SD-card through the DSP installer.

Programming libraries

For arduino on the spresense: `include <Arduino.h> include <arch/board/board.h> include <SDHCI.h> include <File.h> include <Audio.h> include <stdio.h> include <Camera.h>`

Python for spectrums and spectrograms: `import numpy as np import librosa, librosa.display import matplotlib.pyplot as plt`

3.2 Soldering

The soldering performed in this thesis work on the pre-amplification circuit, power supply and buttons was done at the Research and Development laboratories at NTNU with their soldering station and equipment. The equipment used is listed below in table 3.1 from the station with soldering iron, fan and solder with flux.

Table 3.1: Soldering station equipment.

Equipment	Brand	Model	Serial number
Soldering Iron	Metcal soldering system	MX-PS5000	027106
Soldering Fan	Oki Fume Extraction	BVX-200	221806
Solder	FELDER	0.75 mm ISO-Core "EL"	N/A

3.3 Power supply to the preamplifier

The power supply was built with an inexpensive, small and easily available 9 V battery, ANSMANN Alkaline, 550 mAh. The schematic of the power supply circuit is shown in figure 3.2.

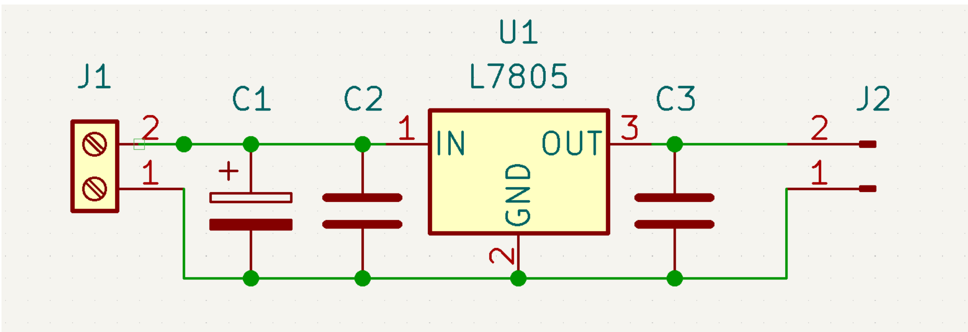


Figure 3.2: Power supply circuit.

The circuit is designed for a 5 V power supply and therefore the power from the battery needed to be limited to 5 V from the provided 9 V. The chosen linear fixed voltage regulator, L7805AC, will step the voltage down to the intended 5 V. The circuit also needs to limit the noise introduced in to the pre-amplifier, since the microphones are powered directly from the power supply. Introducing noise from the power supply, directly affect the recorded audio. First, maintaining a stable output voltage was done by placing an electrolytic capacitor of 22 μF between 9 V power and ground, to store and discharge voltage when the battery is connected through wires and not directly connected to the power supply circuit. A second capacitor of 0.1 μF is also placed between 9 V power and ground a to remove incoming noise into the regulator. The third capacitor, another 0.1 μF capacitor is placed on the output of the regulator to further remove noise out of the system, also connected from 5 V to ground.

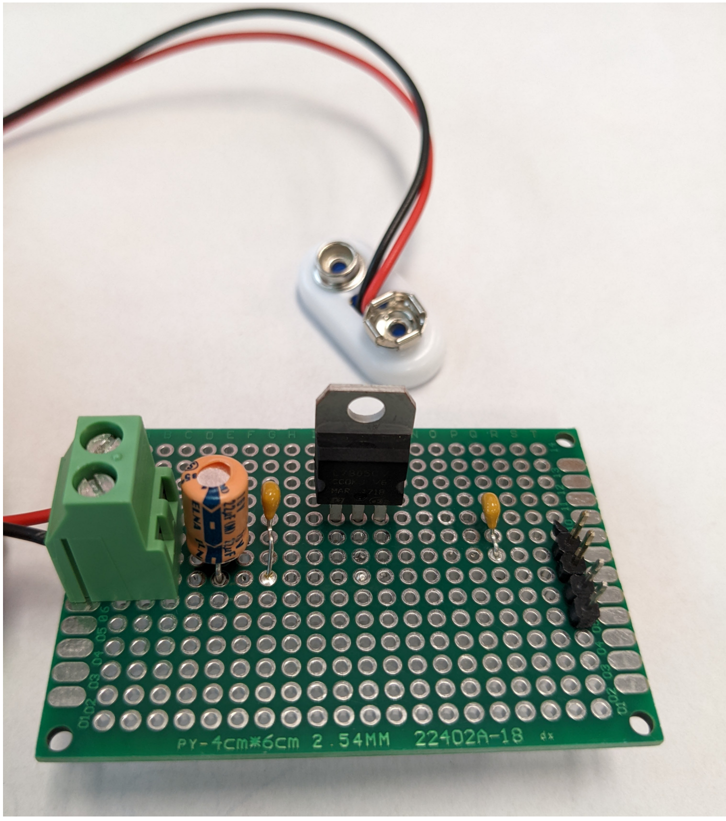


Figure 3.3: Power supply for the preamplifier implemented on perfboard.

The design was implemented onto a perfboard as a prototype. The components were soldered to the perfboard in accordance with the schematic shown in figure 3.2. Using perfboard to create the prototype was done to remove noise, by soldering the components instead of prototyping on breadboard where the connections might be loose.

3.4 Microphone and pre-amplification circuit design

The power supply circuit just discussed is designed for powering the microphones and an inverting amplifier. The schematic of the design is shown in figure 3.4. The low sound pressure emitted by the Hymenoptera discovered in the specialization project [5], resulted in the need of a dedicated amplification circuit. The Spresense has a built in amplifier controlled in software, but for amplifying the signal generated from the Hymenoptera the use of an external analog amplifier was necessary. We used a low noise operational amplifiers to design an inverting operational amplifier. The amplification on the Sony Spresense can be adjusted later if additional amplification is needed. The included amplifier located on the Spresense is not well documented [14]. The design will feature four sets of amplification circuits created in KiCad, A Cross Platform and Open Source Electronics Design Automation Suite, which are shown in figure 3.4. This equals one amplifier for each microphone in the system, the entire schematic is located in appendix B in figure B.1.

Creating the layout and choosing the component for the pre-amplifier is based inverting amplifier circuits, high and lowpass filters and voltage dividers. The actual components used were not added to the schematic in KiCad since the board is soldered by hand and not added in the production. For this reason the prototype board is created with through hole technology for simplifying the soldering process. For production, the types of components, capacitors, resistors, header pins and op-amps are also added. The full list of components and component values are added in Appendix B with a Bill of materials, BOM.

The circuit is powered by a 5 V power supply. The microphones have a pull up resistor R1 of 2.2 kΩ and operate at 4.5 V. The next part is a highpass filter to eliminate the DC signal from the microphones with the C1 capacitor 1 μF and R2 resistor 47 kΩ. The filter calculations are found in equation 3.1.

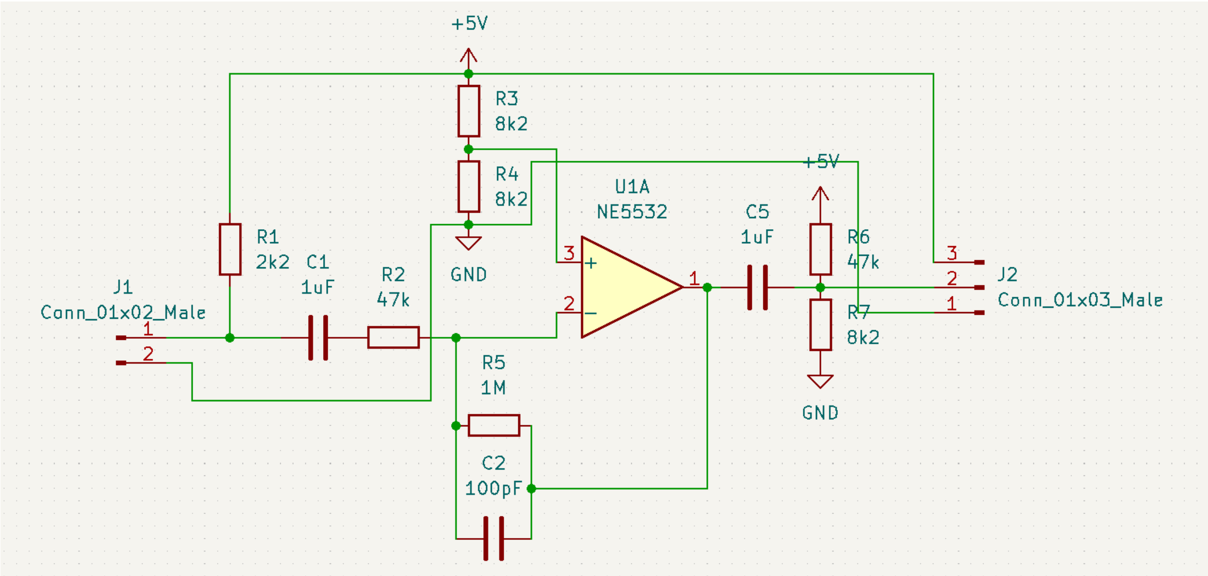


Figure 3.4: Schematic of the pre-amplifier design.

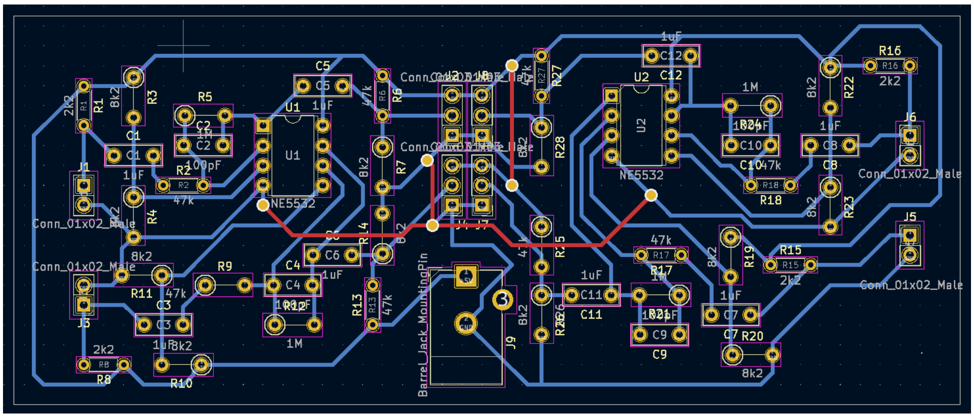


Figure 3.5: PCB of the circuit design.

$$HP = \frac{1}{2\pi \cdot 1\mu F \cdot 47k\Omega} = 3.386Hz \tag{3.1}$$

The lowpass filter is the C2 capacitor 100 pF and R5 resistor 1 MΩ under the OpAmp and creates a cuttoff at ∼16 kHz calculated from equation 3.2. Using a lowpass filter at ∼16 kHz was implemented to remove high frequency noise, while the Hymenoptera has a frequency range that is noticeably lower, the acquisition system is designed to be used for other purposes if desired.

$$LP = \frac{1}{2\pi \cdot 10pF \cdot 1M\Omega} = 15.915kHz \tag{3.2}$$

With the inverting amplification design, capacitors were easily added for gaining filters to clean the audio before recording without removing information. from . From the specialization project [5], post processing of the audio was needed to select the right frequencies to use as features and lowpass filter was placed at 2500 Hz.

The pre-amplifier design uses the NE5532AP, a low noise OpAmp often used in audio applications with two amplification circuits per chip. Using a voltage divider with the R3 and R4 resistors in figure 3.4 to divide the 5 V supply to feed 2.5 V to the positive input contact of the OpAmp. Placing the DC offset in the OpAmp to 2.5 V with the possibility to amplify to an amplitude of 2.5 V, not to clip the output while amplifying. The inverting setup uses the resistors from HP and LP filter to amplify the signal in the feedback loop. From equation 3.3, the amplification formula is presented with a gain of 21 dB. On the output, another capacitor is placed to take away the DC offset and then another voltage divider, with resistors R6 47 kΩ and R7 8.2 kΩ, is placed before the output to match the input bias of the Spresense board.

The amplification of the system is using the resistor in the feedback loop and the input resistor to control the gain of the system. In the equation below the resulting value is the factor of amplification in the circuit.

$$Amplification = \frac{R_f}{R_{in}} = \frac{1M\Omega}{47k\Omega} = 21.28 \tag{3.3}$$

After creating the schematic of the design, the PCB was created in KiCad. The board is designed after the specification from the Electronics and prototype laboratories at NTNU, [17], they are producing prototype PCBs for projects at NTNU. Producing vias, which are holes with wire through the PCB, according to a list of acceptable dimensions of the vias they can produce. The PCB was designed with vias of 1.1 mm inner diameter and 1.6 mm outer diameter. The tracks were also set to a minimum recommended by the electronics and prototype laboratories at 0.5 mm tracks. The vias are placed manually on the prototype, which led to focusing on reducing the number of vias on the board, [17]. This affected the layout of the PCB explained in the next section, which is not optimal for a final product. The PCB layout is shown in figure 3.6.

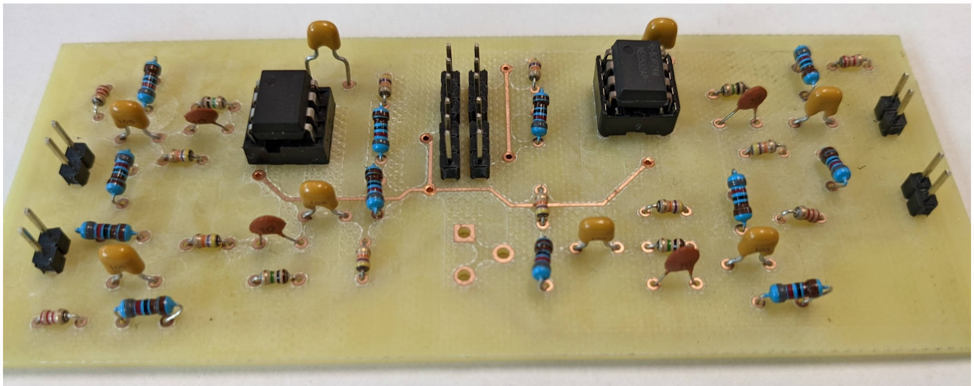


Figure 3.6: Finished PCB with soldered components.

Header pin placement

In an effort to reduce the amount of vias through the board, the pin placements on the board changed from the setup on the Spresense and the microphone inputs. When looking at the twelve central pins, the headers are divided into four sections, one for each channel with the 5 V at the top most position, audio signal in in the middle and ground at the lowest pin. On the left and right side of the PCB, the microphone is connected to 5 V and ground, with 5 V above and ground at the bottom. The orientation of the operational amplifier on each side of the PCB meant that in order to reduce the amount of vias, MICA microphone input is switched on it's head compared to the remaining three microphone inputs, giving MICA ground above and 5 V at the bottom. Figure 3.7 shows the configuration.

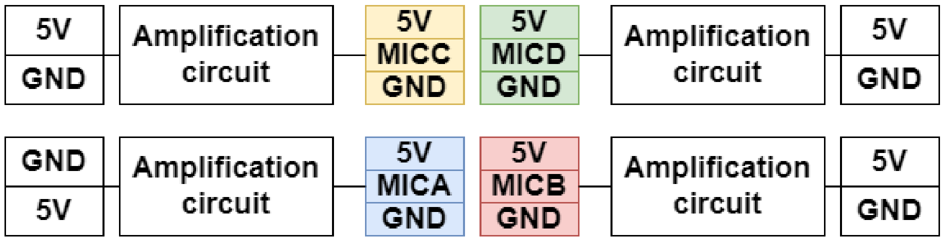


Figure 3.7: Pin header placement on PCB.

3.4.1 Microphones for embedded system

The embedded microphones were chosen for low-cost, low-noise and high sensitivity. Looking at Electret microphones, the microphone selected was the CMA-4544PF-W. The CMA-4544PF-W is a low-cost microphone with frequency range of 20 Hz - 20 kHz, and sensitivity of -44 dB ± 2 dB and SNR of 60 dB.

3.5 Ambisonic mount

Based on the selected microphones, the ambisonic structure was sculpted in OpenSCAD. The script-based CAD software was efficient and easy to learn. The sides between the microphone mounts are 4 cm long and the microphone holes are 10.2 mm in diameter. The three holes on the sides and the hole on top of the structure are mounting holes for the microphone and the hole on the bottom is for cable management and mounting. The microphones are placed in a pyramid structure in order to record in all directions. The 1st order ambisonic mount was 3D-printed in PLA on a Prusa Mini+ and the printed design is shown below, in figure 3.8.

To capture true 360 ° audio, the microphones need to be as close to each other as possible in the tetrahedron shape. For the task of detecting the direction of the insects, the distance is not as important.

Figure 3.8: The 3D printed ambisonic mount.

3.5.1 Button integration with debouncing

The button integration is built as simply as possible to control the system and show the functionality in the system. The button design uses only two buttons, one for starting the recordings and one for stopping the recordings. The design uses the GPIO pins 2 and 4 on the Spresense main board and reacts to the value of the boolean input.

The choice of implementation for the debounce design is a RC circuit to eliminate multiple bounces while pushing the buttons. The capacitor C1 of 0.1 μF is placed over the button with a resistor of 100 kΩ at the bottom between ground and the button. The design was implemented with a RC circuit for each of the two buttons.

3.6 System development

The Arduino IDE was used to program the Spresense and the setup was explained in detail in section 3.1.

3.6.1 Audio

Recording high resolution audio on 4 channels is accessed with `Audio.h` in the Arduino IDE. Using the Arduino IDE to write and edit code, and to compile and flash the scripts to the board. The Spresense comes with an example library for different audio applications, which is useful for looking at different functionalities in the system and how to communicate with the Spresense. Learning the structures of the audio library was still needed to configure functionality for the acquisition system. Looking at the audio, the file is placed as an object that can change states. This object is used for all of the audio related parts of the system. The block diagram in figure 3.9 shows how the audio module works and how to jump between the high level states.

Looking closer at some of the functions, the `theAudio->setRecorderMode()` was used to set the audio object's state to recording audio. The `theAudio->initRecorder()` function takes the codec, sampling rate, bit rate and the number of channels to initialize the recording. For WAV files it is needed to write the WAV header before `theAudio->startRecording` starts the

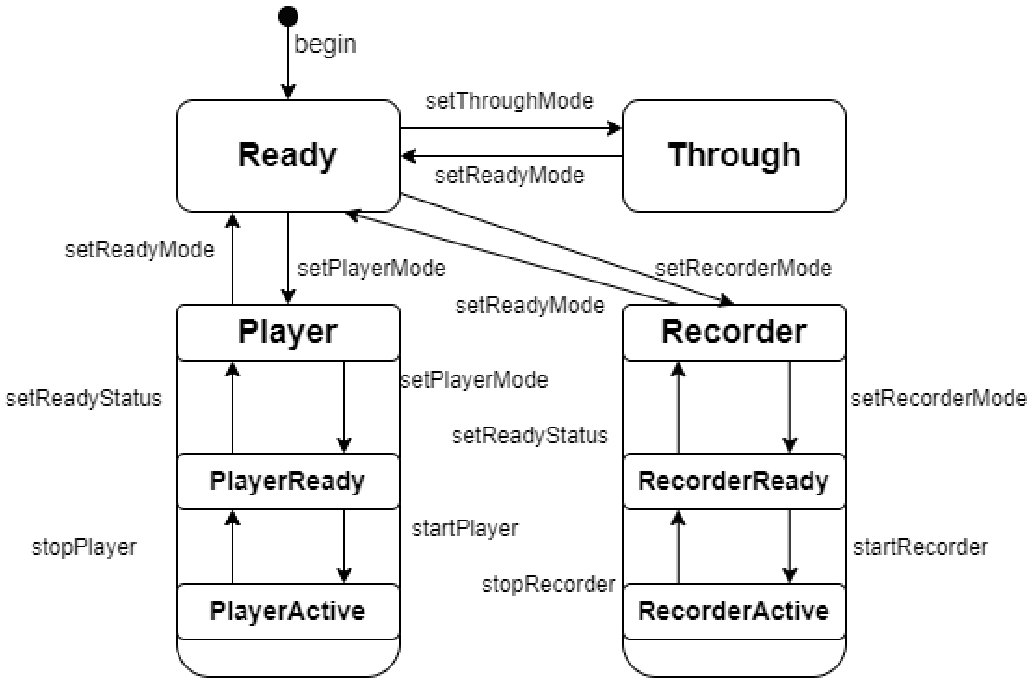


Figure 3.9: The audio states when using the IDE inspired by [18].

recording. Functions for stopping the recording, writing the data to a file and then closing the file, before changing the state back to the ready mode and/or ending the audio object when finished. All the used state changing functions are included in code listing 3.1.

Code listing 3.1: Audio recording code Arduino

```
AudioClass *theAudio;

theAudio = AudioClass::getInstance();
theAudio->begin(audio_attention_cb);

theAudio->setRecorderMode(AS_SETRECDR_STS_INPUTDEVICE_MIC);
theAudio->initRecorder(AS_CODECTYPE_WAV, "/mnt/sd0/BIN", recoding_sampling_rate,
                         recoding_bit_length, recoding_cannel_number);

theAudio->writeWavHeader(myFile);
theAudio->startRecorder();

theAudio->stopRecorder();
theAudio->readFrames(myFile);
theAudio->closeOutputFile(myFile);
```

```
theAudio->setReadyMode();
theAudio->end();
```

3.6.2 Camera

For capturing images on the Spresense with the camera board, the camera library sets the camera up as a class, `theCamera`, for adjusting settings and capturing images at a specified interval. Starting with the example code provided for the camera, `Camera`, the settings of the camera were altered to capture images every second and the number of images to capture was adjusted to match limits of the library. To work with the camera, the camera class was inspected to change settings on the camera and capture the images. In listing 3.2 the camera class and functionality is presented.

Code listing 3.2: Image recording, Arduino

```
theCamera.begin();

theCamera.startStreaming(true, CamCB);

theCamera.setAutoWhiteBalanceMode(CAM_WHITE_BALANCE_DAYLIGHT);

theCamera.setStillPictureImageFormat(CAM_IMGSIZE_QUADVGA_H, CAM_IMGSIZE_QUADVGA_V,
                                     CAM_IMAGE_PIX_FMT_JPG);

CamImage img = theCamera.takePicture();

theCamera.end();
```

Chapter 4

Test and evaluation of the acquisition system

This section will present the tests and evaluations of the acquisition system implemented, starting with the hardware evaluation and verification of the audio recorded. Thereafter the pre-amplifier and filters are evaluated, and lastly the software for recording and image capturing are evaluated.

4.1 Hardware verification

A large part of building the system for data acquisition is the hardware. The hardware includes the Spresense, the pre-amplifier, power supply circuit and the debounce circuit. To ensure the correct operation of the electronics, multiple tests and verification methods were investigated. The Digilent Analog discovery 2 - USB oscilloscope was used to test the electronics. The Digilent was used with its oscilloscope, wave generator and power supply for testing and verification.

4.1.1 Pre-amplifier

Testing the pre-amplifier is divided into two separate parts. First looking at the amplification factor and the amplified signal compared to the input signal. Next the lowpass filter is tested for its functionality.

Amplification

Testing the pre-amplifier aiming for a amplification factor of 21.28 calculated in the implementation.

Amplification was tested with the Digilent, using the signal generator to apply a 125 Hz sine wave with an amplitude of 20 mV at the input of the pre-amplifier. The frequency was chosen to test for amplification in the low frequency spectrum which is tailored for recording of sounds from insects.

Power supply of the Digilent delivers a positive supply of 5 V and ground to power the circuit. The oscilloscope of the Digilent was used to measure the sine wave generated at the input and measure the amplified signal on the output of the pre-amplifier. A list of the measured input signals, output signals and amplification factors are provided in table 4.1.

Table 4.1: Overview of amplification for each of the channels.

Channel	Input amplitude	Output amplitude	Amplification factor
Channel A	20.51 mV	431.5 mV	21.04
Channel B	20.39 mV	426 mV	20.89
Channel C	20.3 mV	428 mV	21.08
Channel D	20.45 mV	425.5 mV	20.81

The signal generated from the Digilent has slight variations on the input of the pre-amplifier and thus variations on the output amplitude. The values presented in the table are selected mean RMS values for the amplitude per channel. Table 4.1 shows the differences in amplification between the channels by dividing the output on the input. The largest difference of 0.27 mV is between channel C and D.

From figure 4.1, an exported image from Digilents oscilloscope, shows the amplification and inversion of the input signal. The axes for channels C1 and C2 are offset to align the sine waves and can be seen on the left hand side of the figure for reference.

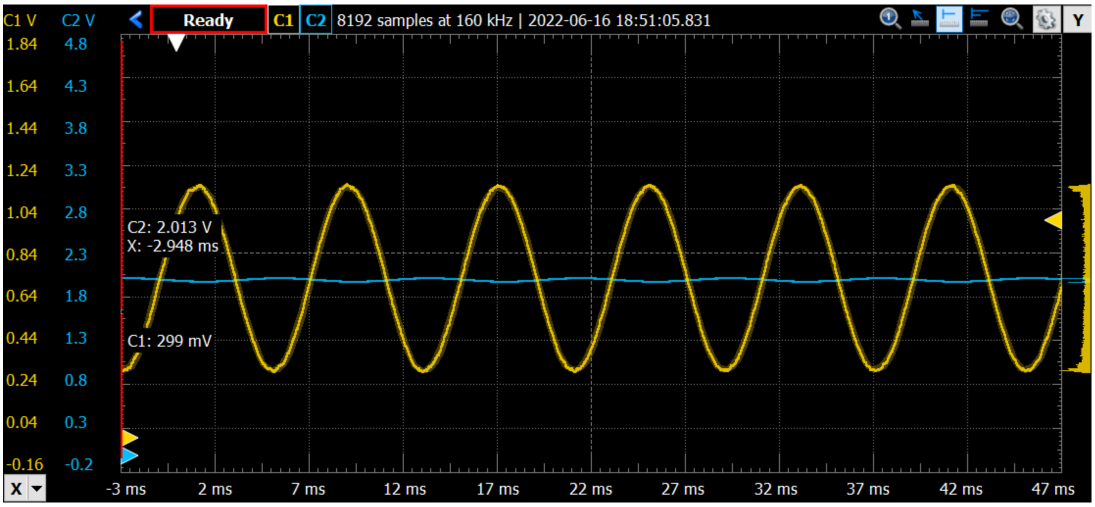


Figure 4.1: Here the blue line shows the signal generator on the input at the microphone pins and the yellow line shows the output in voltage over time on the pre-amplifier.

The overall amplification has low variance between each channel and low deviation from the designed amplification factor of 21.28 times the input signal.

Cutoff

Testing cutoff on the output signal from the pre-amplifier.

The output of the pre-amplifier is designed with a DC offset into the Sony Spresense audio input pin at around 0.73 V. This provides the limits on the input signal. Exceeding an amplified signal with amplitude of 0.7 V might result in clipping and distorted recordings. Testing for the average input amplitude the system is limited to 34 mV on the input, resulting in approximately 0.72 V on the output.

To ensure no clipping in the system, the sound pressure received into the pre-amplifier representing 30 mV or less is recommended.

High- and Lowpass filter

The highpass filter implemented on the pre-amplifier was created to remove the DC from the microphone before entering the operational amplifier. The highpass filter is not evaluated in this thesis. The lowpass filter was created to remove high frequency noise above 15.9 kHz.

Each of the channels were tested by sending a 20 kHz input signal on the input of the pre-amplifier. Each channel on the pre-amplifier has identical filters implemented.

Figure 4.2 shows that the amplification factor is lowered compared to the amplification factor at 125 Hz. Measuring the input signal at 21.45 mV and the output signal at 248.5 mV, gives a amplification factor of 11.59 at 20 kHz, almost half of the highest amplification factor measured for the 125 Hz test signal, which is around 21.

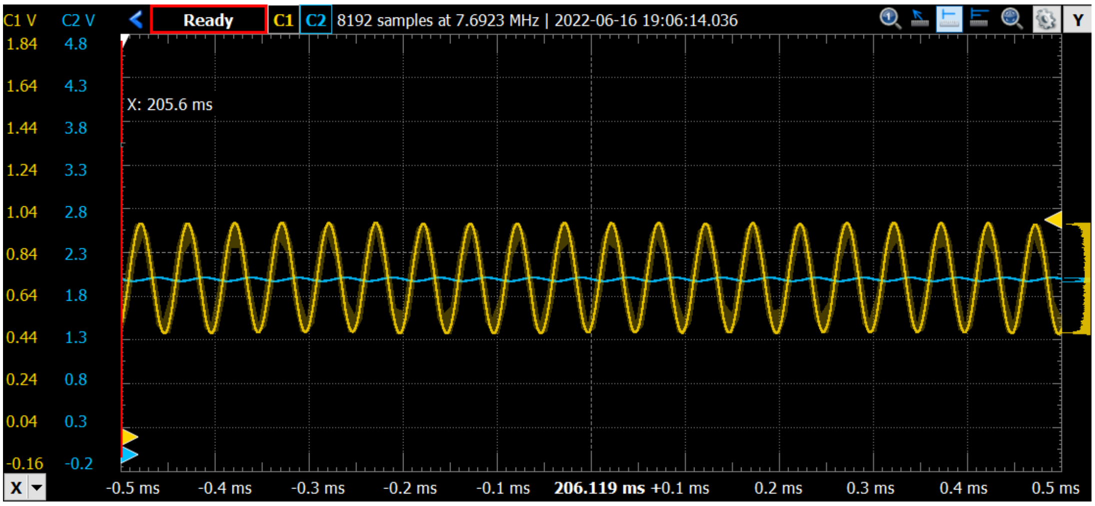


Figure 4.2: The lowpass filter tested on a frequency of 20 kHz input signal, outside the 15 kHz cutoff.

Using 20 kHz resulted in a decrease in the amplification due to the lowpass filter.

4.1.2 Power supply

Testing a stable low noise power supply at 5V.

Testing the power supply circuit, the ANSMANN Alkaline, 550 mAh, battery was connected to the circuit and the oscilloscope channels on the Digilent was connected to the board. Channel C1, was placed on the output of the circuit to evaluate the voltage delivered to the system. Channel C2, was connected to the input and ground to the L7805AC regulator. The measurements on the battery and the output of the linear voltage regulator are presented in figure 4.3.

Looking at the measurements, the battery delivers ∼8.78 V to the regulator with low noise

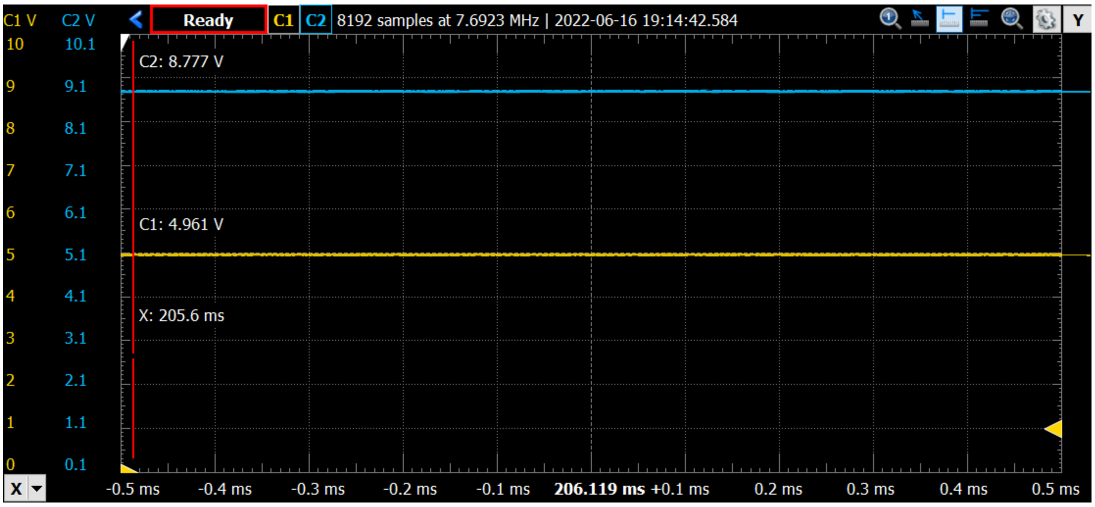


Figure 4.3: Power supply measured with Digilent. The blue line, C2, shows the voltage on the input of L7805AC and the yellow line, C1, shows the output after the fixed voltage regulator.

levels. The output signal to the system is delivered at ∼4.96 V measured at the output of the linear fixed voltage regulator to ground. This nearly equals the desired 5 V power supply, also providing a fairly clean ∼5 V DC power supply to the microphones and pre-amplifier.

4.1.3 Buttons

Testing the debounce functionality on the buttons by looking at measurements from pushing and releasing the button with the debounce circuit.

The Digilent was used to supply 3.3 V simulating releasing the button. And connected to ground simulating pushing the button. The C1 oscilloscope was connected to the output of the button. Measurement of the button press and release are shown in figure 4.4.

The measured time for a stable high signal to a stable low signal and vise versa was deduced from the time axis in the oscilloscope. Charging the capacitor in the RC circuit and rising the voltage to a high signal takes around 20 ms. While discharging back to a low signal takes around 6 ms.

From the measurements, no multiple peaks when pushing the button was detected. So the bounce was successfully removed.

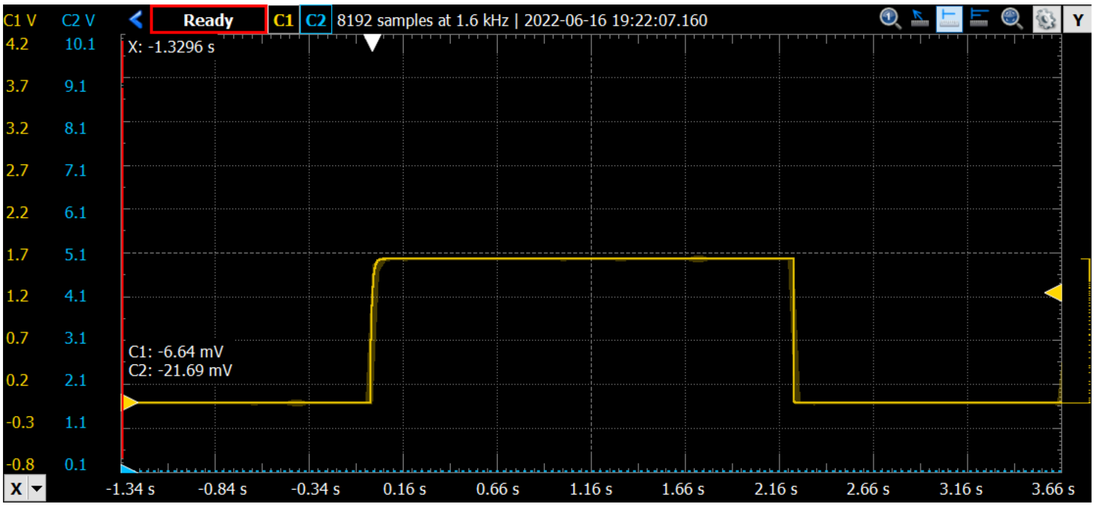


Figure 4.4: Button design with debouncing measured with the Digilent.

4.2 Audio capturing

Testing the quality of the audio capturing solution focusing on noise and frequency spectrum and response. To see the actual performance of audio capturing solution consisting of the microphones, pre-amplifier, Spresense, powered by the power supply.

Testing of the audio solution was done in the anechoic chamber at Norwegian University of Science and Technology. EASRA, Electronic and Acoustic System Evaluation and Response Analysis, was used to deliver the sounds through a loudspeaker placed 96 cm above ground, aligned with the microphones and 100 cm between the microphones and loudspeaker. The microphones were placed linearly with 3 cm between each, in front of the loudspeaker. The test signals were set to -18 dB for all tests. The test set-up is shown in figure 4.5.

The equipment used for the test setup in the anechoic chamber at NTNU is shown under in table 4.2.

Table 4.2: Anechoic chamber equipment.

Equipment	Brand	Model	Serial number
Loudspeaker	Genelec	1029A	029A041044
Sound card	Lynx	Aurora 2	N/A

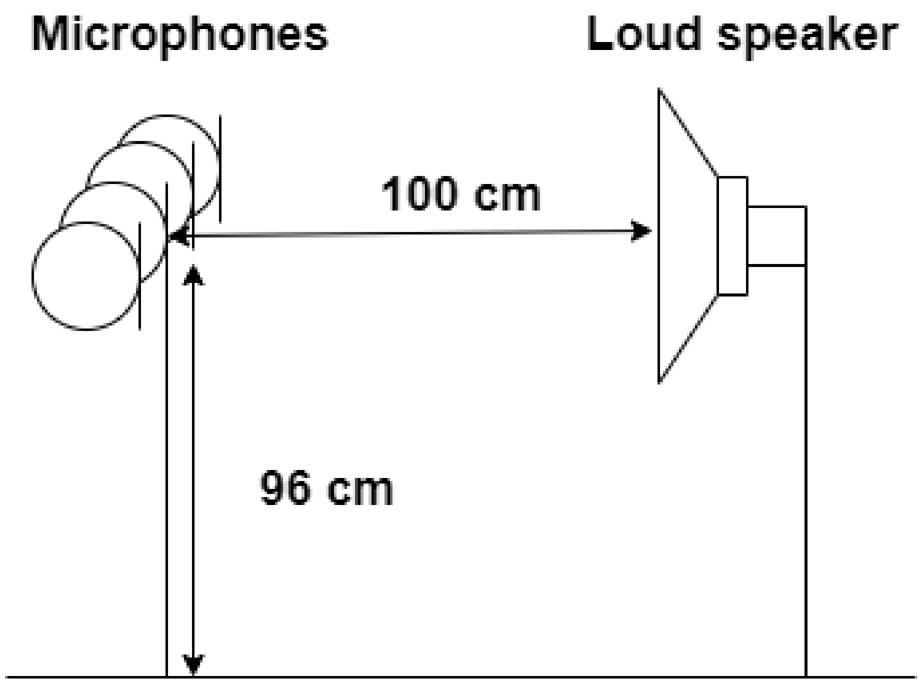


Figure 4.5: Test setup in the anecohic chamber with 4 microphones in one line towards the loud speaker and loud speaker to the right.

Intrinsic noise

Evaluation of the intrinsic noise in the system as a whole was done to see where the noise floor existed and the amplitude at different frequencies. Recording of the intrinsic noise in the system was carried out by recording without any sound in the chamber. The measurements were done on all 4 channels at the same time. All 4 channels produced similar responses when plotted with the magnitude (dB) versus the frequencies (Hz). In figure 4.6 the frequency spectrum of channel A of the intrinsic noise is plotted, and all four channels is added to appendix A in figure A.1.

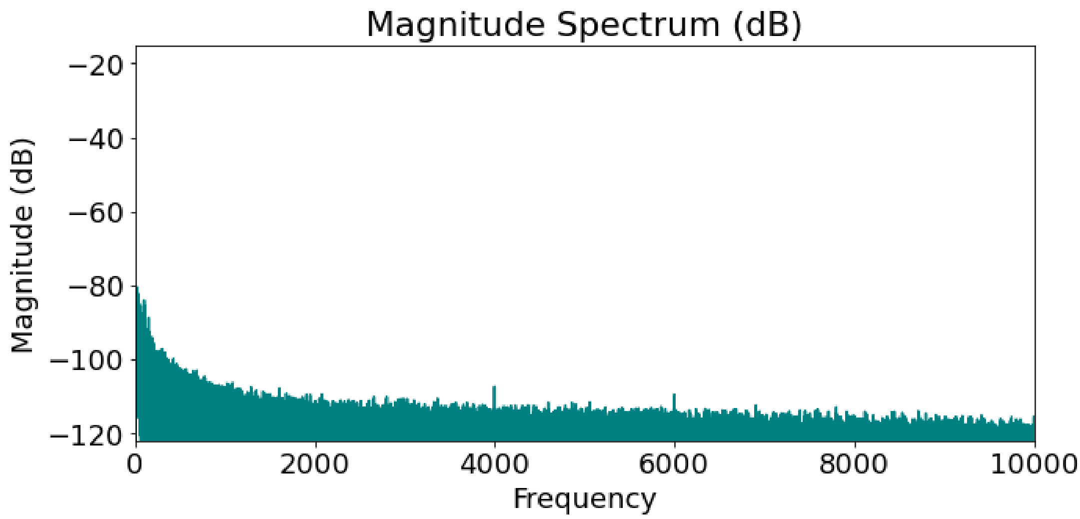


Figure 4.6: Magnitude spectrum of intrinsic noise in the system, recording with no signal in the anechoic chamber.

Noise is present across the entire spectrum up to 10 kHz visualized in figure 4.6, with more noise present at the lower frequencies. Most noise is present in the lower frequencies of the spectrum, and measured to maximum -80 dB with a decay of the amplitude towards the higher frequencies.

The measurements find the noise generated by the system as a whole to be lower than -80 dB.

Frequency spectrum - 125 Hz and 1 kHz test

Evaluating the audio recordings for single frequency testing, and comparing recordings between the microphones and the signal to noise ratio for each channel. The 125 Hz signal closely resembles the fundamental frequency of the insects from the specialization project.

First, testing with a 125 Hz sine wave signal generated in EASRA, played on the loudspeaker in the anechoic chamber. From figure 4.7, the plotted spectrum displays the 125 Hz frequency and the harmonics from channel A. The noise floor recorded and evaluated above is visible in the recording measured for 125 Hz signal.

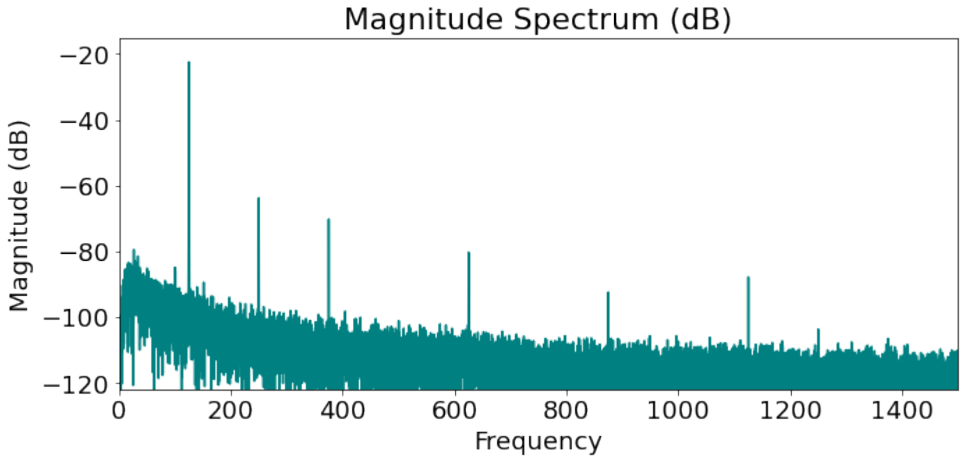


Figure 4.7: Magnitude spectrum of 125 Hz played at -18 dB, 1 m from the microphone. Magnitude [dB] plotted versus Frequency [Hz]

The frequency spectres from all 4 channels are similar in shape and response, and are added in appendix A in figure A.2.

A common test for microphones includes looking at the response with a 1 kHz sine wave. The recorded signal was transformed into a magnitude spectrum with a larger frequency axis to see the harmonics of the 1 kHz signal visible at 2 kHz and 3 kHz. Figure 4.8, shows that the noise floor is similar to the 125 Hz test.

All 4 channels produced similar responses and all 4 can be found in appendix A in figure A.3.

As seen in figure 4.7 and 4.8, the non-linear behavior of loudspeakers cause harmonics to

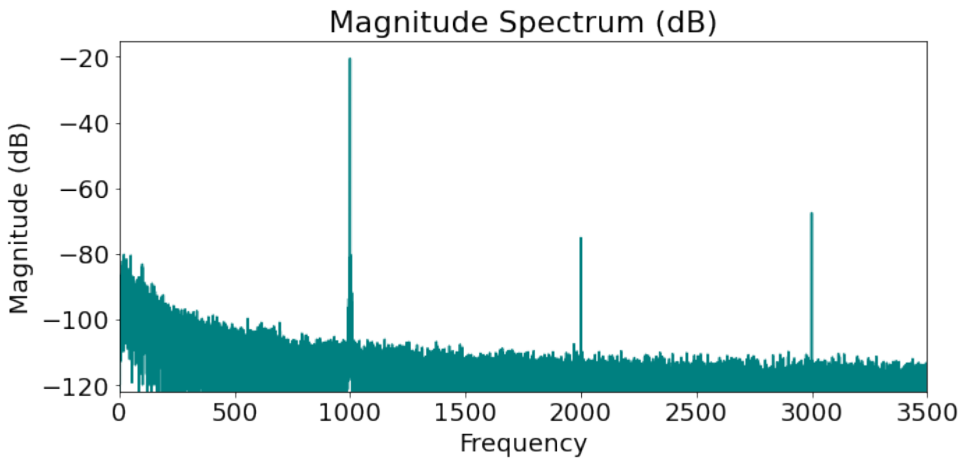


Figure 4.8: Magnitude spectrum of 1000 Hz played at -18 dB, 1 m from the microphone.

appear while reproducing the frequencies. These harmonics overtones are visible at many of the consecutive multiples of the fundamental frequency.

Frequency response - Logarithmic Sine Sweep

To check the frequency response in the audio capturing solution, a logarithmic sine sweep was played in the anechoic chamber in front of the microphones.

The spectrogram recording on channel A, of the logarithmic sine sweep is shown in figure 4.9. We found similar looking spectrograms for each channel, all four spectrograms are located in appendix A in figure A.5.

Looking closer at two of the frequency spectrums of the logarithmic sine sweep recorded on channel C and channel B, we see a notch form around 6 kHz for channel B, meaning that channel B is particularly bad at recording frequencies around 6 kHz. All four channel measurements are located in appendix A in figure A.4.

From the sine sweep recordings we want to find the impulse response. First, channel A was chosen to be the reference microphone. The sine sweep generated in EASRA was extracted and an inverse filter was added to it. To find the impulse response, the signal recorded on channel A and the inverse sine sweep was convoluted to obtain the impulse response. The three signals described are shown in figure 4.11.

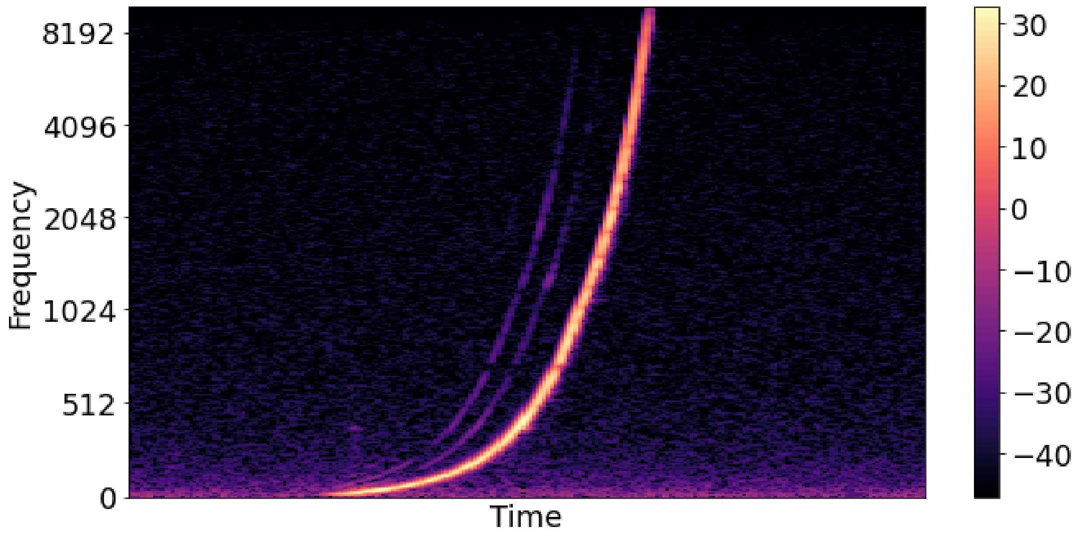
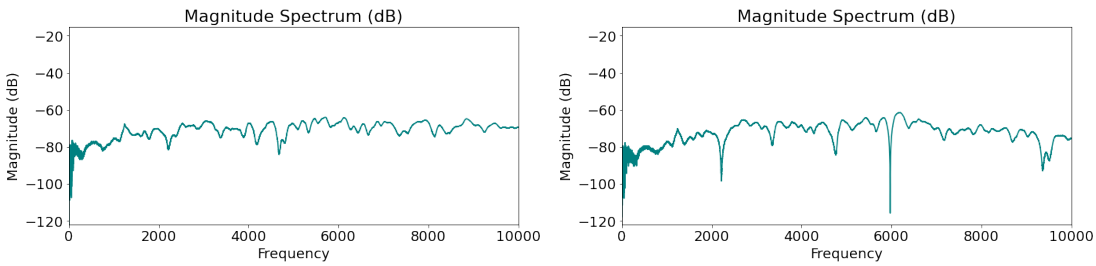


Figure 4.9: Spectrogram of channel1 of recorded logarithmic sine sweep.



(a) Frequency spectrum of channel3.

(b) Frequency spectrum of channel2.

Figure 4.10: Frequency spectrums recorded on channel3 and channel2 of the logarithmic sine sweep.

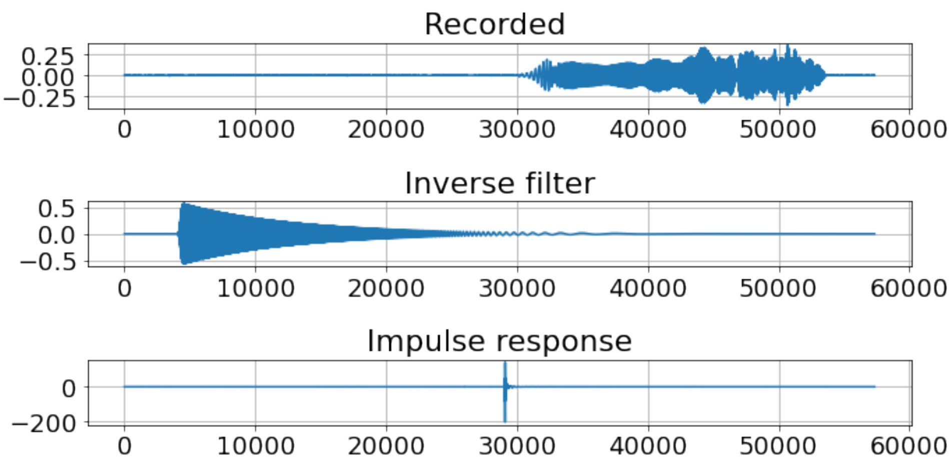


Figure 4.11: Channel 1 recording at the top, inverse filter in the middle and impulse response at the bottom.

Taking the Fourier transform of the impulse response, the magnitude-frequency response found, shown in figure 4.12. From this response we see that the impulse response is not flat as expected, this is from channel A's imperfections.

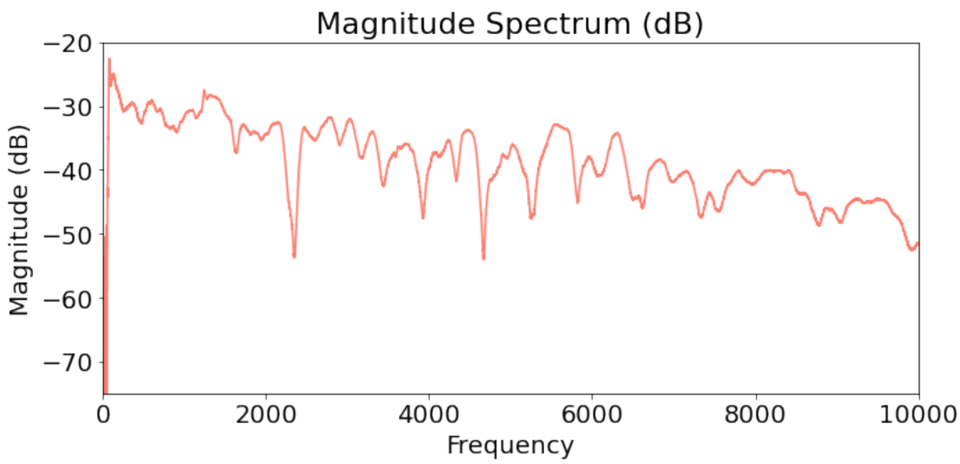


Figure 4.12: Frequency spectrum for the impulse response.

4.3 Software verification

Evaluating the audio recording and images capabilities of the Sony Spresense and programming it with the Arduino IDE.

4.3.1 Testing audio recording length

For testing audio recording length, the example code provided with the Arduino Spresense Library was used. The script $recorder\_wav$, example was used to modify settings and match the desired recording setup. The settings for audio recording are shown in table 4.3 from section 2.

The recording setup was flashed from a computer and used to power the Spresense with the USB on the device. The Spresense board was connected to the pre-amplifier with the microphones attached, powered by the 9 V battery through the power circuit.

During recording of WAV-files on the Spresense the program stores a WAV header to each recording. The header contains information about the recorded track with sample rate, bit

Table 4.3: Audio recording settings.

Specification	Settings
Sampling rate	48000 Hz
Number of channels	4
Bit depth	16
File	WAV

depth and number of channels to mention a few. The size of the header per recording is 31 bytes. To calculate the size of the recording, the sampling rate, number of channels, bit depth and recording time are used and divided by 8 to get the size in bytes.

$$RecordingSize = \frac{Fs \cdot \#channels \cdot bitDepth}{8} \cdot seconds \tag{4.1}$$

Testing the limitations in recording length on the Spresense with recordings of the following lengths, 5 min, 10, min, 20 min and 1 hour. The 10 min recording spectrogram is shown in figure 4.13.

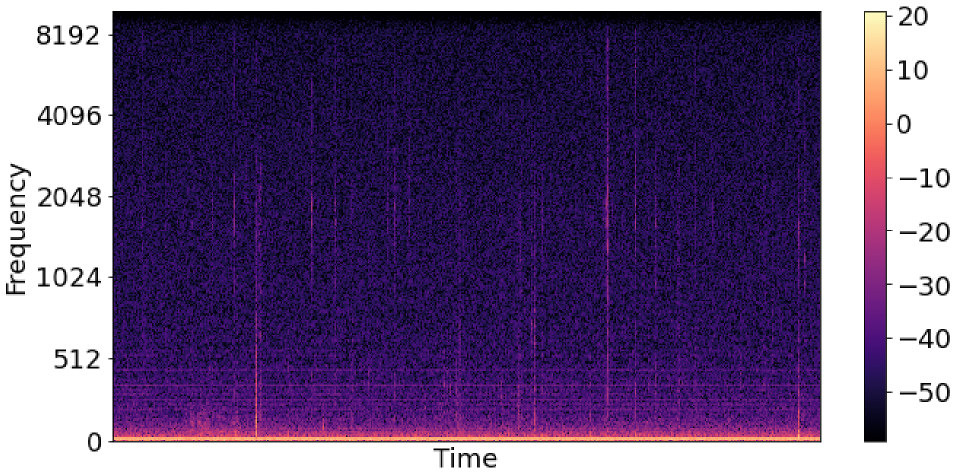


Figure 4.13: Spectrogram of 10 minutes of recording.

No problems arose during recordings of the three first tests, but recording 4 channels for 1 hour resulted in problems with the recorder sending an *Attention: module[4][0] attention id[1]/code[6] (objects/media_recorder/audio_recorder_ sink.cpp L84)* message. By inspecting the audio recorded, however, the audio sounded normal. Running tests with a lower sampling rate, 16000 Hz, or only 2 channels, worked for 1 hour recordings, indicating a memory managing dilemma.

The test showed that the Spresense was able to record up to at least 20 minutes with the chosen set-up. Not reaching 60 minutes which was the target in the specifications.

4.3.2 Testing image capturing

For testing the image capturing, the example code provided for the camera, was used with the Arduino Spresense Library. Using the example code with the standard size, white balance (WB), and format provided. The settings for capturing images are shown in table 4.4.

Table 4.4: Image capture settings.

Specification	Setting
Height	960 pixels
Width	1280 pixels
WB	Daylight
Bit depth	24
Format	JPG

The setup for capturing images was executed running the code script from a computer and powering the Spresense with the USB port on the device. The camera board was attached to the Spresense main board and the tests included capturing 1 image per second for 20 minutes.

Recording JPG-files on the Spresense at 5 megapixels does not exhaust resources with file sizes at 73 kB per image.

Looking at the images produced by the camera board, the Depth of Field (DoF) of the camera is limiting the distance from the target and generating a blurry image at closer proximity to the subject.

(a) Image captured further than 77.5 cm from the flower. **(b)** Image captured closer than 77.5 cm from the flower.

Figure 4.14: Captured images of flowers outside with camera module.

4.3.3 Testing audio and image capturing together

To enable audio and image capturing on the Spresense, the desired solution was designed for simultaneous data acquiring. Unfortunately, the Spresense Arduino library is not currently capable of running audio and camera libraries on the subcores [19]. The high sample rate also made it impossible to capture images between recording samples.

This resulted in recording audio and capturing images separately, recording audio for 3 seconds, stopping the recording and starting image capture before repeating. For running both libraries the memory size of the main core was increased from 1024 kB to 1280 kB. The memory space still did not allow for high quality recordings and capturing images, resulting in that we had to lower the sampling frequency to 16000 Hz and recording only one channel. The buttons were integrated to control the system, starting and stopping data acquisition by polling using if-statements.

Running the tests without microphones connected was successful without faults and errors. Running with both microphones and camera connected resulted in a memory issue between the required space for capturing images and the audio recording FIFO on the board. The camera module prompts `Error: No memory` when having the memory at 1024 kB. Not able to change the size of the recorder module the size allocated to capturing images on the main core was increased to 1280 kB. Setting the memory above 1280 kB results in a memory error for the audio module with the prompt `ERROR: FIFO area is not allocated`, meaning the FIFO does not have adequate memory available outside the main core to store data.

The implemented system was unable to record audio and capture images at the same time with the setup created in this Master's thesis.

Chapter 5

Discussion of the acquisition system and results

This master thesis has described the specification, implementation and evaluation of the proposed acquisition system. The process has included research into different technologies and specialization fields, working with a new MCU, hardware design and testing, acoustics through microphones and measurements, 3D modeling and printing, and programming functionality.

In this section the contents of the master thesis work is discussed, looking at the technologies and choices, discussing the design and all of its parts. The section starts with the hardware used and developed, then looking at the audio functionality and results. Both limitations and improvements are discussed and presented.

5.1 Hardware discussion

A substantial part of the master thesis was dedicated to developing and testing the hardware. In this hardware discussion section the hardware from the MCU to the developed circuits are discussed.

5.1.1 Sony Spresense

Starting with the Sony Spresense platform and extension boards. The Sony Spresense has a detailed setup guide for the different work environments. Arduino with the Arduino IDE, the Spresense SDK and a separate version for CircuitPython. All three setups were tested and functioned with hello world examples.

Spresense with the Arduino IDE was used for testing the capabilities of the system while also developing the rest of the system. The experience of using the Arduino IDE with the Sony Spresense was close to using it with an Arduino device. For prototyping, the Arduino IDE made it convenient to compile and run the scripts to the Sony Spresense board.

The installed Spresense Arduino Library comes with example code for many different operations to show functionality and comments for learning how to program the Sony Spresense. The examples also made it easier to test the system at a fast pace. Using a new MCU often comes with a learning curve to operate the features and understand how to control it. Using the documentation provided from Sony in the beginning of the process was, in the author's opinion, hard to use making it difficult to find information quickly. During the Master's thesis, the website has been updated for a better experience navigating to relevant information.

After reassembling the main board with the extension board, the micropins between the boards did not connect correctly and the connection was loose. This resulted in a small section of the connections not working and resulted in a bad connection to the SD-card. Connecting the boards with care is advised not to damage the B-2-B connection.

Software on the Spresense

The intention was to capture audio and images simultaneously, but the test showed that it only was possible to do this in sequential. The intention of recording while capturing verification images did not pan out. While the Sony Spresense has 6 cores to work with, the Spresense Arduino Library is not optimized for running the camera- or the audio library on the 5 sub-cores at the moment of writing the report. These restrictions do not apply to the Spresense SDK and are likely to be the way forward for simultaneous data acquisition.

Audio inputs

The audio inputs on the extension board are useful and accessible for the user. By default the pins are arranged to work for 4 channel analog recordings, but the extension board can be modified for 8 channels with digital microphones. From the manufacturer, the ground pins in JP10 are recommended for low noise operations when recording. With the header pins on the extension board, developing a shield based on the pre-amplifier will decrease the need for cables between the boards and clutter when recording.

Camera

The specified system does not place hard restrictions on the camera's performance. The camera is an important part of the system for enabling crucial verification of the insects. The camera's performance from the test photos needs a decent amount of light to be optimal, and should not be a problem for recording insects during daytime. The depth of field on the camera extension board means that it's not possible to focus closer than 77.5 cm. The Hymenoptera are small insects and might be hard to detect from that distance. While working on this Master's thesis, Sony launched a HDR camera extension board with variable depth of field, and a close up function down to 5.1 mm. On paper this is a solution to be looked at in the future.

Insights and further work

The software developed and examples used from Arduino IDE are too constraining on memory for data acquisition and not optimized for the libraries. Exploring the Spresense SDK for simultaneous recording needs to be a next step in development and further work. The depth of field of the camera is not suitable for the intended task. Investigating the use of the Sony Spresense HDR camera should be explored further.

Insight in to the Spresense platform have been a learning process, getting to know the hardware and software solutions was a big part of getting the project started. Learning that better research into functionality in regard to the subcores would have saved a lot of time and steered the project into another direction earlier.

5.1.2 Ambisonic microphone mount

The ambisonic mount was sturdy and the microphones fit snugly into the mould, but the mould by itself needs a stand to keep it upright. There was unfortunately no time to test the ambisonic setup with direction detection in this Master's thesis.

Insights and further work

3D printing a case for the electronics with a stand for the ambisonic mount are recommended improvements to make it easier to deploy the system in real life use cases. For optimal results in direction detection, calibration of magnitude and phase is important for further work in this area.

5.1.3 Microphones

The microphones chosen for this project, CMA-4544PF-W, showed promise to the specifications with low noise, low cost and a reasonable sensitivity. From the frequency response measurements, the differences on some of the microphones are visible, especially channel B.

Insights and further work

With cheaper microphones the calibration and testing is important to make sure the microphones operate with similar responses. Proper calibration should be performed during testing. Also testing more microphones and selecting the best ones to get the quality needed.

5.1.4 Power supply

When working on powering the pre-amplifier multiple approaches were explored. Firstly, the barrel jack connector was implemented on the pre-amplifier for testing with a wall mounted power supply securing a low noise power supply. This was not necessary in the later stages of the design with the selected power supply circuit. During testing of the 5 V output power from the Spresense board, we found it to be a switching signal that didn't generate a clean supply to the microphones.

The next approach was using a boost converter to amplify the voltage from the microphone BIAS of the audio ports. But by boosting the 2 V bias to 5 V, the current was not strong enough to drive the circuit. The last step was to use regulators to step down the voltage of a 9 V battery. Firstly by using a switching regulator circuit, but that generated too much noise. Then the circuit design of the linear fixed voltage regulator was created as described in section 3.

From measuring the input and output from the linear regulator, L7805AC, we see a small irregularity from the intended 5 V. The power delivered to the pre-amplifier is 39 mV lower than 5 V, but does not contribute to a vital difference to the system as a whole. Importantly, the power supply circuit significantly reduces the noise into the system compared to the other solutions tested.

Insights and further work

A natural next step for the power circuit is to add the power circuit directly to the PCB to remove additional cables and plug the battery directly into the board with a screw mount as presented here or with a barrel jack. Removing more cables reduces connection issues and noise added to the audio recordings through the circuitry.

5.1.5 Pre-amplifier

The hardware tests of the pre-amplifier were successful in recreating an amplification close to a factor of 21 in gain, with a slight difference due to natural variation in the components used. The lowpass filter tested fulfilled the intended use case of removing the DC signal from the audio before amplification and reduce high frequency noise. The fundamental frequencies of the Hymenoptera acquired in the specialization project are mostly located at less than 300 Hz. For the specialization project a lowpass filter was implemented in post processing to remove high frequencies without removing harmomics, and was placed at 2500 Hz. In specialized circumstances of only recording Hymenoptera the lowpass filter of the pre-amplifier can be adjusted by changing the capacitors, to remove frequencies closer to 2500 Hz. In the case of this general system with the lowpass filter at 16 kHz, post processing can be applied to the data before adding it to a database.

A large part of the pre-amplifier design was to reduce noise into the system for recording high quality audio. The dual low noise operational amplifiers, NE5532AP, was smoothly implemented to the system and operated to they're specifications. Contributing with good amplification and low noise.

The PCB prototype was designed following the recommendation and instruction of Electronics and prototype laboratories. Creating the PCB locally at NTNU made the production time faster, but generated a suboptimal result of the layout.

As shown in section 3, the channel A input is switched on its head compared to the three other channels. This does not affect the functionality of the pre-amplifier, but makes it more inconvenient to remember the polarity of the input when in use. The central connection pins are also flipped in comparison to the Spresense audio input pins, where the signal input is placed in the middle and not the top.

Insights and further work

For future improvements the production of the PCB should be factory made without the restriction on manually placed vias on the board to make it more convenient and possible to mount the PCB directly to the Spresense as a shield.

5.1.6 Buttons and debounce

The debounce RC-circuit ensures that the button push is only registered once. From the tests, charging- and discharging time for the RC-circuit is short enough not to be a problem.

Insights and further work

Further work on a controllers can include software gain controller for adjusting additional gain if needed in certain situations.

5.2 Audio capturing

From the measurements on intrinsic noise, the noise floor is located beneath -80 dB in the frequency range for the Hymenoptera. This shows that signals above this threshold will be possible to locate and distinguish from noise.

The 125 Hz test and the 1 kHz test on all four microphones result in similar frequency spectrum for magnitude and frequency. Supporting that the microphones pick up the same signal with a low degree of variation on these frequencies and the harmonic overtones.

The logarithmic sine sweep test was performed on all four microphones and the spectrograms look similar, looking further into the frequency spectrum the differences are easier to point out. Channel B has a notch around 6 kHz that can affect audio in that frequency region.

Insights and further work

For future endeavors, using a calibrated microphone is recommended. From the data captured in the anechoic chamber, future efforts to calibrate the microphones will be necessary working with low-cost electret microphone. The impulse response calculated in section 4, the transfer function can be derived to find a correction filter to calibrate the microphones with similar frequency and phase responses.

Chapter 6

Conclusion and final remarks

Through the Master's thesis, the primary question was to test and design a data acquisition system for for the insect order Hymenoptera. This report have focused on different parts of the system and the prototypes created. In the introduction of the thesis a research question was introduced. **Is it possible to create a low cost embedded acquisition system that records multichannel audio of high resolution, and with the capability of capturing images?**

With all of the parts put together, the system as a whole is incomplete without a well functioning platform that can record audio and capture images simultaneously or sequentially. The Spresense is optimal for recording high quality audio on four channels, but falls short using the Arduino IDE for both image and audio acquisition as discussed. Further work using the Spresense SDK for programming the Spresense and unlocking the capabilities of simultaneous data acquiring needs to be verified before the Spresense is verified as a good MCU for this type of data acquisition system.

The selected solution designed, implemented and evaluated does not positively confirm the research question. But it is my hope that other students and researchers continue this important work and successfully build a low cost embedded acquisition system that can contribute to preserving the biodiversity in the world.

Bibliography

[1] A. P. Hill, P. Prince, E. P. Covarrubias, C. P. Doncaster, J. L. Snaddon and A. Rogers, 'AudioMoth: Evaluation of a smart open acoustic device for monitoring biodiversity and the environment,' *Methods in Ecology and Evolution*, vol. 9, no. 5, N. Isaac, Ed., pp. 1199–1211, Jan. 2018. DOI: 10.1111/2041-210x.12955.

[2] N. E. Miller-Struttmann, D. Heise, J. Schul, J. C. Geib and C. Galen, 'Flight of the bumble bee: Buzzes predict pollination services,' *PLOS ONE*, vol. 12, no. 6, R. M. Borges, Ed., e0179273, Jun. 2017. DOI: 10.1371/journal.pone.0179273.

[3] S. G. Potts, J. C. Biesmeijer, C. Kremen, P. Neumann, O. Schweiger and W. E. Kunin, 'Global pollinator declines: Trends, impacts and drivers,' *Trends in Ecology &amp Evolution*, vol. 25, no. 6, pp. 345–353, Jun. 2010. DOI: 10.1016/j.tree.2010.01.007.

[4] F. Sánchez-Bayo and K. A. Wyckhuys, 'Worldwide decline of the entomofauna: A review of its drivers,' *Biological Conservation*, vol. 232, pp. 8–27, Apr. 2019. DOI: 10.1016/j.biocon.2019.01.020.

[5] K. Øverli, 'Machine learning for large scale bioacoustic monitoring of Norwegian Hymenoptera,' 2021.

[6] D. Nirmal, 'Machine Learning is Everywhere: Preparing for the Future,' Last accessed: 29. June 2022, 2017. [Online]. Available: https://www.datanami.com/2017/07/03/machine-learning-everywhere-preparing-future/.

[7] J. J. Noda, C. M. Travieso-González, D. Sánchez-Rodrıguez and J. B. Alonso-Hernández, 'Acoustic classification of singing insects based on MFCC/LFCC fusion,' *Applied Sciences*, vol. 9, no. 19, p. 4097, Oct. 2019. DOI: 10.3390/app9194097.

[8] T. Oikarinen, K. Srinivasan, O. Meisner, J. B. Hyman, S. Parmar, A. Fanucci-Kiss, R. Desimone, R. Landman and G. Feng, 'Deep convolutional network for animal sound classification and source attribution using dual audio recordings,' *The Journal of the Acoustical Society of America*, vol. 145, no. 2, pp. 654–662, Feb. 2019. DOI: 10.1121/1.5087827.

[9] N. Volkmann, B. Kulig, S. Hoppe, J. Stracke, O. Hensel and N. Kemper, 'On-farm detection of claw lesions in dairy cows based on acoustic analyses and machine learning,' *Journal of Dairy Science*, vol. 104, no. 5, pp. 5921–5931, May 2021. DOI: 10.3168/jds.2020-19206.

[10] S. Kawakita and K. Ichikawa, 'Automated classification of bees and hornet using acoustic analysis of their flight sounds,' *Apidologie*, vol. 50, no. 1, pp. 71–79, Jan. 2019. DOI: 10.1007/s13592-018-0619-6.

[11] A. R. S. Parmezan, V. M. A. Souza, I. Žliobaitė and G. E. A. P. A. Batista, 'Changes in the wing-beat frequency of bees and wasps depending on environmental conditions: A study with optical sensors,' *Apidologie*, vol. 52, no. 4, pp. 731–748, Jun. 2021. DOI: 10.1007/s13592-021-00860-y.

[12] A. Gradišek, G. Slapničar, J. Šorn, M. Luštrek, M. Gams and J. Grad, 'Predicting species identity of bumblebees through analysis of flight buzzing sounds,' *Bioacoustics*, vol. 26, no. 1, pp. 63–76, May 2016. DOI: 10.1080/09524622.2016.1190946.

[13] 'Spresense Hardware Documents,' Last accessed: 29. June 2022: Sony Corporation. [Online]. Available: https://developer.sony.com/develop/spresense/docs/hw_docs_en.html#_spresense_hardware_comparison.

[14] 'CXD5247GF technical manual,' in *page 23.*, Last accessed: 29. June 2022: Sony Corporation. [Online]. Available: https://www.sony-semicon.co.jp/products/common/pdf/CXD5247GF_technical_manual_02.pdf.

[15] F. Adriaensen, 'A Tetrahedral Microphone Processor for Ambisonic Recording,' 2007. [Online]. Available: https://kokkinizita.linuxaudio.org/papers/tetraproc.pdf.

[16] 'Spresense Arduino Library Getting Started Guide,' Last accessed: 29. June 2022: Sony Corporation. [Online]. Available: https://developer.sony.com/develop/spresense/docs/arduino_set_up_en.html.

[17] 'List of things to consider when ordering PCBs from Elektronikk- og Prototypelaben,' Last accessed: 29. June 2022: Electronics and prototype laboratories. [Online]. Available: https://i.ntnu.no/documents/1306938287/1307152215/List+of+things+to+consider+when+ordering+PCBs+from+Elektronikk.pdf/91858740-a5f1-4399-b1b4-a4369e28ce88?t=1510247874018&status=0.

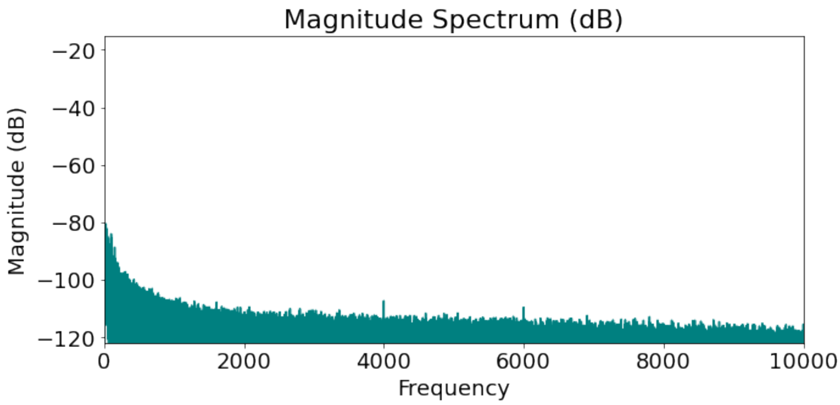
[18] 'Spresense Arduino Library Developer Guide,' in *Section 2.1.4*, Last accessed: 01. July 2022: Sony Corporation. [Online]. Available: https://developer.sony.com/develop/spresense/docs/arduino_developer_guide_en.html#_about_layer_structure.

[19] 'Spresense Arduino Library Developer Guide,' in *Section 2.11.2.8*, Last accessed: 01. July 2022: Sony Corporation. [Online]. Available: https://developer.sony.com/develop/spresense/docs/arduino_developer_guide_en.html#_mp_library.

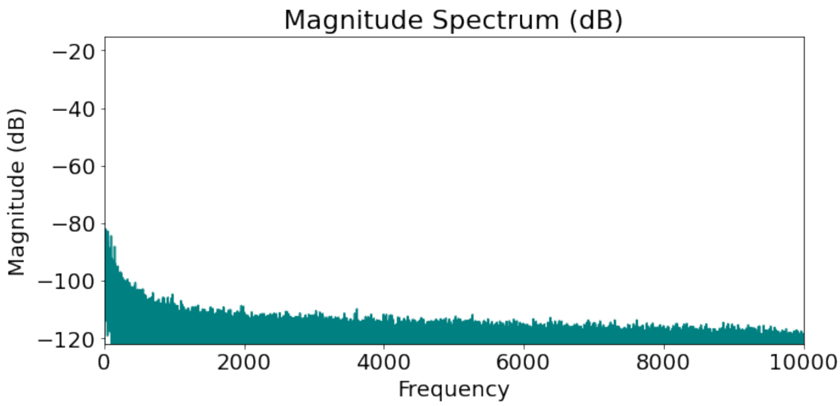
Appendix A

Measurements

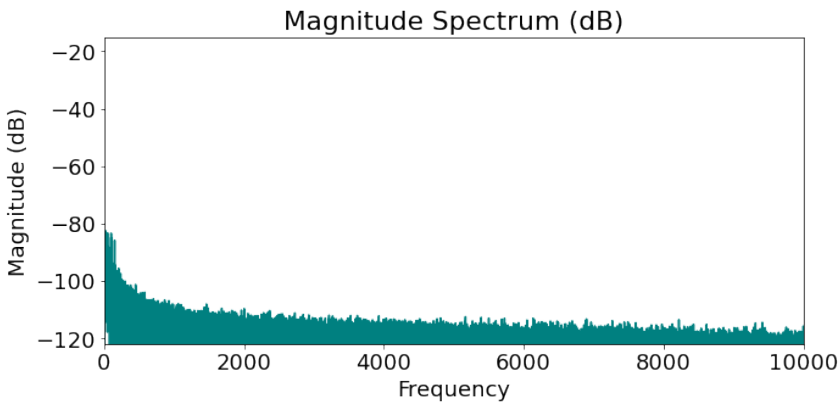
The measurements from recordings on all four channels are placed in this appendix. The fist figure is of intrinsic noise, then the 125 Hz and 1 kHz spectrums. Thereafter is the magnitude spectrum of the logarithmic sine sweep and at the end the spectograms of the logarithmic sine sweep.



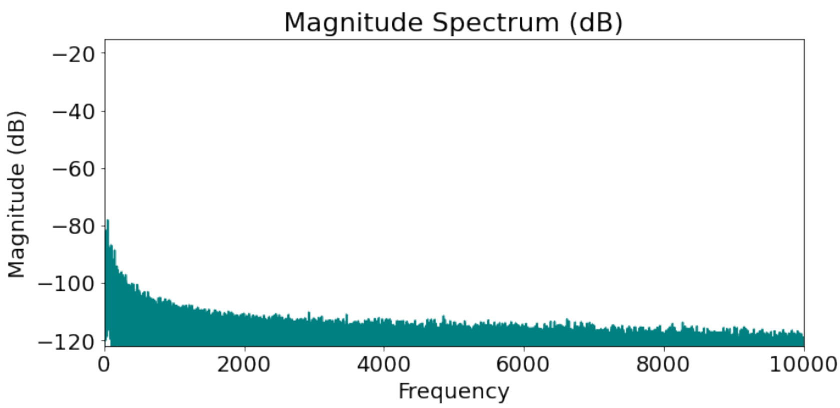
(a) Intrinsic noise recorded on channel A



(b) Intrinsic noise recorded on channel B

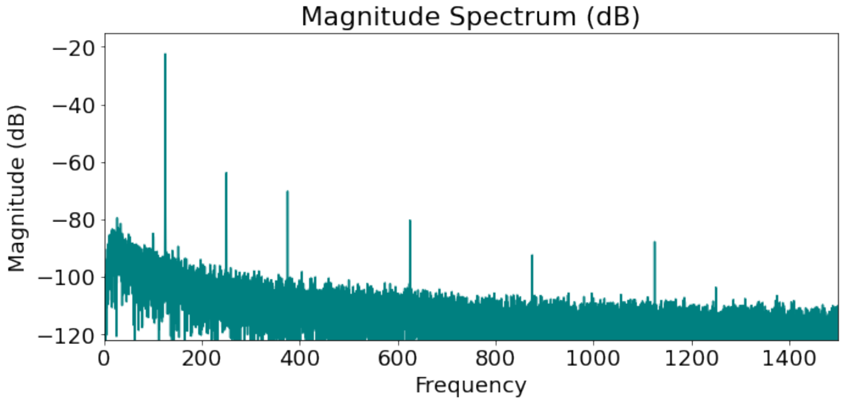


(c) Intrinsic noise recorded on channel C

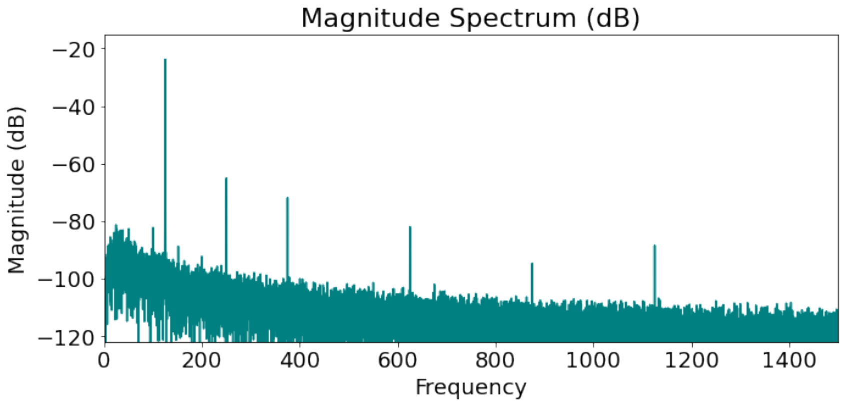


(d) Intrinsic noise recorded on channel D

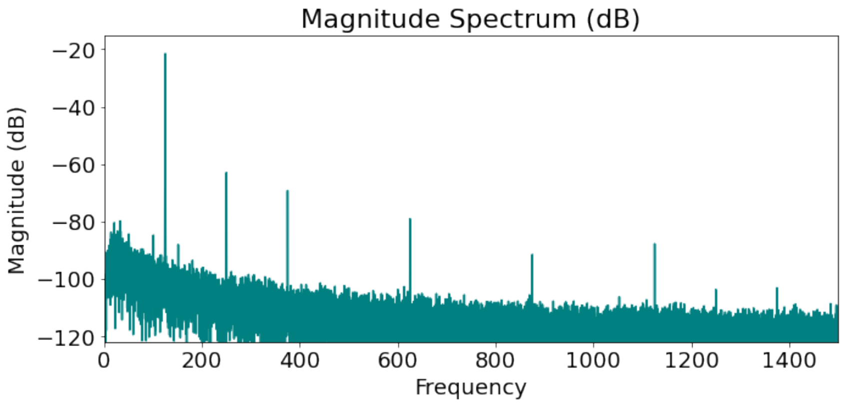
Figure A.1: Intrinsic noise recorded on all four channels.



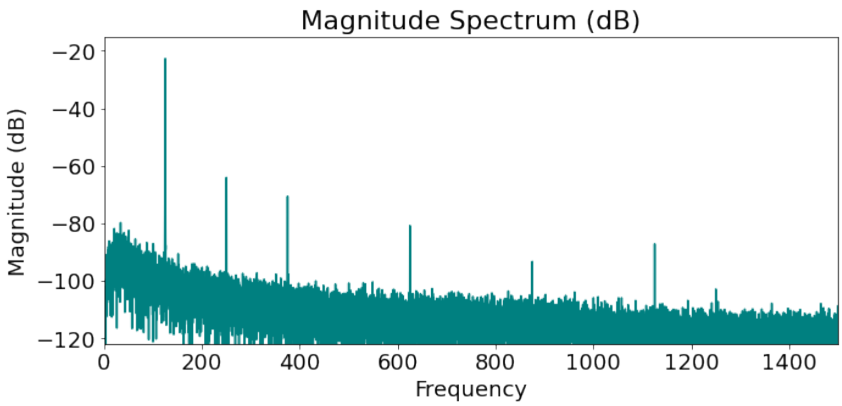
(a) 125 Hz signal recorded on channel A



(b) 125 Hz signal recorded on channel B

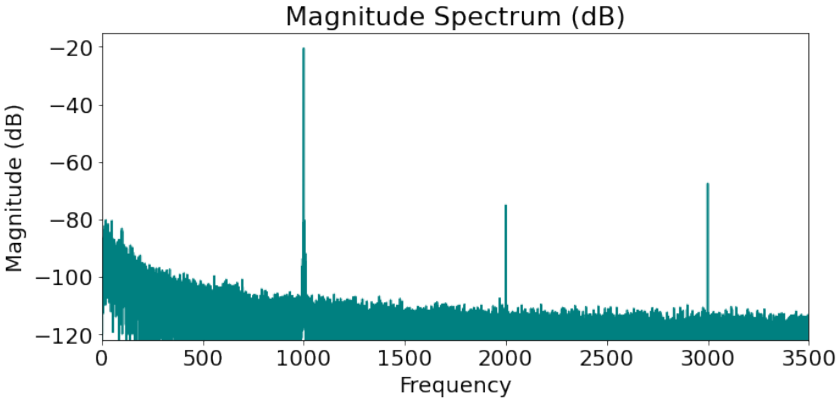


(c) 125 Hz signal recorded on channel C

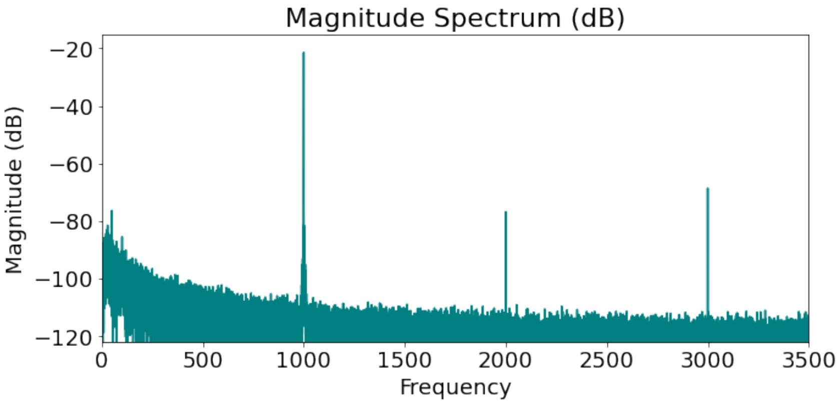


(d) 125 Hz signal recorded on channel D

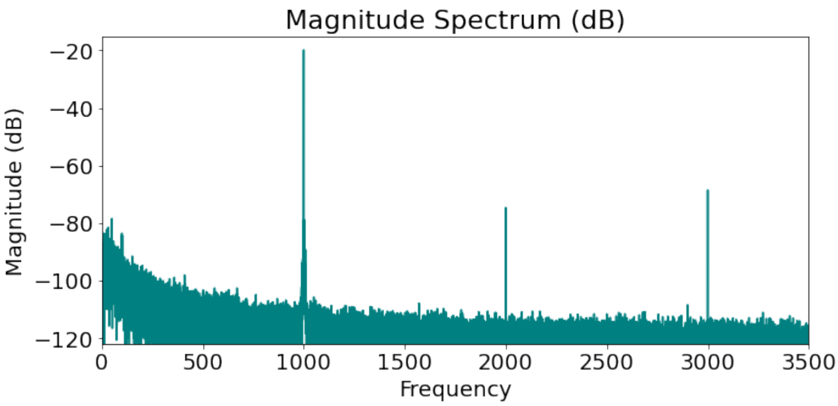
Figure A.2: 125 Hz signal recorded on all four channels.



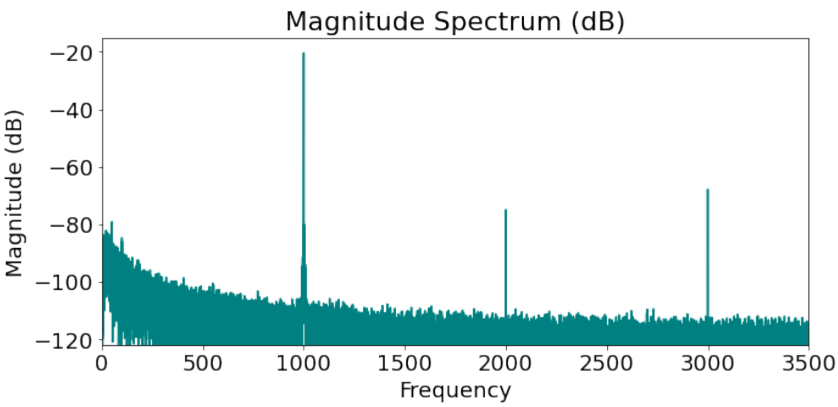
(a) 1 kHz signal recorded on channel A



(b) 1 kHz signal recorded on channel B

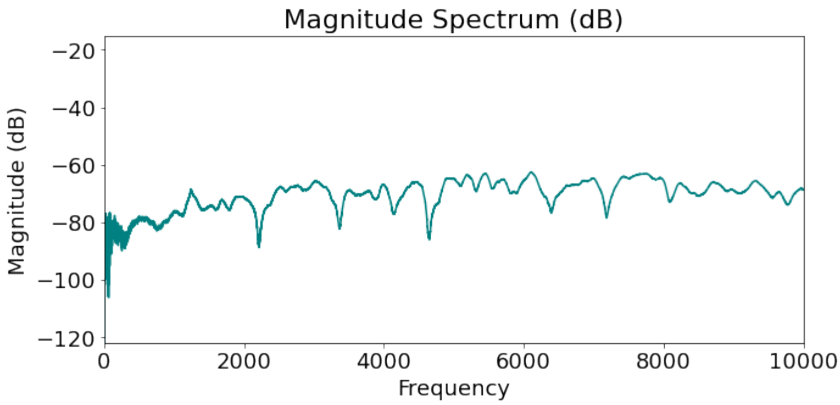


(c) 1 kHz signal recorded on channel C

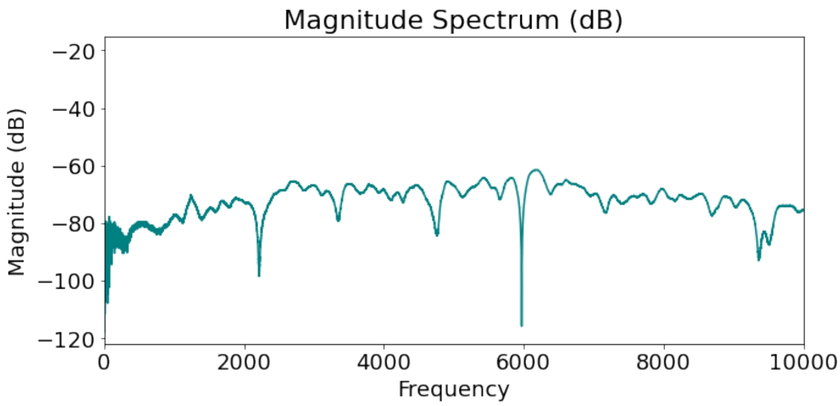


(d) 1 kHz signal recorded on channel D

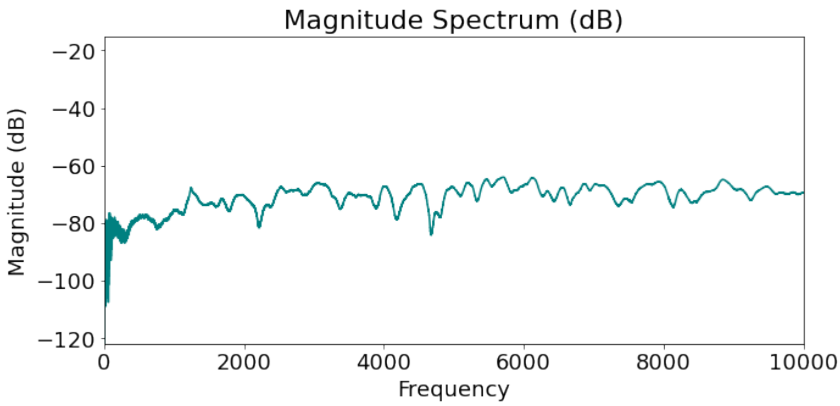
Figure A.3: 1 kHz signal recorded on all four channels.



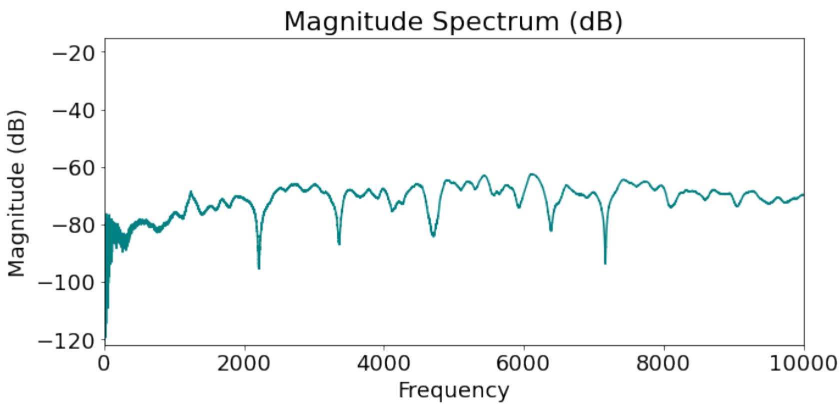
(a) Magnitude spectrogram of recorded log sine sweep on channel A



(b) Magnitude spectrum of recorded log sine sweep on channel B

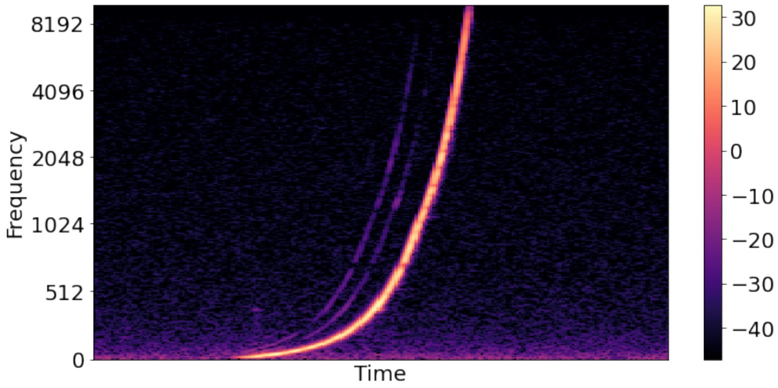


(c) Magnitude spectrum of recorded log sine sweep on channel C

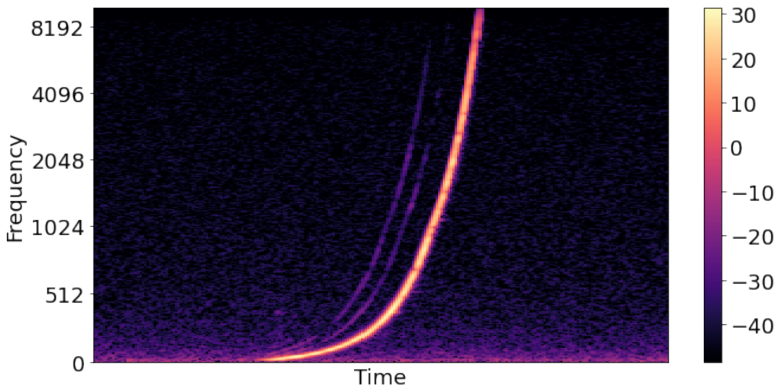


(d) Magnitude spectrum of recorded log sine sweep on channel D

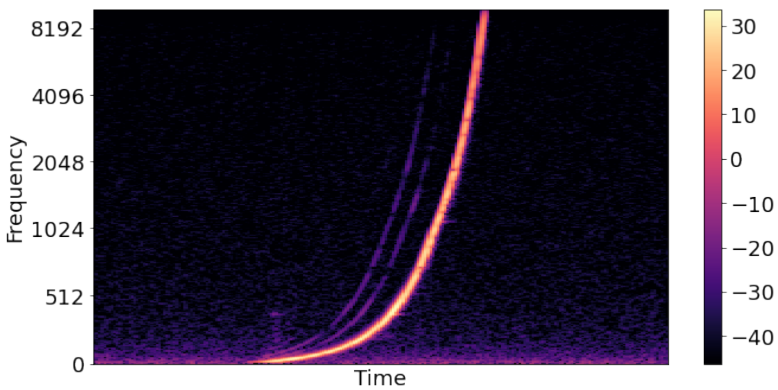
Figure A.4: Magnitude spectrum of recorded log sine sweep on all four channels.



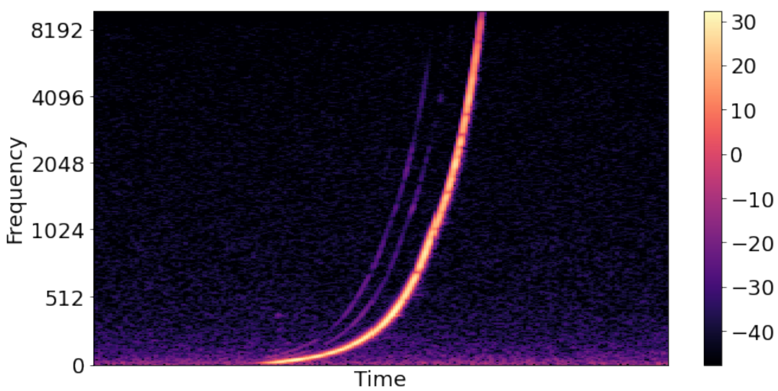
(a) Spectrogram of recorded log sine sweep on channel A



(b) Spectrogram of recorded log sine sweep on channel B



(c) Spectrogram of recorded log sine sweep on channel C



(d) Spectrogram of recorded log sine sweep on channel D

Figure A.5: All four channel of recorded Logarithmic sine sweep plotted as a spectrogram with MEL-scaling on the y-axis.

Appendix B

Bill of Materials (BOM) and schematic

In figure B.1 the entire schematic of the pre-amplifier is shown under.

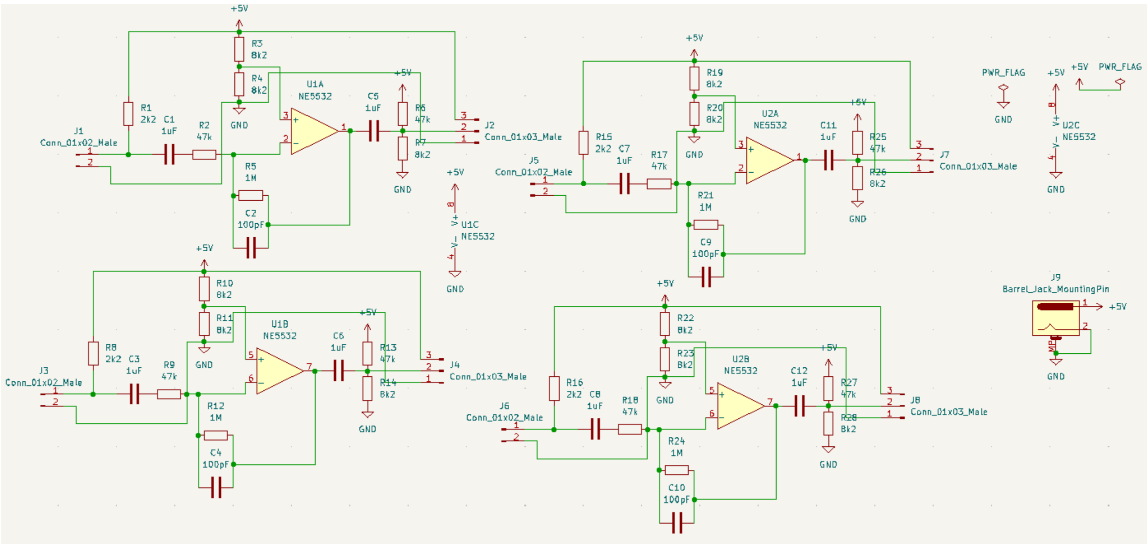


Figure B.1: Full schematic of the pre-amplifier.

Below is selected colums from the generated BOM file from KiCad, showing the component, value and reference.

Reference(s)	Value	Component
C1	1uF	Capacitor
C2	100pF	Capacitor
C3	1uF	Capacitor
C4	100pF	Capacitor
C5	1uF	Capacitor
C6	1uF	Capacitor
C7	1uF	Capacitor
C8	1uF	Capacitor
C9	100pF	Capacitor
C10	100pF	Capacitor
C11	1uF	Capacitor
C12	1uF	Capacitor
J1	Conn_01x02_Male	Conn_01x02_Male
J2	Conn_01x03_Male	Conn_01x03_Male
J3	Conn_01x02_Male	Conn_01x02_Male
J4	Conn_01x03_Male	Conn_01x03_Male
J5	Conn_01x02_Male	Conn_01x02_Male
J6	Conn_01x02_Male	Conn_01x02_Male
J7	Conn_01x03_Male	Conn_01x03_Male
J8	Conn_01x03_Male	Conn_01x03_Male
J9	Barrel_Jack_MountingPin	Barrel_Jack_MountingPin
R1	2k2	Resistor
R2	47k	Resistor
R3	8k2	Resistor
R4	8k2	Resistor
R5	1M	Resistor
R6	47k	Resistor
R7	8k2	Resistor
R8	2k2	Resistor
R9	47k	Resistor
R10	8k2	Resistor
R11	8k2	Resistor
R12	1M	Resistor
R13	47k	Resistor
R14	8k2	Resistor
R15	2k2	Resistor
R16	2k2	Resistor
R17	47k	Resistor
R18	47k	Resistor
R19	8k2	Resistor
R20	8k2	Resistor
R21	1M	Resistor
R22	8k2	Resistor
R23	8k2	Resistor
R24	1M	Resistor
R25	47k	Resistor
R26	8k2	Resistor
R27	47k	Resistor
R28	8k2	Resistor
U1	NE5532	Amplifier
U2	NE5532	Amplifier

NTNU

Norwegian University of
Science and Technology