Master's thesis

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

Khadija Laajab

Detection of Bowel Sounds using machine learning and the ability to use it in early meal detection

Master's thesis in Electronic Systems
Supervisor: Dag Roar Hjelme
Co-supervisor: Salman Ijaz Siddiqui
June 2022

NTNU
Norwegian University of
Science and Technology

Khadija Laajab

Detection of Bowel Sounds using machine learning and the ability to use it in early meal detection

Master's thesis in Electronic Systems
Supervisor: Dag Roar Hjelme
Co-supervisor: Salman Ijaz Siddiqui
June 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

This thesis describes the implementation of a bowel sound detector using machine learning and the ability to use it in early meal detection. Such a method can be used to optimize advanced glucose meters, such as continuous glucose monitoring (CGM) for diabetic patients so that there will be no need for meal announcements, which are necessary today. The data sets used in this thesis were openly available recordings from Youtube and recordings provided by the Artificial Pancreas Trondheim (APT) group. It was implemented two classifiers with different data sets; the first one was taken as a starting point which contained data from youtube and the second one added collected data to the first created data set. Mel-scaled spectrograms were used as features, the CNN was chosen as a classifier, and parameter tuning was done to find the best possible performance. The models were tested on the test set and on the collected recordings which presents a more 'real condition' to evaluate its performance. The final created model managed to identify the classes, BS (bowel sound) and NBS (non-bowel sound) with an AUC of 95%. This supports previous findings that suggest the feasibility of distinguishing between bowel sound and non-bowel sound. However, the detector has not been evaluated on contaminated recordings and should be focused on in further work.
The final implemented bowel sound detector was used to analyze further collected recordings when different subjects followed a given protocol. Typically, the detected bowel sound was more frequent right before or/and during a meal which supports human physiology. Also, the acoustic features spectral centroid (SC), spectral bandwidth (SBW), and the duration of the detected bowel sound were extracted from the different states followed in the protocols. It was not observed a particular trend in SC and SBW during the fasting and eating period. However, it was observed often, that the duration of the detected bowel sounds increased right before or/and during the meal period. The same acoustic analysis was done to see if there was possible to differentiate between a hard meal and a soft meal. The total duration of the detected bowel sounds per minute was longer when the subjects ate a soft meal. Also, the SC and SBW tend to be higher when the subjects ate a hard meal and the recordings were collected from the left lower quadrant (LLQ). Some of the subjects, had an increase in the occurrence and duration of detected bowel sounds while watching food videos, meaning there is a need for further acoustic analysis to differentiate between the bowel sounds found in this state and the ones during the meal.

Sammendrag

Denne oppgaven beskriver implementering av en tarmlyddetektor ved bruk av maskinlæring og muligheten til å bruke det i tidlig måltidsdeteksjon. En slik metode kan brukes til å optimalisere avanserte glukosemålere, som for eksempel kontinuerlig glukosemåling (CGM) for diabetespasienter, slik at det ikke er behov for måltidskunngjøring, som er nødvendig i dag. Datasettene som ble brukt i denne oppgaven var åpent tilgjengelige opptak fra youtube og opptak levert av APT-gruppen. Det ble implementert to læringsmodeller med forskjellige datasett; den første ble tatt som utgangspunkt som inneholdt data fra youtube og den andre la innsamlet data til det første opprettede datasettet. Mel-skalert spektrogrammer ble brukt som egenskaper (eng. features), konvolusjonalt nevrale nettverk (CNN) ble valgt som læringsalgoritme og parameterjustering ble gjort for å finne best mulig ytelse. Modellene ble testet på testsettet og på innhentet opptak som presenterer en mer "reell tilstand" for å evaluere ytelsen. Den endelige modellen klarte å identifisere klassene BS og ikke-BS med en areal under "receiver operating characteristic curve" (AUC) på 95%. Dette støtter tidligere funn som antyder muligheten for å skille mellom tarmlyd og ikke-tarmlyd. Detektoren er imidlertid ikke evaluert på kontaminerte opptak og bør fokuseres på i videre arbeid.
Den endelige implementerte tarmlyddetektoren ble brukt til å analysere ytterligere innsamlede opptak når forskjellige forsøkspersoner fulgte en gitt protokoll. Vanligvis var den detekterte tarmlyden hyppigere rett før eller/og under et måltid som støtter menneskets fysiologi. Dessuten ble de akustiske egenskapene spektralsentroide (SC), spektral båndbredde (SBW) og varigheten av den detekterte tarmlyden hentet ut fra de forskjellige tilstandene som ble fulgt i protokollene. Det ble ikke observert en spesiell trend i SC og SBW under faste- og spiseperioden. Imidlertid ble det observert ofte at varigheten av de oppdagede tarmlydene økte rett før eller/og under måltidsperioden. Den samme akustiske analysen ble gjort for å se om det var mulig å skille mellom et hardt og et mykt måltid. Den totale varigheten av de påviste tarmlydene per minutt var lengre når forsøkspersonene spiste et mykt måltid. SC og SBW hadde også en tendens til å være høyere når forsøkspersonene spiste hardt måltid og opptaket ble samlet inn fra venstre nedre kvadrant (LLQ) av abdomen. Noen av forsøkspersonene hadde en økning i forekomsten og varigheten av detektert tarmlyd mens de så matvideon, noe som betyr at det er behov for ytterligere akustisk analyse for å skille mellom tarmlydene funnet i denne tilstanden og de under måltidet.

Preface

This thesis is part of the Master of Science degree in the field of Engineering Electronic Systems at the Norwegian University of Science and Technology (NTNU). The study was suggested by the APT research group which provided the data set containing recordings as WAV files. My supervisor, Dag Roar Hjelme, and co-supervisor, Salman Ijaz Siddiqui, handed relevant background literature on early meal detection and theory about bowel sounds and diabetes during the initial stages of my study. In addition, my co-supervisor participated weekly meetings where we discussed my ideas, and he shared his expertise in acoustic signals and digital signal processing. They both suggested a report structure.

I would like to thank my co-supervisor, Salman Ijaz Siddiqui, for his valuable feedback and guidance during the semester. Finally, I want to thank my family and friends for their never-ending support during my time as a student in Trondheim.

Nomenclature

AI	Artificial Intelligence
ANN	Artificial Neural Network
APT	Artificial Pancreas Trondheim
AUC	Area Under the receiver operating Characteristic
BGL	Blood Glucose Level
BP	Backpropagation
CGM	Continuous Glucose Monitoring
CRS	Continuous Random Sound
CSII	Continous Subcutaneous Insulin Infusion
DFT	Discrete Fourier Transform
FFT	Fast Fourier Transform
FN	False Negative
FP	False Positive
HS	Harmonic Sound
LLQ	Left Lower Quadrant
LUQ	Left Upper Quadrant
MB	Multiple Burst
MBGD	Mini Batch Gradient Descent
MFCC	Mel-frequency cepstral coefficients
NTNU	Norwegian University of Science and Technology
PNCC	Power-normalised Cepstral Coefficient
ReLU	Rectified Linear Unit
RLQ	Right Lower Quadrant
RMS	Root Mean Square
RNN	Recurrent Neural Networks

ROC	Receiver Operating Characteristic
RUQ	Right Upper Quadrant
SB	Single Burst
SBW	Spectral Bandwidth
SC	Spectral Centroid
SGD	Stochastic Gradient Descent
SNR	Signal-to-Noise Ratio
STFT	Short-time Fourier Transform
SVM	Support Vector Machine
TN	True Negative
TNR	True Negative Rate
TP	True Positive

Contents

1 Introduction

1.1 Background

1.1.1 Diabetes mellitus

Diabetes mellitus is a group of metabolic diseases that are caused by the production of the hormone insulin being reduced or completely absent. The production of insulin happens in the pancreas and allows the cells in the muscles, fat, and liver to absorb glucose that is in the blood [1]. If the production is reduced, this can result in too high blood glucose, also known as hyperglycemia, and cause health problems such as the risk of heart disease, nerve problems, kidney disease, and vision problems [2].
The most common types of diabetes are type 1 diabetes mellitus (DM1), type 2 diabetes mellitus (DM2) and gestational diabetes. DM1 occurs often unexpectedly at a young age and is a condition that happens when the body's immune system attacks the cells in the pancreas that are involved in producing insulin. This results in insufficient production of the hormone, so people with this condition usually need to take insulin every day to stay alive. DM2 is the most common type and is diagnosed when the body does not make or use insulin well. This condition is associated with obesity and a poor diet high in fat, calories, and cholesterol. This condition usually develops in adults over the age of 45, but can also occur in younger age groups. The treatment of this condition involves changing to a healthier lifestyle along with insulin medications. Gestational diabetes is a condition that usually occurs during pregnancy, usually in the second and third trimesters. The need for insulin is usually high during pregnancy and if the body is not able to produce enough insulin, the blood glucose can be too high and lead to complications for mother and child. This condition is treated through a combination of regular physical activity, meal plans, and including insulin injections if needed. The aim is to keep the blood glucose levels of pregnant women at the same level as those who do not have this disorder. Often, the condition disappears after birth [3].

1.1.2 DM1 and DM2 treatments

People who are suffering from DM1 and some who have DM2 are dependent on exogenous insulin to control their blood glucose level (BGL). The insulin has to be injected into the blood by using a fine needle or insulin pen, it can not be eaten as it interferes with stomach enzymes. The BGL has to be monitored and controlled, there are various treatments for doing this. The BGL can be monitored by pricking the finger to get a small drop of blood. The results can be read using an electronic glucose meter which tells in what range the BGL are and if there is a need to inject insulin. How often the insulin has to be injected

differs from person to person, some people need at least 2 insulin shots in a day, others need 3 or 4 [4]. This manual insulin therapy is time-consuming and can be stressful for the patients.

Another treatment is the use of an insulin pump, which contains a small device carried on the body and a tube that connects it to a catheter. The device is set to deliver insulin in small doses every hour throughout the day. This simplifies insulin treatment by reducing the use of syringes and making it easier to measure blood sugar within a day. The pumps have a bolus calculator that calculates the recommended insulin intake. Patients must enter how many carbohydrates he/she eats and their blood sugar. Also, they need to insert insulin each time the glucose level is too high, or they have eaten a meal, by pressing a button on the device [5].

An advanced diabetes treatment that is more effective, is the artificial pancreas (AP). The AP consists of a Continuous glucose monitoring (CGM), a process control algorithm and a continuous subcutaneous insulin infusion (CSII) unit. The first component, CGM, continuously monitors a person's blood glucose levels every few minutes and is placed in the subcutaneous tissue. The second component, the control algorithm, calculates how much insulin is needed and alerts the insulin pump when it needs to be delivered. The algorithm decides the amount of insulin that should be injected based on the CGM value along with various static parameters such as the body weight, insulin sensitivity, and the glucose target range. The last component, insulin infusion delivers the insulin when the blood glucose levels are not in the target range [6]. However, the glucose kinetics and insulin absorption are slow in subcutaneous tissue, leading to limitation of glucoregulatory outcomes and causing larger postprandial glucose excursions. Therefore, patients are required to announce whenever having a meal.

1.2 Motivation

Despite the advantages of advanced diabetes treatments, the patient's involvement in the therapy is still important, which can be a burden. Another limitation of CGM systems is that it takes about 30 to 40 minutes from the beginning of a meal until it is detected which can cause hyperglycemia. Therefore, meal announcements are required by clinically tested systems for glucose control. This can cause mental occupation about the disease for the patient. An automated meal detection system could help minimize the time between a meal start and when the system has noticed it and remove the need for meal announcement.

For centuries, physicians have used gut noises to diagnose gastrointestinal conditions. The intensity, location, and frequency are considered important factors in assessing a condition [7]. Gut sounds higher than 900 Hz can indicate intestinal obstruction, a phenomenon that does not occur in patients with normal conditions.

Earlier research has used different machine learning algorithms to detect bowel sounds. A study from Tsinghua University in 2018 developed a bowel sound recognition using support vector machine (SVM) classification [8]. The results indicate that the proposed method provides greater than 90% accuracy and specificity, and greater than 85% sensitivity. A study from Tokushima University developed an automatic bowel sound detector when using an artificial neural network (ANN) [9]. The proposed method achieved an accuracy of approximately 90%. However, for both studies, when testing the model in an environment containing noise and machinery sounds the results degrades. Another study focused on LSTM classifier [10], the proposed method achieved a sensitivity of 90.92% and total accuracy of 92.56%. However, when testing in a different environment, the sensitivity was affected and dropped to 62%. A newer study from Tsinghua University in 2020 developed a bowel sound detector using CNN [11]. The results were promising, with an accuracy of higher than 90%. The only downside is that the segment has to be at least 1 second long. The most common type of bowel sound are 20-40 ms long and will not be detected. Recent research from Australia 2022 a bowel sound detector was developed by Wang [12]. This research showed that the detector was effective in detecting all types of bowel sounds and provides an accurate time stamp for each type of BS. However, it was not investigated thoroughly the possibility of using this in an early meal detection. Earlier studies such as Kölle [13] has shown promising results with an average meal detection time of 10 minutes. However, the system produced many false positives.

This present study aims to design a bowel sound detector and study if the activity and the features of the detected bowel sounds are changing significantly enough shortly after meal intake and during a meal to allow reasonably early detection for use in an artificial pancreas. Also, the bowel sound detector would be used in the analysis for meal differentiating and meal simulation.

1.3 Objective

This master's thesis implements a bowel sound detector inspired by Wang [12] and takes it a step further to see the feasibility of using such a detector in early meal detection. Two different data sets were generated: bowel sound from a

youtube channel and added collected data. For the data acquisition, usually, four microphones were used to capture bowel sounds, swallowing and chewing sounds. The latter two recordings were used for another work and therefore discarded here.
Specifically, this study consists of the following work:

- Label bowel sounds and non-bowel sounds in different recordings to generate useful data sets.
- Implementation of a bowel sound detector using machine learning. This will include processing raw data, feature extraction, and finding the best parameter combination to create an optimal model.
- Evaluation of the bowel sound detector using both: the test set and the collected data presenting a 'real condition'.
- Analyse collected data using the final implemented bowel sound detector. This includes extracting different acoustic features to look at the feasibility of early meal detection and distinguishing between meal type and how it performs on meal simulation.

1.4 Artificial Pancreas Trondheim

This study is written in collaboration with Artificial Pancreas Trondheim (APT). This is a research group based at the Norwegian University of Science and Technology (NTNU) at Trondheim and consists of people with high competence in various fields. These can be control engineering, biomedical engineering, biosensors, applied clinical research, endocrinology, anesthesia, intensive care medicine, pharmacology, biotechnology, mathematical modeling, biochemistry, and chemometrics [14]. The goal is to develop a robust artificial pancreas that can support patients with DM1 and DM2 diabetes.

1.5 Outline of the master's thesis

This master's thesis is divided into 11 chapters. First, an introduction is given including its motivation and objectives. Further, Chapter 2 presents relevant theory on bowel sound characteristics, preprocessing audio signals, feature extraction and machine learning methods. Chapter 3 describes the two different data sets that are used in this thesis. Next, Chapter 4 describes an attempt of implementing a bowel sound detector using only data from youtube. Chapter 5 adds the collected recordings by the ATP group to the data set and retrains the model created in Chapter 4. In Chapter 6, the latest implemented bowel sound

detector is used to detect the bowel sounds, various acoustic features are extracted to look at the possibility of early meal detection. Chapter 7 is studying the feasibility of differentiating two types of a meal when extracting the same features as in the previous section. Chapter 8 is focusing on if there are any detected bowel sounds and their acoustic features when different subjects watch food videos. Finally, Chapter 9 provides an overall discussion of the thesis before a conclusion is given in Chapter 10, and suggestions for future work are provided in Chapter 11.

2 Theory

2.1 Bowel Sound

Bowel sounds are typically created when food, liquids or gas are passing through the gut. The hollowness of the intestines causes these sounds to propagate through the stomach, which can be reminiscent of the sound of water pipes[15]. These sounds can occur from the entire abdomen which is typically divided into four different areas named right upper quadrant (RUQ), left upper quadrant (LUQ), left lower quadrant (LLQ) and right lower quadrant (RLQ) as shown in Figure 2.1. These sounds may be classified as normal, hypoactive, or hyperactive. Hypoactive is when the bowel sounds are almost absent and occurs normally during sleeping or right after abdominal surgery [16]. On the other side, hyperactive is when bowel sounds happen frequently. This occurs normally after eating a meal when the food is leaving the stomach and enters the small intensities in LUQ as it uses muscles to move the food until it enters the large intestine in RLQ. The sounds should continue as the large intestines absorb water and nutrients and are responsible for pushing the food [17].

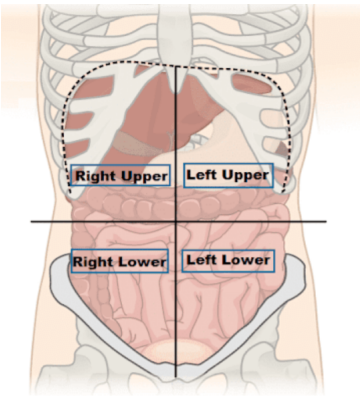
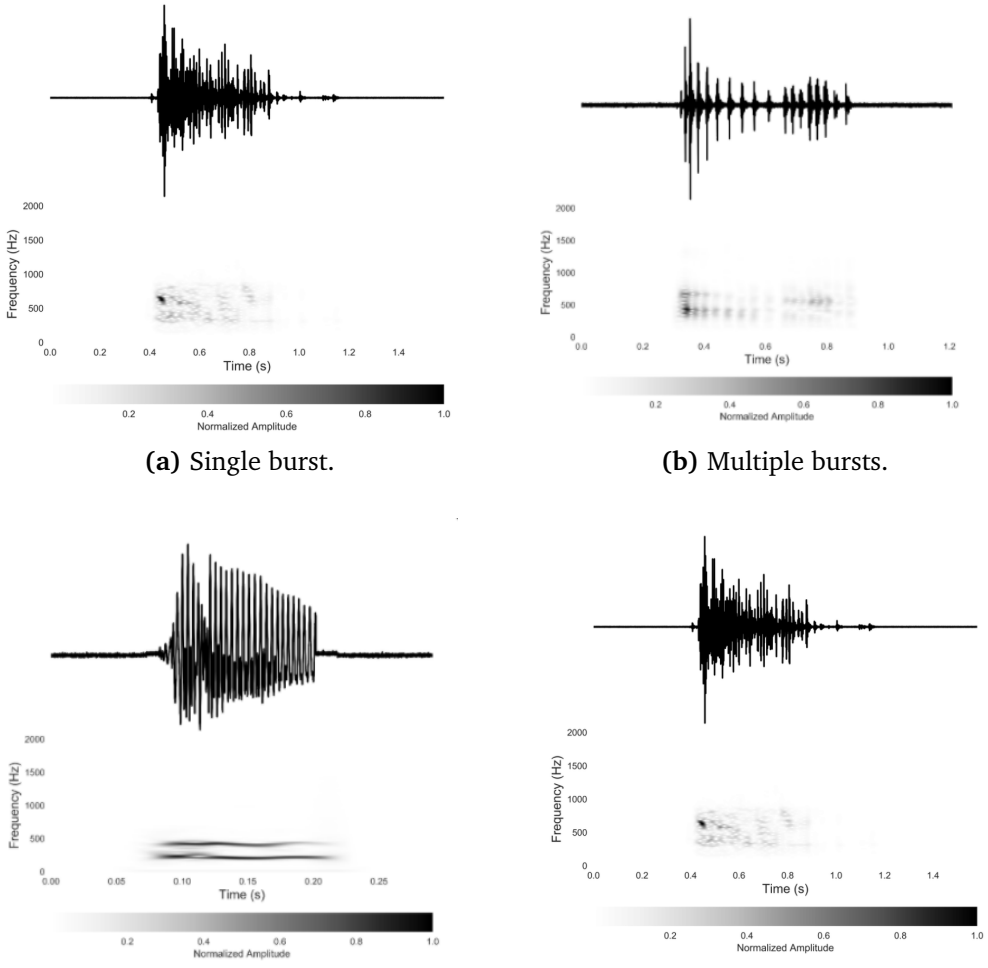


Figure 2.1: The four abdominal quadrants. Source: [18]

2.2 Charecterstics of Bowel Sound

Bowel sounds occur in an irregular pattern and according to earlier research [12], these are typically divided into subtypes, including single burst (SB), multiple bursts (MB), continuous random sound (CRS), and harmonic sound (HS). Figure 2.2 presents these different types found by the study [19] with the corresponding spectrogram on the bottom, the characteristics are quite different. SB is a simple pulse shown in Figure 2.2a, that can be caused by a single contraction in the bowel muscle and lasts for only 10-30 ms [20]. The typical frequency range lies between 200 Hz to 1000 Hz. MB shown in Figure 2.2b can

be described as multiple consecutive SB with a shorter interval time between adjacent components. The duration is therefore much longer, ranging from 40-1500 ms, and has the same frequency range as SB. It has also previously been reported that SB and MB are the most common types of bowel sound [19]. HS shown in Figure 2.2c, are rhythmic noises and can be described as whistling sweeps with a typical duration of 50-1500 ms and have one to multiple of frequency components. The highest frequency found in the study [19] was up to 3000 Hz. On the other hand, CRS shown in Figure 2.2d, does not have any pattern or clear rhythm and is therefore often recognized as a random sound. This subtype is also difficult to distinguish from background noise as the sound is random. Also, the duration lasts over long periods ranging from 200-400 ms without silent periods.



(a) Single burst.

(b) Multiple bursts.

(c) Harmonic sound.

(d) Continous random sound

Figure 2.2: Different types of bowel sounds. Source: [19]

2.3 Data preprocessing

Data preprocessing is an operation that converts raw data into data that contains relevant and understandable information for the system. It is an important part of a machine learning system as it affects the ability of the model to learn by for example removing information that is not necessary [21]. In this section, different types of relevant preprocessing techniques will be presented.

2.3.1 Quantization

In digital signal processing, quantization is a process that involves mapping values from a large set (continuous values) to a smaller set (discrete, predetermined levels). When a signal is quantized, its bit depth, known as the number of bits per sample, is reduced. This makes it difficult to output the values from continuous values and leads to rounding the values in the signal which reduces the sharpness of the signal. This operation gives a lower Signal-to-noise ratio (SNR), as a result of cutting some of the signal's highest peak [22].

2.3.2 Downsampling

If the relevant information lies in a lower frequency range, then downsampling can be done without the loss of important information. Also, the processing will become faster as the samplings frequency is lower, which makes the processing time more effective. However, it is important to fulfill Nyquist Sampling Theorem to avoid aliasing. The theorem states that the sampling frequency must be selected more than 2 times the highest frequency component in the signal,

$$f_{nyq} = 2 \cdot f_{high} \tag{1}$$

where f_{high} is the highest frequency component in the signal [23].

2.3.3 Digital filter

Filtering audio signals in preprocessing means often removing unwanted noise, reducing the processing time, or keeping a wanted frequency band.

A digital filter in signal processing refers to suppressing unwanted components on the input frequencies [24]. The frequency response, $H(w)$, of the filter, determines which frequencies will be removed. There are different types of filters such as low-pass, high-pass and band-pass filters. It is the threshold, f_c that determines which frequencies should be removed. The low-pass and high-pass filters attenuate high and low frequencies, respectively. A band-pass filter passes frequencies within a certain range and rejects frequencies outside this range. An ideal filter has a maximum gain at the passband region and zero

gain in the stopband region. An ideal high-pass filter is shown in Figure 2.3, y-axis shows the amplitude response $|H(w)|$ and x-axis shows the frequencies. However, these filters are not possible to implement in reality as they are discontinuous at the cutoff frequency. The response of practical filters would be continuous at the cutoff frequency resulting in having a passband and a stopband that are not perfectly flat. To make the different filters different designs can be used, one common type is the Butterworth filter which is designed to have a frequency response that is as flat as possible in the passband. Different orders result in how flat the stopbands are. Figure 2.4 shows the response of different orders of the Butterworth low-pass filter together with the ideal filter in black.

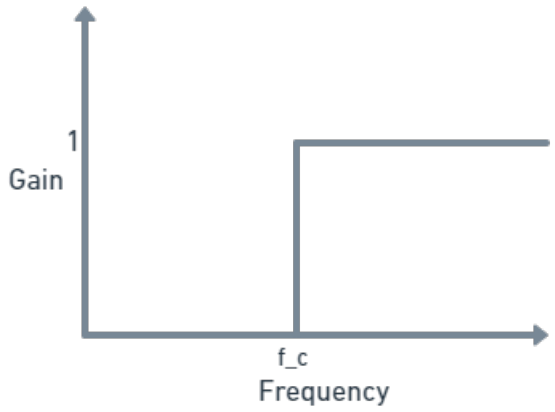


Figure 2.3: Amplitude response of an ideal high-pass filter.

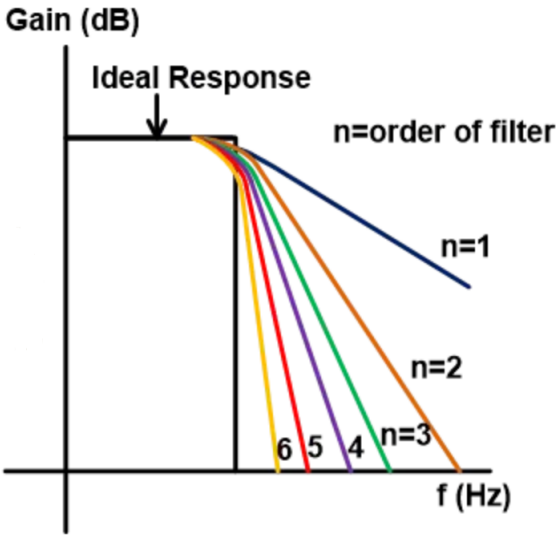


Figure 2.4: Different orders of Butterworth low-pass filter. Source: [25]

2.3.4 Normalization

When having different audio signals, the amplification can vary. To be able to combine the different value sets, the data has to be scaled without distorting differences in the ranges of values or losing information [26]. One approach is to map the recordings to the range [-1,1], the relative information within a signal is then considered, rather than their absolute amplitudes.

2.4 Feature representations

Audio signal recognition can use different types of quantities, called features, to describe the signal. A feature is an individual measurable property or characteristic of a phenomenon [27]. These features can describe the difference across classes better and make the classification better. This section will present different types of feature representation.

2.4.1 Fourier Transform

A complex audio signal contains multiple frequency components. These can be found by a mathematical expression called Fourier transform. For continuous signals, this is obtained by

$$X(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft}\, dt \tag{2}$$

where $X(f)$ is the Fourier transform, f is the frequency of the signal and $x(t)$ is the time-varying signal.

2.4.2 Discrete Fourier Transform

As the computers only work with discrete signals, there will be a need for Discrete Fourier Transform (DFT). Assume having a discrete signal $x(n)$ with the length of $n = 0, ..., N-1, N$. The DFT of the signal will be

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{N}kn}, k = 0, ...N-1 \quad (3)$$

where $X(k)$ is the DFT of the input signal $x(n)$. DFT can be calculated using the algorithm, Fast Fourier Transform (FFT).

2.4.3 Short-Time Fourier Tranform

Non-periodic signals vary over time and the standard Fourier Transform provides the frequency information averaged over the entire signal time interval. There is a need for another operator that provides the time-localized frequency content of the signal. Short-Time Fourier Transform (STFT) is computed by participating the signal into shorter overlapping segments of equal length. The Fourier transform is then computed in each separate segment [28]. This can be obtained mathematically by

$$X(m, n) = \sum_{k=0}^{L-1} x(k) g(k-m) e^{-j\frac{2\pi}{N}kn}, k, m = 0, 1..., N-1 \quad (4)$$

where $x(k)$ denotes a signal and $g(k)$ denotes an L-point window function.

The computed FFTs can be stacked on top of others and form a spectrogram. Figure 2.5 an example of how the procedure is done. A spectrogram is a visual representation of the strengths of a signal over time [29]. It allows one to visualize the energy levels that vary depending on the frequency of the signal. These are typically two-dimensional graphs, the x-axis and y-axis show the time and the log scale of the frequency, respectively, and with a third variable, amplitude in decibels represented by colors. This can be used to for example distinguish between different types of sounds/vibrations and determine how the frequency changes over time.

2.4.4 Mel-scaled spectrogram

Mel-scale is a nonlinear transformation of the frequency scale. The idea behind the mel-scale is that it reflects how humans hear differences in frequencies. Sounds of equal distance on the mel-scale are perceived to be of equal distance to humans. Humans are better at noticing differences in lower frequencies than

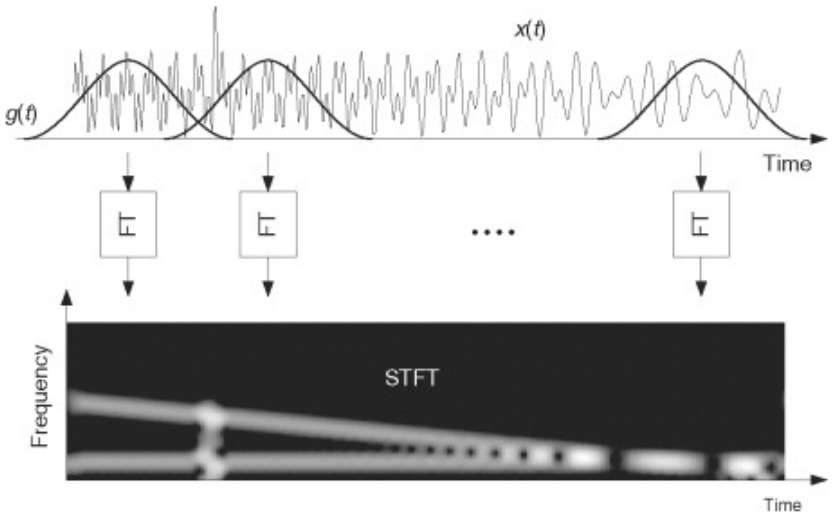


Figure 2.5: An example of how the STFT are computed and realized in a spectrogram. Source: [30]

in higher frequencies, for example, most can tell the difference of sounds in frequencies between 300 Hz and 400 Hz, but for larger frequencies such as 1100 Hz and 1200 Hz it is much harder [31]. The transformation from frequency to mel-scale is

$$mel(f) = 1127 \cdot log(1 + \frac{f}{700}) \tag{5}$$

The plot of the Mel-scale formula 5 is shown in Figure 2 the x-axis is the frequency, ranging from 0 to 25kHz and the y-axis shows the mel-scale. A mel-

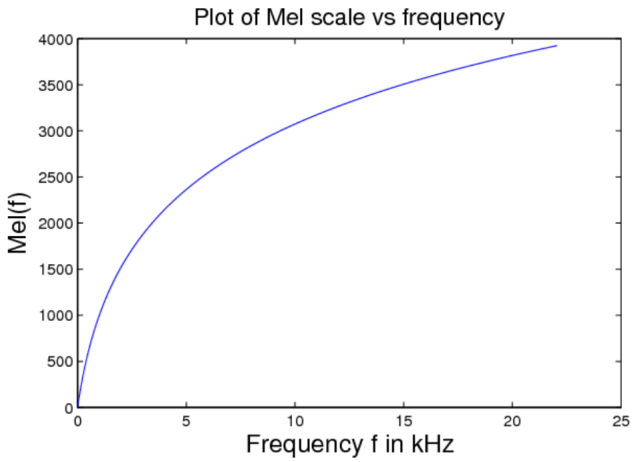


Figure 2.6: Plot of mel-scale function vs frequency.

scaled spectrogram is a spectrogram where the frequencies are converted to

the mel-scale on the y-axis [32]. The function *melspectrogram* from package [33] does this. The inputs *y*, *sr*, *n_fft*, *hop_length* and *center* indicate the audio signal, sampling rate of the audio signal, window length of the short segment when computing STFT, the number of samples between each successive segments and where the frame begins, respectively. If the center is False, the frame t begins at y[t*hop_length] while if the center is True, the signal is padded so that frame t is centered at y[t*hop_length] [34]. The function returns the spectrogram in the shape (n_mels, t).

2.4.5 Spectral Centroid

The spectral centroid (SC) indicates the center mass of the spectrum and is described in [35] "the center of gravity of the amplitude spectrum of a signal" and defined as

$$S_{centroid} = \frac{\sum_{k=0}^{\frac{K}{2}-1} kA(k)}{\sum_{k=0}^{\frac{K}{2}-1} A(k)} \tag{6}$$

where $A(k)$ is the spectrum amplitude given a time window so, $A(k) = |X(k)|$. X(k) is the DFT.

2.4.6 Spectral Bandwidth

The spectral bandwidth (SBW) is described in [35] "the extent of the power transfer function around the center frequency" and defined as

$$B_{rms} = \sqrt{\frac{\sum_{k=0}^{\frac{K}{2}-1} (k - S_{centroid})^2 A^2(k)}{\sum_{k=0}^{\frac{K}{2}-1} A^2(k)}} \tag{7}$$

where $A(k)$ is the spectrum amplitude given a time window so, $A(k) = |X(k)|$. X(k) is the DFT.

2.5 Machine Learning

Machine learning is a subfield of artificial intelligence (AI) that focuses on learning how to fit data into models that can be utilized by humans. The various tasks that are performed in machine learning are generally classified into broad categories based on how the learning is received and how feedback is given to the system. There are different types of machine learning algorithms where the two most widely adopted methods are supervised learning and unsupervised learning [36]. Supervised learning trains on input data that is labeled by humans/experts while unsupervised allows the algorithm to explore the data without being bound by its labeled data. This master's thesis will focus on supervised machine learning. The other methods are also comprehensively described in the article from Guru99 [37].

2.5.1 Supervised Learning

In supervised learning, the data set is labeled such as both the input vector X and the output vector Y are known for the algorithm. The input X is the features of the observations while the output Y is the label which is the ground truth. The labels are provided by a supervisor which can either be a human or a machine. Table 1 taken from [38] shows five unlabeled data examples that can be labeled based on various judgments. The goal is to find the target function f that fulfills $Y = f(X)$. This can be done by using various machine learning algorithms, if this is found, the function can be used to predict the output of unseen data which is also called a test set. There are two subgroups of supervised learning; regression and classification [39]. The differences between these groups are that the former uses a method to predict continuous values such as age, price, and salary, while the latter focuses on discrete values such as Male or Female, True or False, Spam or Non-Spam. This study will focus on classification problems. The latter method is comprehensively further explained in a tutorial from [36].

Unlabeled Data Example	*Example Judgment for Labeling*	*Possible Labels*	*Possible Supervisors*
Tweet	Sentiment of the tweet	Positive/negative	Human/machine
Photo	Contains house and car	Yes/no	Human/machine
Audio recording	The word football is uttered	Yes/No	Human/machine
Video	Are weapons used in the video?	Violent/nonviolent	Human/machine
X-ray	Tumor presence in X-ray	Present/absent	Expert/machine

Table 1: Unlabeled Data Examples along with Labeling Issues. Table taken from [38].

2.5.2 Artificial Neural Networks

The following paragraph is partly reused from the specialization project [40]:

Artificial Neural Networks (ANNs) are structured networks consisting of real-valued elements called neurons [41]. The network is as described by an article from Forbes [42] "an attempt to simulate the network of neurons that make up a human brain so that the computer will be able to learn things and make decisions in a humanlike manner".

A common type of Neural Network is the Multilayer perceptron (MLP) which contains of multiple layers of perceptrons. A perceptron is a neural network unit that is a precursor to larger Neural Networks [43]. A single layer perceptron is the first proposed neural model created and is shown in Figure 2.7. Here, different inputs $x_1, x_2, ...x_n$ are multiplied with their corresponding weighted variables $w_1, w_2, ...w_n$, summed together with a bias function and applied to a activation function which is the Heaviside step function. This simple model can only learn linearly separable patterns as the activation function is linear. Given an input of $x_1, x_2, ...x_n$ the expression for the output y is

$$y(x_1, x_2, ..., x_n) = \begin{cases} 1, \text{ if } bias + w_1x_1 + w_2x_2 + ... + w_nx_n > 0 \\ -1, \text{otherwise} \end{cases} \tag{8}$$

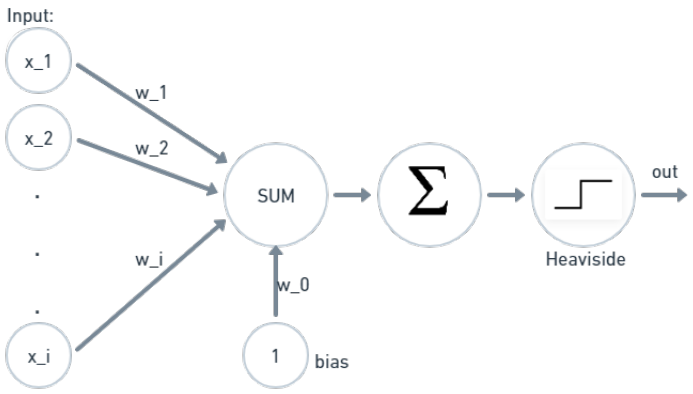


Figure 2.7: A single layer perceptron.

Having multiple single-layer perceptrons in a row results in a layer, and having several of these layers in columns results in ANN. Figure 2.8 shows an example of ANN consisting of one input layer, two hidden layers and one output layer. The input layer is the exposed part of the network and shows how the input is to the network which is typically one neuron per input value or column. The hidden layer consists of multiple perceptrons and is the layers in the middle. It is called hidden because it does not get exposed to the input [43]. The final hidden layer, called the output layer is responsible for outputting a value or vector of values that correspond to the format required for the problem. The number of the input neurons, hidden neurons, and output neurons has to be set before training a neural network. Also, the number of epochs has to be determined. Epoch means training the network with all the data for one cycle. The number of epochs can be determined in the learning process, meaning the process will stop when it reaches a certain iteration. Also, it can be determined by stopping the learning process when it reaches a minimum error value.

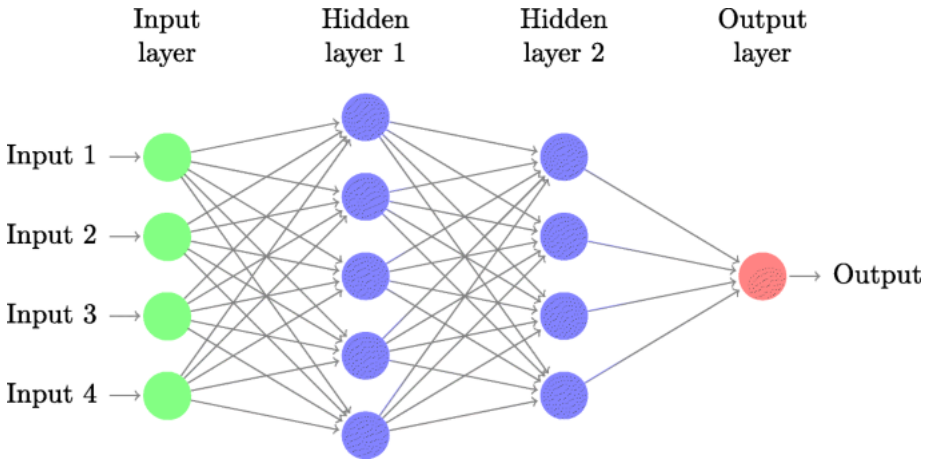


Figure 2.8: Example of an ANN network with two hidden layers. Source:[44]

As mentioned, the single-layer perceptron can only learn linearly separable patterns. However, this is not the case for MLPs as it uses a back propagation (BP) algorithm for training the network. The main processes are feedforward and backpropagation. Feedforward is an algorithm for calculating the output values based on the input values while the BP algorithm changes the weights based on errors obtained from the output values [43]. The learning process is based on looking for the minimum value of the loss function with the respect to the weights by using a technique called gradient descent. The loss function can be computed using the BP algorithms with respect to the weights in the network. This algorithm requires and prefers differentiable activations functions which can be

$$\textbf{Rectified linear unit (ReLU):}\ f(z) = \begin{cases} 0, & \text{if } z \leq 0. \\ z, & \text{otherwise} \end{cases} \tag{9}$$

$$\textbf{Sigmoid:}\ f(z) = \frac{1}{1 + e^{-z}} \tag{10}$$

$$\textbf{Hyperbolic tangent:}\ ReLU(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}} \tag{11}$$

where $z = bias + w_1x_1 + w_2x_2 + ...w_nx_n$. The BP algorithm makes it possible to train the network using gradient descent. This is a method that updates the weights and the bias so the loss function minimizes through the training process. The loss function is a function that returns the cost associated with the model and measures how well the model is doing on the training data [45]. The cross entropy and binary-cross entropy is typically used as the loss function for classification problems with more than two classes and binary classifier, when the number of classes equals two, respectively. The binary-cross entropy is defined as

$$L(y, \hat{y}) = -\frac{1}{N}\sum_{n=1}^{N} y_n log\hat{y_n} + (1 - y_n)log(1 - \hat{y_n}) \tag{12}$$

Where the $\hat{y_n}$ is the predicted probability of being 1 after the network gets fed the input. N is the total samples and y_n is the true label, so either 1 or 0 in the case of binary classification. If the output layer of the network for example uses hyperbolic tangent as the activation function, the prediction is then $\hat{y} = \frac{e^{z_n} - e^{-z_n}}{e^{z_n} + e^{-z_n}}$, where $\mathbf{w} \cdot \mathbf{x_n}$.

The idea of training a network is to minimize the loss of function. As we can see from the equation 12, it is only the weights w that can be changed as the data set is defined and given to the network. Therefore, the goal is to update these values so the global minimum of the function gives the minimum loss. This can be done by finding the steepest direction by derive the loss function

$$\nabla L(\mathbf{w}) = \left[\frac{\partial L}{\partial w_0}, \frac{\partial L}{\partial w_1}, \frac{\partial L}{\partial w_2}, ..., \frac{\partial L}{\partial w_N}\right] \tag{13}$$

This turns to a vector which gives the direction of where the function is steepest. The training rule algorithm will be

$$\mathbf{w} \leftarrow \mathbf{w} + \Delta\mathbf{w} \tag{14}$$

where,

$$\Delta\mathbf{w} = -\eta\nabla L(\mathbf{w}) \tag{15}$$

Where η is a positive constant called learning rate and determines the speed of learning which can be selected to a respectively low value. The gradient descent training rule begins with initiating the value of the weight vector, then calculating the initial loss function, next find δw_i for each weight as in Equation 15 and updates each weight by using the Equation 14. Find the new loss function and repeat until it converges to a weight vector with a minimum loss function. The greater value the learning rate is, the faster the machine is learning but it can overstep the minimum leading to difficulties to get optimum results. On the other hand, the smaller the learning rate, the slower the machine is learning but is a higher chance of getting optimum results [45].

This paragraph is reused from the specialization project [40]:

"From equation 14 and 15, one can see the algorithm considers all the samples which can be time-consuming for large data sets and computationally expensive. There are various training algorithms where the goal is to overcome this problem. Stochastic gradient descent (SGD) considers only one sample to take a single step, while gradient descent considers all of the samples. This leads to calculations per weight update and the computation is less expensive
Mini batch gradient descent (MBGD) is a combination of the previous two. The algorithm selects a set of samples from the training data to get a mini-batch. This mini-batch is presented to the network, and the weights are updated using the mean gradient of the loss function.
The *Adam* optimization algorithm is an extension of SGD. It is an Adaptive Learning System that takes the first and the second moments of the past gradients to determine the learning rate for the weights in the network. The first is the mean and the second moment is the uncentered variance [46]."

2.5.3 Convolutional neural network

A convolutional neural network (CNN) is a type of neural network that focuses on learning from data that has a gridlike structure such as a digital image. A digital image is a representation of a visual image that has a set of values that represents how bright each pixel should be. This can be signals, videos, images, etc. The role of the CNN is to reduce the data into a form that is easier to process, without losing features that are critical for getting a good prediction.
A CNN typically consists of three layers: a convolutional layer, a pooling layer, and a fully connected layer. Figure 2.9 shows the architecture of the network.

The core building block of a CNN is the convolution layer, which carries the main portion of its computational load. This layer performs a dot product between two matrices, where one matrix is the set of learnable parameters which

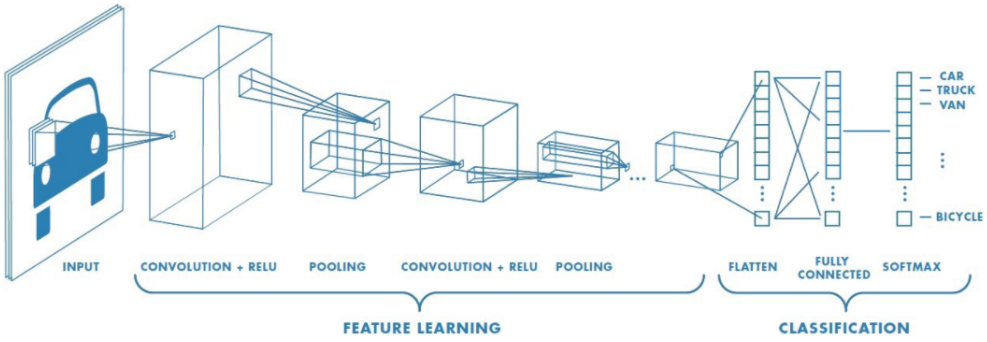
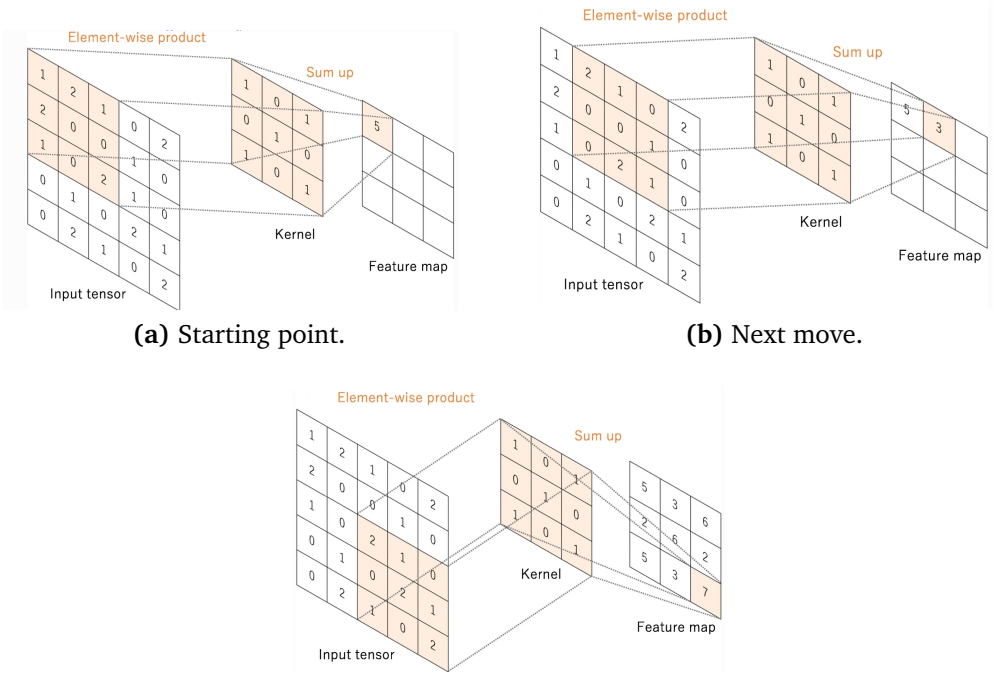


Figure 2.9: The architecture of CNN predicts the vehicle represented in the input image. In the first part, the convolutional layers learn a good feature representation for separating the classes. In the second part, fully connected layers take over the classification and produce the output which is the predicted target. Source: [47]

is known as a kernel, and the other matrix is the restricted portion of the receptive field i.e. a small region of the input to that layer [48]. The kernel is much smaller than the input data but can have high depth, meaning that if the image contains multiple channels such as an RGB image have 3 channels. The width and length of the kernel will be small in spatial but the depth will extend up to all three channels. If the input image is a spectrogram of an audio signal, the channel is one as it is a grayscale image therefore in this case, the kernel will have a small depth.
During the first operation, the kernel starts at the top left corner and moves to the right with a specified number, defined by *Stride Value*, until it parses the complete width. The kernel moves down to the beginning of the image with the same stride value and repeats the same process until the whole image is traversed. Figure 2.10 shows how this process is done where the stride value is one.

As the convolutional layer, the pooling layer reduces the spatial size of the feature map. This is done so the required amount of computation and weights are reduced. The pooling operation is processed on every given size of the feature map individually and is used to extract dominant features which are where the dominant frequencies are in the case of when having a spectrogram as an input image. This leads to maintaining the process of effectively training the model.

There are two different types of pooling operations: Average pooling and Max Pooling. Average Pooling returns the average of all the values from the portion of the image covered by the pooling size which results in a noise suppress-



(a) Starting point.

(b) Next move.

(c) Ending point.

Figure 2.10: Movement of the kernel when the stride value is one. Source: [49]

ing mechanism. On the other hand, Max Pooling returns the maximum value from the portion. This operation behaves as a Noise Suppressant as it discardes the noisy activations althogether and also performs de-noise along with dimensionality reduction. Therefore, the Max pooling are doing much better than the Average Pooling. Figure 2.11 shows how the operations works.

The convolutional layer together with the pooling layer represents one layer in a CNN as shown in the architecture from Figure 2.9. How many of these layers a CNN contains is depending on the complexities in the images. However, this is at the cost of more computational power.

Now, the model understands the features. Further, the final output will be flattened out and fed to an ANN which is described earlier. Over a series of epochs, the model can distinguish between the dominant and the low-level features in images.

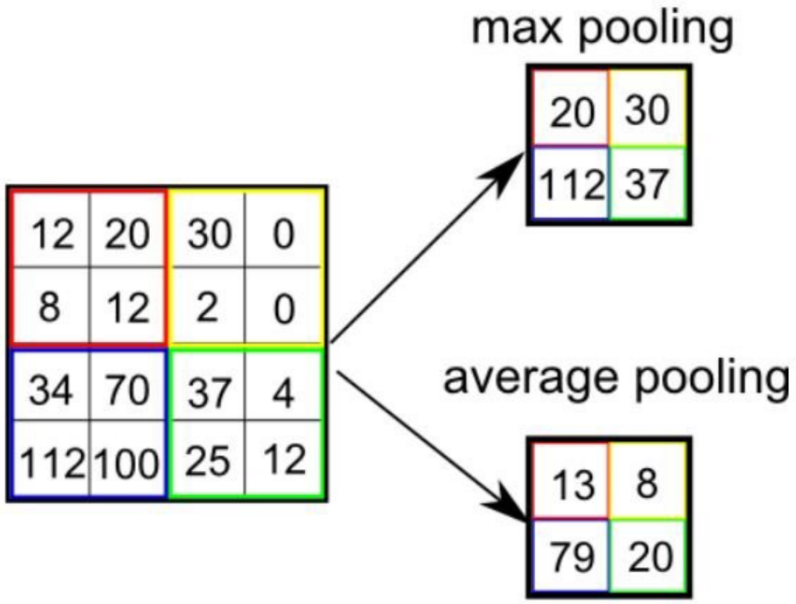


Figure 2.11: Max pooling and average pooling. Source: [50]

2.5.4 Overfitting

The following paragraph is partly reused from the specialization project [40]:

In supervised learning, the algorithm can get overfitted, meaning that a model learns the random fluctuations and noise in the training data to the extent that it negatively impacts the performance of the model on unseen data [51]. Overfitting is more likely to occur when having small data set as the model is more likely to see patterns that do not exist [52]. One way to observe overfitting is to look at the loss on a validation set during training. If the loss of the validation set starts to increase while the loss on the training set continues to decrease, it is a sign that the models start to overfit.

In ANNs, a common technique, called *early stopping* is used to prevent overfitting. The idea is to stop training when the validation loss does not decrease after a fixed epoch. [53]. Also, another way to avoid overfitting is to use the dropout layer which randomly drops some of the connections between layers during training. If this is done, the different networks will result in overfitting in different ways which in the total effect of dropout will reduce overfitting [54].

2.6 Software

2.6.1 Python

In this master's thesis, the machine learning algorithms and acoustic analysis are implemented in Python. Python is a well-known programming language that has plenty of libraries useful for data analysis and Machine Learning [55]. These are also well documented which makes them easy to use. These include libraries for statistical data analysis such as NumPy to Machine Learning in Keras. It is also great for visualizing massive data sets. This makes Python a well-suited programming language for this study.

2.6.2 SciPy

SciPy is an open-source library in Python which is used to solve mathematical and scientific tasks [56]. This library is built on several extensions such as NumPy, Pandas and Matplotlib, which contain high-level commands that can manipulate and visualize data.

2.6.3 Scikit-learn

Scikit-learn is an open-source library that is used for Machine Learning. It is built on NumPy, Pandas and Matplotlib and contains high-level functions for data modeling, Machine Learning algorithms and evaluation [57].

2.6.4 Keras

Keras is an open-source library in Python that simplifies the creation and evaluation of deep learning models. It has extensive documentation and developer guides. It also provides clear and actionable error messages [58].

2.6.5 Librosa

The Librosa package is a Python library that provides the necessary building blocks for audio analysis and retrieval systems [59].

2.7 Audacity

Audacity is a free open-source cross-platform audio software. The program supports 16-bit, 24-bit and 32-bit signals. Editing such as cutting, copying, pasting and deleting is easy to do. Viewing a spectrogram to visualize the frequencies of an audio signal is also possible to do[60].

2.8 Performance Metrics

Various parameters/metrics can be used to judge the performance of a machine learning model. When having a binary classification problem, the outcome can only be one out of two classes. When predicting the labels, four different situations can occur:

- True Positive (TP)- number of samples of the positive class that has been predicted correctly.
- False Negative (FN)- number of samples of the positive class has been predicted incorreclty.
- False Positive (FP)- number of samples of the negative class has been predicted incorrectly.
- True Negative (TN)- number of samples of the negative class has been predicted correctly.

All these situations makes up the confusion matrix shown in Table 2.

	Predicted Positive Class	**Predicted Negative Class**
Actual Positive Class	TP	FP
Actual Negative Class	FN	TN

Table 2: Confusion matrix when having a binary classification.

Different metrics can tell how good or bad the model performs and can be derived from these factors. The most relevant for the present study is Accuracy, Precision, Recall, F1-score, AUC and ROC. All of the metrics are described under and can have a value between 0 and 1, where the latter and former is the best case and worst case, respectively.
Accuracy is a measure that tells the number of correct predictions divided by the number of evaluated instances and is defined as

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{16}$$

Precision is the ratio of true positives and total positives predicted so

$$Precision = \frac{TP}{TP + FP} \tag{17}$$

Recall/Sensitivity or True Positive Rates (TPR) tells the ability of the classifier to find positive samples and is defined as

$$Recall = \frac{TP}{TP + FN} \tag{18}$$

Specificity or True Negative Rate(TNR), tells the ability of the classifier to find negative samples and is defined as

$$Specificity = \frac{TN}{FP + TN} \tag{19}$$

F1-score is the weighted harmonic mean of the precision and recall and is defined as

$$F1 - score = \frac{2 \cdot precision \cdot recall}{precision + recall} \tag{20}$$

Another useful metric when evaluating a binary classifier is the Receiver Operating Characteristic (ROC) curve. It is a graphical representation of a binary classifier's diagnostic capabilities.

It plots the true positive rate and false positive rate of a given y-axis relative to the x-axis. The curve should bow to the top left corner to have a skilled classifier. A no-skill classifier will have a diagonal line from the bottom left to the top right. The area under the ROC-curve (AUC) gives a measure of how good the classifier is. The score is between 0.5 and 1, where a no-skill model will have an AUC of 0.5 and a perfect skilled model with an AUC of 1.

Depending on if the test set is balanced or imbalanced, different metrics will be useful for each case. For the former, there are equal samples of the classes. Therefore, accuracy will be a good metric as the data is balanced and the metric will tell how much of the test input got classified correctly. In the latter case, there is an imbalance between the classes. In this case, the accuracy will in general, not be a good metric as the high accuracy will be obtained by only predicting the majority class. Therefore, Recall can be an important metric, as it measures how well the model predicts the positive class, which is often the minority class. The results in this study will be presented by several of the metrics above.

2.9 Box plot

A box plot is a type of chart commonly used in data analysis. It shows the distribution of numerical data and its skewness by displaying the data quartiles and averages. In addition, box plots also provide a representation of the overall trend. Figure 2.12 shows different parts of a box plot when the data values are normally distributed. The minimum score, excluding the outliers, is shown at the left whisker. In the first Quartile, Q1, 25% of the scores fall below this quartile. The median line marks the mid-point of the data. In the third quartile, Q3, 75% of the score fall below this quartile. Thus, 25% of data is above this value [61].

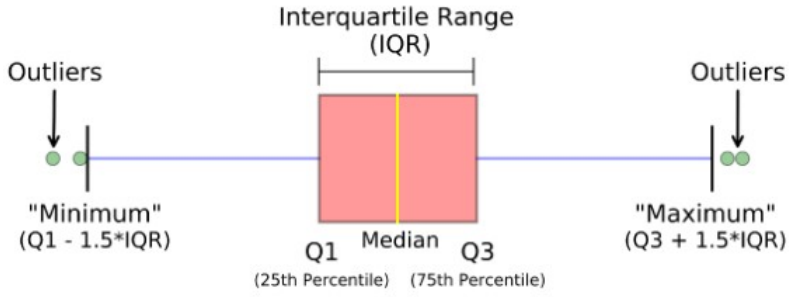


Figure 2.12: Different part of a box plot. Source: [62]

3 Data preparation - Youtube recording and data collection

This section describes the different type of data which has been used for designing the bowel sound detector and the analysis. The data sets are from the Youtube channel, *stomach and intestines sound* [63] and APT. The former is openly available, while the latter is collected by the research group, APT, by following an expert-approved protocol.

3.1 Data set - Stomach and intestines sound

Stomach and intestines sound is a youtube channel where the owner posts different recordings containing sounds from the abdomen after eating a meal. The meal can be anything such as soybean, oatmeal, or banana. Three different recordings were selected randomly to have different types of bowel sounds as these may be dependent on the meal. These recordings are named *Belly sounds 20211010*, *Belly sounds 20211022* and *Belly sounds Mix water-soluble dietary fiber and carbonated water* and are described by the owner as in Table 3. The recordings were analysed using Audacity [60]. It was easy to determine whether it was a bowel sound or not in a given period as the non-bowel sound was often silence or background noise.

Recording title	Description
Belly sounds 20211010	The sound of digesting soybeans and oatmeal.
Belly sounds 20211022	Digested food moves in my intestines.
Mix water-soluble dietary fiber and carbonated water	A few hours after ingesting psyllium husk and water, I drank carbonated water and recorded.

Table 3: Table showing the description of the selected recordings from Youtube [63].

3.2 Data set collected by APT

The experiment took place at NTNU Gloshaugen campus and Øya campus. The audio data were collected in a non-clinical setting during the Fall of 2021 and spring of 2022. Three different protocols have been followed with different aims; the first one is for early meal detection, the second one is for differentiating between soft and hard meals, and the last one is for meal simulation. All these protocols have been approved by experts [64].

3.2.1 Data Equipment

The diagram of how the recordings were collected is shown in Figure 3.1. Figure 3.2 shows a picture of the complete acquisition system in real. The sound made by the body is recorded by four SPM0687LR5H-1 microphones. These are connected to a power source and the signal was then sent to a sound card (Roland Octa-Capture 24-bits 48 kHz) [65]. The output of the sound card is connected to a computer, where the recordings are saved.

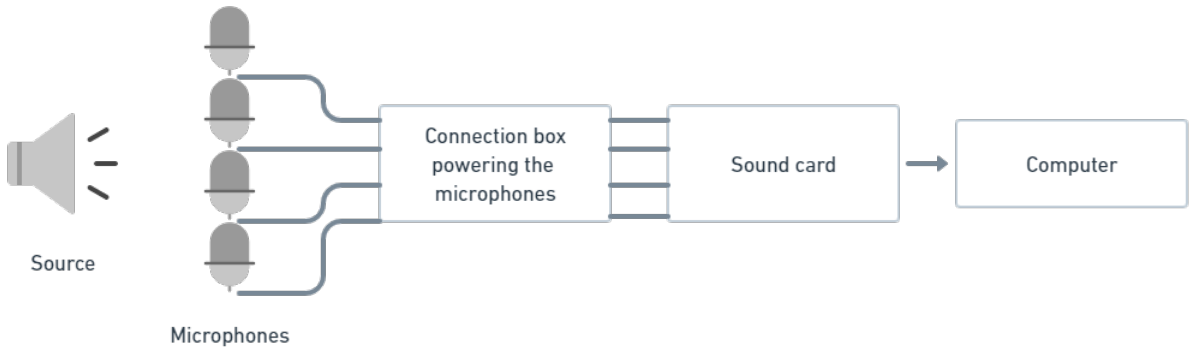


Figure 3.1: Diagram of how the recordings were collected.

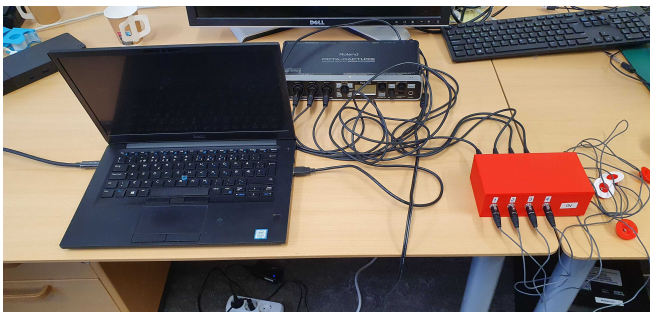


Figure 3.2: The complete acquisition system. The red box is the connection box for powering the microphones. The black box is the sound card for digitizing the sound signal and a laptop for saving the data.

The microphones were placed in a stethoscope-shaped microphone holder as shown in Figure 3.3. To hold the microphones on the skin during the recording they were attached with a ring-shaped double-sided tape.

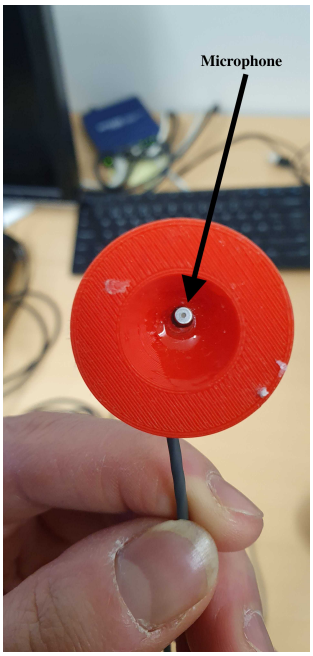


Figure 3.3: Microphone set up.

3.2.2 Protocol for the meal recordings

For each recording session, four microphones were recorded simultaneously which results in 4 different recordings. The different positions of the microphones are shown in Figure 3.4. The first microphone was used to record chewing sounds and therefore placed right under the right ear, while the second

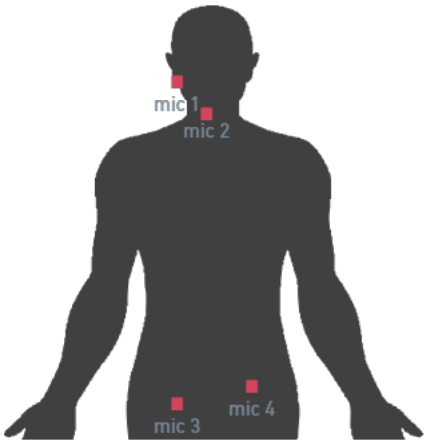


Figure 3.4: The placement of the microphones.

microphone was used to record swallowing sounds and placed over the collar bone on the neck. Both of these recordings were discarded as it was not used in this study. The third and fourth microphones were placed on the RLQ and LLQ, respectively, to capture bowel sounds.

The protocol of the experiment is shown in Figure 3.5. The subjects were asked to fast at least 3 hours before the start of the recording session. The recording session started with the subject continuing fasting for 15 minutes, followed by eating a meal of their choice for 10-15 minutes, and ending with at least 45 minutes for digestion. The subject remained seated for the whole recording.

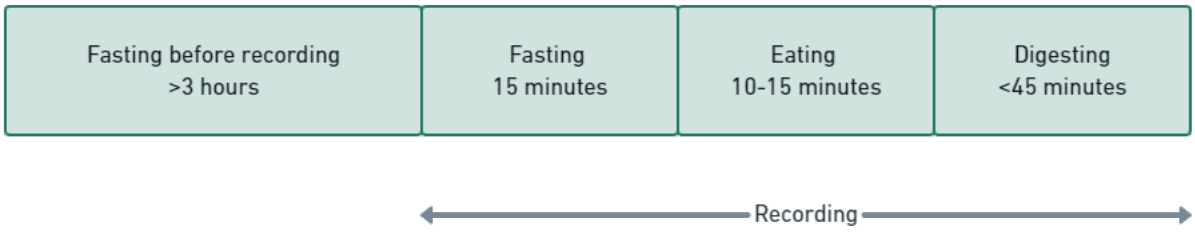


Figure 3.5: Protocol of meal recordings.

3.2.3 Protocol for the meal type recordings

The placement of the microphones is the same as for the meal protocol as shown in Figure 3.4. The recordings that contained swallowing and chewing sounds were discarded. The protocol is shown in Figure 3.6 and the subjects were asked as earlier, to fast at least three hours before the start of the recording session. The recording started with the subject continuing fasting for 20 minutes, followed by eating a soft or hard meal for less than 10 minutes, and ending with 5 of digestion. The soft and hard meals contained oatmeal and salad, respectively. The subject remained seated for the whole recording.

Figure 3.6: Protocol for meal type.

3.2.4 Protocol for the meal simulation recordings

The placement of the microphones was a bit different in the protocol for the meal simulation. The Figure 3.7 shows the different positions of the microphones. For each recording session, microphone 1 and microphone 2 in LLQ and RLQ, respectively, were recorded simultaneously which results in 2 different recordings.

The subjects were asked, as earlier, to fast at least three hours before the start of the recording session. The session started with the subject continuing fasting for 15 minutes, followed by watching a food video of their own choice in 15 minutes, followed by 15 minutes of being idle, and ending with eating a meal in less than 15 minutes. As earlier, the subject remained seated for the whole recording.

Figure 3.7: The placement of the microphones when following meal simulation protocol.

Figure 3.8: Protocol for meal simulation.

4 Design of BS detector using data from *Stomach and intestines sound*

A bowel sound detector has been implemented using data from Youtube described in Section 3.1. The system implementations are described, including preprocessing of audio data, feature extraction, classifier setup, and choice of training parameters such as batch size, epochs, kernel size, optimizer, neurons in dense layer and dropout rate.
Figure 4.1 shows the workflow of this thesis. The two first sections contains of how the different BS detectors have been implemented and a short discussion of the results. Also, the detectors have been tested on the collected data described in Section 3.2 to see how it performs in a 'real condition' during fasting, meal and digestion. The aim is to design a BS detector that can be used for analysing recordings when different protocols are followed. The next three sections use the final implemented BS detector to see the ability of early meal detection, distinguish between hard and soft meal, and how it performs on meal simulation.

Due to the need to process large amounts of data, Google Colab Pro was used [66]. The RAM size of the computer used was of 8 GB while the Google Colab can have a RAM size of 24 GB. Also, the software had powerful GPUs which is an advantage when training deep neural networks.

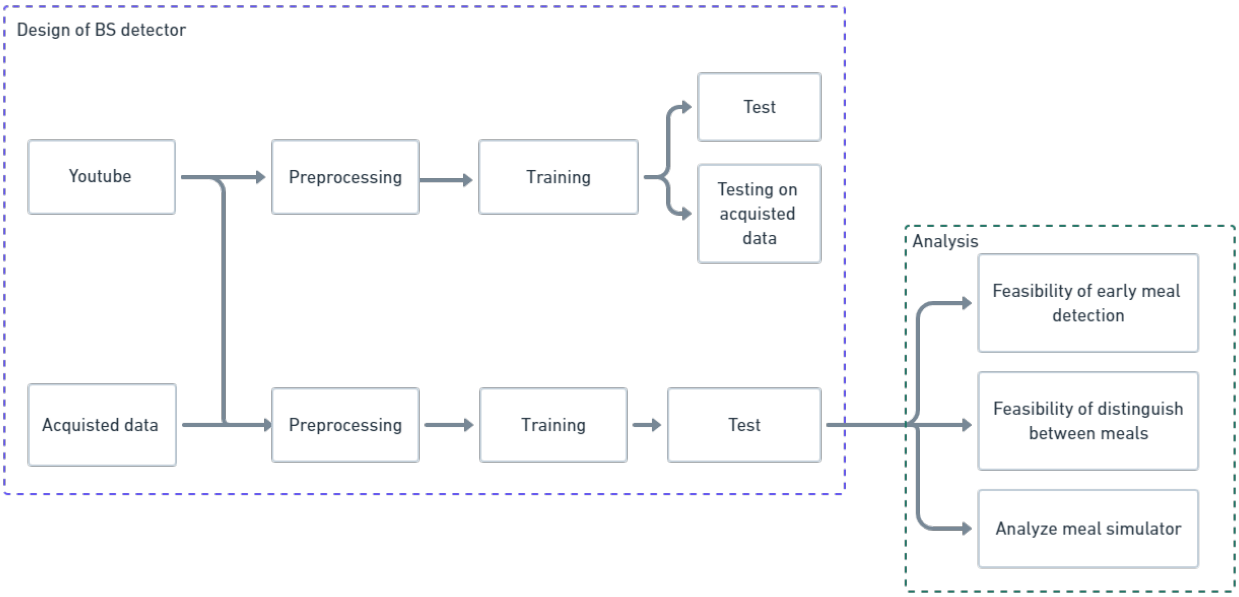


Figure 4.1: Workflow of the BS detector design and analysis.

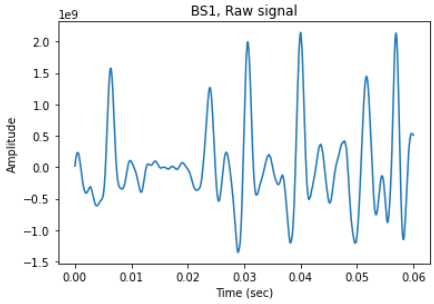
4.1 Preprocessing

The data is from the youtube channel stomach and intestines sound described in Section 3.1. The generated data set is balanced with 248 labeled segments as BS or NBS (non-bowel sound) by using Audacity [60]. The segments lasted at least 60 ms and were randomly selected in the recording. This means that a BS segment could be at the start, middle, or end of a long bowel sound. As mentioned earlier, SB lasts typically 10-30 ms, there was not taken any accounts for where these could occur in the labeled segment. It was noticed that NBS was often a silence/specific background noise. The number of labeled BS and NBS from the different youtube recordings is shown in Table 4. TThe bowel sounds and non-bowel sound segments got named *BS1, BS2,..., BS124 and NBS1, NB2,..., NBS124*, respectively. These were saved as new recordings and resulted in 248 balanced labeled recordings containing either a bowel sound or not.

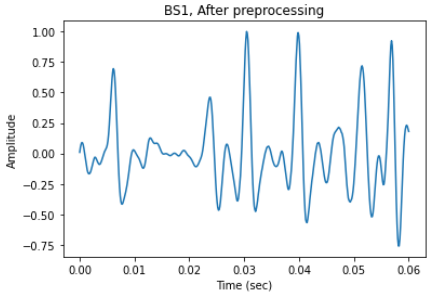
Recording title	**Total number of BS and NBS**
Belly sounds 20211010	86
Belly sounds 20211022	82
Mix water-soluble dietary fiber and carbonated water	80
Total	248

Table 4: Table showing total number for BS and NBS from each recording.

The labeled recordings were sliced so it only contained 60 ms of the beginning to ensure the uniform size of the input data for the classifier. These were downsampled from 48 kHz to 8 kHz using the *decimate* function from [67]. The downsampling reduces the sample rate of the signal so the computation accelerates without loss of relevant information. As presented in Section 2.2, bowel sounds usually have a frequency of at least 200 Hz and vary up to 3000 Hz. The signal was therefore filtered by a second-order Butterworth high-pass filter using the *butter* function from [67]. The cutoff frequency was set to 80 Hz to remove non relevant low frequencies. Figure 4.3 shows the frequency response of the filter. As the different audio signals had different amplitude values, the signal was normalized to the range [-1,1]. This is to ensure that the relative changes within the signal were considered as explained in Section 2.3.4. The raw and the preprocessed signal of the first labeled bowel sound, BS1, is shown in Figure 4.2a and 4.2b, respectively.



(a) Raw signal of the first labeled bowel sound.



(b) Preprocessed signal of the labeled bowel sound.

Figure 4.2: The raw and preprocessed signal of BS1 to the left and right, respectively.

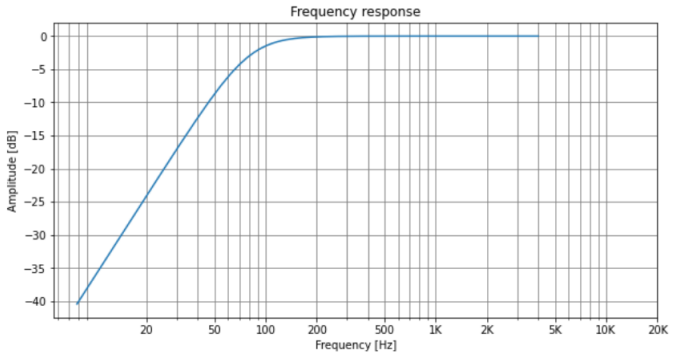
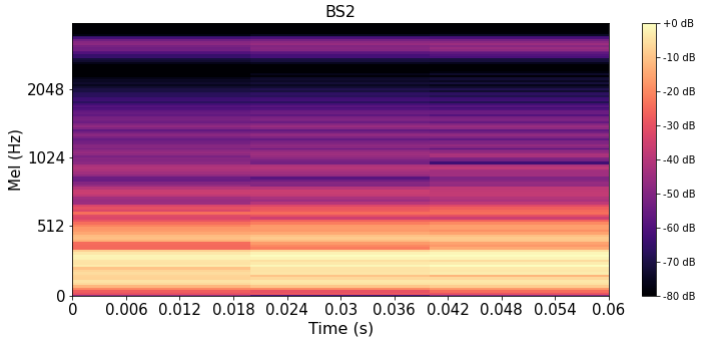


Figure 4.3: Frequency response of the second-order Butterworth high-pass filter.

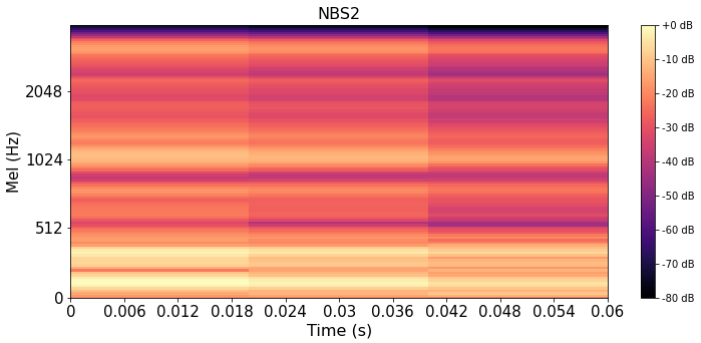
4.2 Feature Extraction

For each of the preprocessed labeled recordings, the Mel-scaled spectrogram got calculated using the function *melspectrogram* from librosa package [33]. These features were chosen because they are commonly used in audio classification [32] and are used in the study [12] which made it possible to distinguish between a bowel sound and a non-bowel sound. The frame length and hop length were set to 50 ms and 5 ms, respectively and the center to False. Hence, for each segmented signal, the shape of the input tensor is 128x3. Figure 4.4a and 4.4b shows an example of how the features looks like for BS2 and NBS2, respectively. Typically, the Mel-scaled spectrogram for a bowel sound has a higher amplitude value in lower frequencies than in higher frequencies. Therefore, one can expect brighter color on the bottom of the spectrogram and a darker color on the top. The non-bowel sounds are typical of noise, the spectrogram of these should be flat, especially if the noise is even. It is shown from the figures, the spectrogram of the bowel sound has a black stripe on top while the non-bowel sound has orange color evenly distributed. The obtained feature matrix has a dimension of 248x128x3 where one row corresponds to one labeled recording.

Class labels were originally "BS" and "NBS" and were encoded to 1 and 0, respectively, to be able to feed them to the classifier.



(a) Mel-scaled spectrogram of BS2.



(b) Mel-scaled spectrogram of NBS2.

Figure 4.4: Representation of the features for a bowel sound and non-bowel sound at the top and bottom, respectively.

4.3 Training

The data set was split into 70% training data, 10% validation data, and 20% test data. The ML classifier chosen for this feature set is the CNN described in Section 2.5.3. The classifier is good at finding patterns within images which is in our case, the spectrograms. It also gave great results in the research [12] which developed a bowel sound detector. The network setup was therefore inspired by this.
A CNN network was created using the Keras library [58] and is consisting of convolution layers, pooling layers and fully connected layers as described in Section 2.5.3. The architecture of the model designed is shown in Figure 4.5. The input dimension was set to (128, 3), as one labeled recording had a feature dimension of 128x3. Three convolutional layers with the rectified linear units and a maxpooling layer were used. To prevent overfitting, a drop out layer was used. The output was then flattened and sent to a dense layer. Lastly, a fully connected sigmoid layer was used to get the prediction in probability to either be BS or NBS, for the input sample. The binary-cross entropy was used as a loss function as the case was of binary classification. To get optimized results, parameter tuning was done manually by trying different parameter combinations shown in Table 5.
To avoid overfitting and unnecessary long training time, the early stopping technique was used. If the validation loss did not decrease over a time period of 5 epochs, the training was stopped.

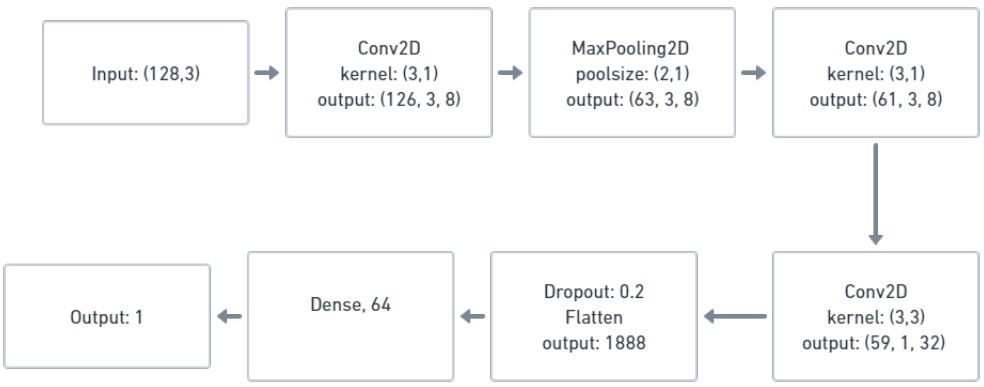


Figure 4.5: Architecture of the CNN network.

Batch_size	Optimizer	Neurons in dense layer	Dropout rate
[32, 64, 128]	[Adam, Adadelta]	[32, 64]	[0, 0.2, 0.5]

Table 5: Various parameters and their possible values that has been tried out to get optimized results.

4.4 Results and testing

The CNN model was evaluated in two different ways: testing on the test set and testing the model on the collected data described in Section 3.2.2, which are presented in Section 4.4.1 and 4.4.2, respectively.

4.4.1 Test set

The best parameter combination were tuned manually and found to be: batch_size=128, optimizer= Adam and dropout= 0.2. A CNN network was trained on these parameters and evaluated on the test set. The confusion matrix and confusion report is shown in the upper part and lower part of Figure 4.7, respectively. The classification report shows precision, recall, and F1-score obtained for each class. This means for the BS row, BS is regarded as the positive outcome, while in the NBS row, NBS is regarded as a positive outcome. Obtained metrics for both of the classes were 100% for all of the metrics. The training history is shown in Figure 4.6 and shows the classifier has no difficulties to find a pattern for each class. The model was saved after 12 epochs as the training loss and validation are at 0. This was done by using the function *dump* from [68]. The ROC-curve in Figure 4.8 of the CNN network is in orange and bows towards the top left corner which means the classifier is skilled and has beaten the no-skill model in blue. The CNN network obtained an AUC of 100% meaning that the classifier is perfect.

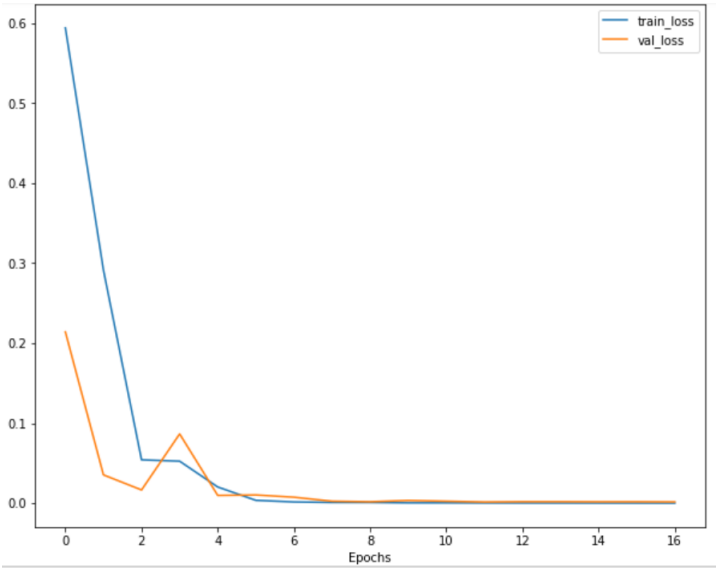


Figure 4.6: Training history for the CNN network showing the validation and training loss. The model was trained for 12 epochs.

	precision	recall	f1-score	support
NBS	1.00	1.00	1.00	29
BS	1.00	1.00	1.00	29
accuracy			1.00	58
macro avg	1.00	1.00	1.00	58
weighted avg	1.00	1.00	1.00	58

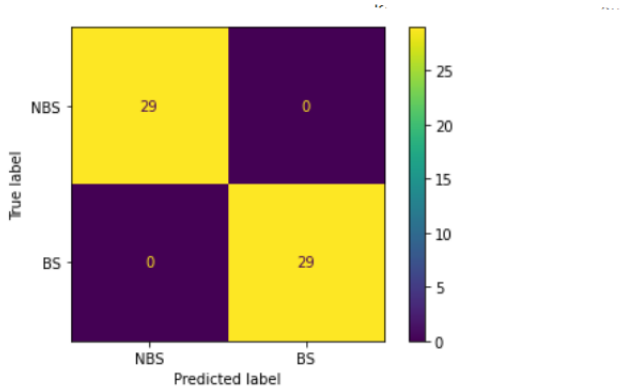


Figure 4.7: Confusion report and confusion matrix.

4.4.2 Testing on collected data

The results were great when testing on the test set, it can be interesting to see how the BS sound detector performs on more 'real condition' data. Two subjects followed the protocol described in Section 3.2 where the first subject did

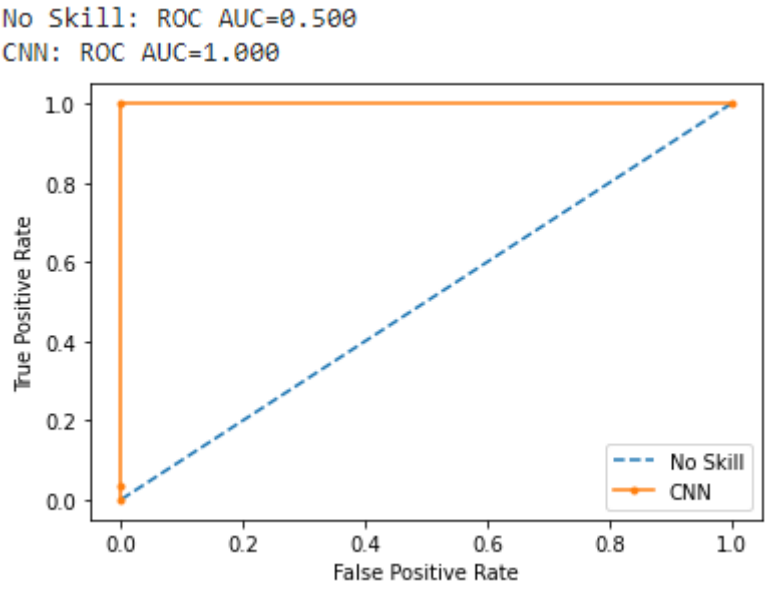


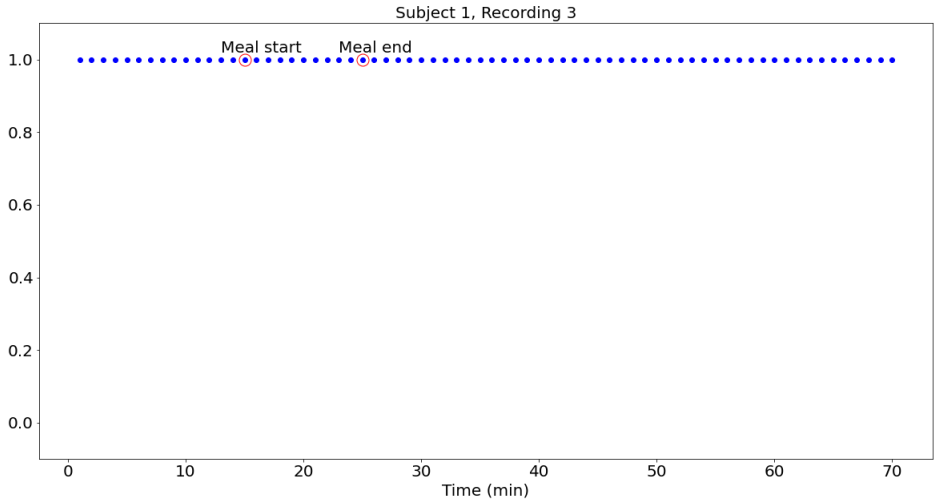
Figure 4.8: ROC-curve for the CNN network. The curves show the true and false positive rates for different thresholds between 0.0 and 1.0. AUC is shown as the area.

the experiment twice. This produces in total 6 recordings as two microphones were placed in two different positions, RLQ and LLQ. The implemented BS detector would be used to identify bowel sounds on the collected recordings. The recordings had a bit depth of 24 bits and since Scipy package [67] can not read this data type, the recordings got quantized to 16 bits using Audacity [60]. To have more continuity between different time segments, an overlap was carried out. The recordings got segmented in each minute by using the function *frame* from librosa package [59]. The window length and the hop length were set to 60 ms and 50 ms, respectively. This means for each minute, the number of frames is 1199. The same preprocessing was done as earlier: each frame got downsampled, filtered and normalized. The Mel-scaled spectrogram was calculated for each frame and fed into the model in time order, which produces a time series of predictions (BS or NBS). The table 6 shows the proportion of detected BS and NBS frames for each recording, SNRM stands for Subject N, Recording M. Subject 1 and subject 2 are the same individual. The microphone 3 and 4 correspond to the places LLQ and RLQ, respectively as described in Section 3.2.2. In all of the recordings, RLQ was the position the model detected most BS frames, with a probability of at least 80%. However, in position LLQ, for all of the recordings, the detection of BS frames was much less, a maximum of 30%. The recordings were also partially listened to and it has been noticed that there is a lot more noise and sound on the recordings collected from the microphone 4. From earlier research [12], it has been shown that long bowel

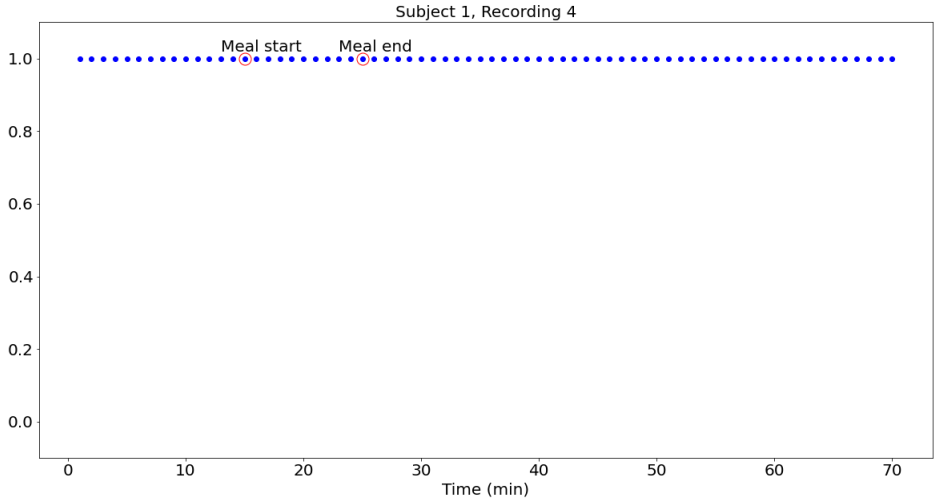
Subject and microphone position	BS	NBS
S1R3	22.6%	77.45%
S1R4	92.8%	7.2%
S2R3	27.5%	72.5%
S2R4	93.7%	6.3%
S3R3	14.1%	85.9%
S3R4	86.4%	13.6%

Table 6: Proportion of BS and NBS detected by the model for each recording.

sound has a higher potential to distinguish between fasting and food intake. Therefore, three consecutive BS frames corresponding to a time period of 160 ms were set to identify one bowel sound. Figure 4.9a and 4.9b shows how the detection of bowel sound is during fasting, meal intake and digestion for S1R3 (subject 1, recording 3) and S1R4 (subject 1, recording 4), respectively. The x-axis shows the time in minutes while the y-axis shows if the current minute contains a bowel sound (1.0) or not (0.0). The red circle at 15 minutes and 25 minutes marks the meal start and meal end, respectively. The detector recognizes at least one bowel sound each minute. This was done to the rest of the recordings and the result are the same for all of them.



(a) Plot of detected BS in each minute of the subject 1, recording 3.



(b) Plot of detected BS in each minute of the subject 1, recording 4.

Figure 4.9: Plot of how the model performs during the fasting, eating and digestion period. If the model predicts three consecutive frames as BS, then it indicates that the whole minute contains BS.

Due to time constraints, all the segments in the recording were not labeled as BS and NBS before feeding to the model, therefore another evaluation method had to be done. The collected recordings were partly listened to and every time a sure bowel sound and the non-bowel sound were found, the timepoint was saved and converted to frame number using the equation 21. It was noticed

that the model predicts most movements and noise as bowel sounds.

The formuala for converting time to frame number is given as

$$FrameNumber = N \cdot 1199 + \frac{t - 0.06}{0.005} + 1 \tag{21}$$

where N is the number of minutes and t is the time in seconds. The function, called *sec_to_frame* shown in Listing 1 was created to do the calculation effective. The function has the input time in seconds and returns the corresponding frame number.

```
def sec_to_frame(time):
  if time>=60:
    minu=int(time//60)
    sec=time-minu*60
    if sec==0:
      frames=minu*1199
    else:
      frames=minu*1199+(sec-0.06)/0.05+1
  return frames
```

Listing 1: Python code that converts time in seconds to corresponding frame number.

4.5 Discussion

The results obtained when testing the model on the test set were promising. The weighted average of the precision, recall and F1-score for the two classes were 100%, which is a perfect model. The ROC-curve shows also that the model performs ideally on the test set. The classifier did not find any difficulties to distinguish between the BS and NBS frames. However, when testing on the collected data that presents a more 'real condition, the model performs badly. Table 6 shows a high proportion of BS frames was found, especially from the recordings collected at LLQ. The highest occurrence of bowel sound frames was achieved by subject 1, microphone 4, with a proportion of 94% which is unrealistic high. According to [69], normally, the bowel sounds should occur 5-30 times per minute. The recordings were partly listened to and it was noticed that all sounds in the recording, whether it was a bowel sound or not, were predicted as a bowel sound. Also, the recordings collected from LLQ contained more sound which often could be reminiscent of a movement, so sometimes it was harder to tell for sure if a time period was of bowel sound or not. The sounds may be CRS, a type of bowel sound reminiscent of noise. The plots of detection of bowel sounds in each minute do not show the distribution of bowel sounds registered through session time, which leads to difficulties in saying whether there is less number of BS during the fasting state as found in [13]. The analysis should be more investigated even though the detections are

generally too high. A clear limitation of the data used is the non-bowel sound in the Youtube recordings is often a silent period or one specific background noise. The generalization of non-bowel sounds is limited as it has not seen anything but a silent period as NBS. Also, the data is only collected from one person with an unknown microphone and environment than the collected recordings from APT.

5 Design of BS Detector including collected data

From the previous section, it is shown that the model performs badly on the collected recordings as it contained a different type of noise. Therefore, a new detector was designed which is trained on the noise contained in the experimental recordings. The goal and motivation is to make the classifier more robust and improve the results when testing on the acquired recordings.

5.1 Preprocessing

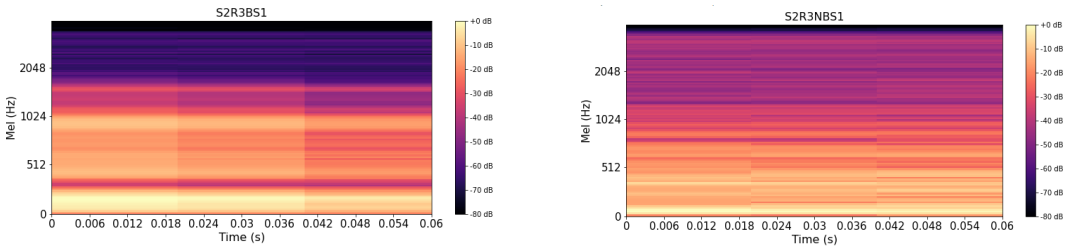
7 different subjects followed the protocol described in Section 4.4.2, two of the subject did the protocol twice. This produces in total 9*2=18 recordings, named as SNRM where N and M are the number of subject and microphone, respectively. For the record, the subjects that did the experiment twice got named as different subject numbers meaning $N = 1, 2, .., 9$. Subject 1 and 2, and subject 3 and 9 are the same person, respectively. Recordings from subject 4, 5 and 8 failed and got discarded, resulting in 12 recordings being taken care of. Random segments lasting at least 60 ms and containing BS or NBS were cut from the recordings collected by three different subjects to avoid making a subject-depending model. This was done by a non-expert listening to different random parts of the recordings using Audacity [60]. The segments could be at the start, in the middle, or at the end of a long BS. For short bowel sounds such as a single burst, no account was taken of where it could occur in the segment. As the previous detector worked poorly on the recordings from microphone 4, more segments were labeled from this position than from microphone 3. Table 7 shows the total number of BS and NBS segments from the different subjects. The segments were sliced to only contain the first 60 ms to ensure the uniform size of the input data for the classifier. In total, 800 balanced segments got labeled as BS and NBS. The NBS were carefully not chosen as a silent period as this is already included in the recordings from youtube. The segments were saved as new recordings with the name SNRMBSO (Subject N, Recording M, BSO) and SNRMNBSO (Subject N, Recording M, NBSO) for BS and NBS, respectively. N is the number of subjects (N=2, 6, 7), M is the position of the microphone (3 or 4) and $O = 1, 2, ...$ is the number of the labeled segment. The previous data set was added which results in a total of 1048 recordings. The recordings got quantized to 16 bits using Audacity. The rest of the preprocessing followed the same procedure as in Section 4.1: Each recording got downsampled to 8 kHz, filtered by a second-order Butterworth high-pass filter and normalized to the range [-1,1].

Subject and microphone position	Number of BS and NBS
S2R3	100
S2R4	300
S6R3	100
S6R4	100
S7R3	100
S7R4	100
Total	800

Table 7: Total number of BS and NBS from different recorings.

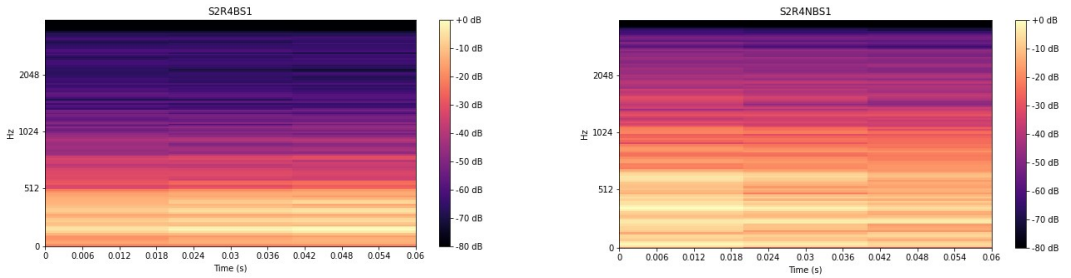
5.2 Feature Extraction

For the feature calculation, the same procedure was followed as in Section 4.2: each generated recording containing either a BS or NBS, the Mel-scaled spectrogram got calculated with a frame length and hop length of 50 ms and 5 ms, respectively. The resulting feature set has a dimension of 1048x128x3. Figure 5.1a and 5.1b shows how the features looks like for S2R3BS1 and S2R3NBS1, respectively. Figure 5.2a and 5.2b shows how the features looks like for S2R4BS1 and S2R4NBS1, respectively. It is easy to see the difference between the features on a bowel sound and a non-bowel sound, also, these features are reminiscent of the features from Figure 5.2; For a BS, they both have high amplitude in the lower part of the spectrogram than in the higher part of the image. For a NBS, they both have merged different colors everywhere. The class labels "BS" and "NBS", got as earlier, encoded to 1 and 0, respectively.



(a) Mel-scaled spectrogram of S2R3BS1 **(b)** Mel-scaled spectrogram of S2R3NBS1

Figure 5.1: An example of how the features looks like for a bowel sound and non-bowel sound to the left and right, respectively, from recording 3 and subject 2



(a) Mel-scaled spectrogram of S2R4BS1 **(b)** Mel-scaled spectrogram of S2R4NBS1

Figure 5.2: An example of how the features looks like for a bowel sound and non-bowel sound to the left and right, respectively, from recording 4 and subject 2

5.3 Training

The data set was split into 70% training, 15% validation and 15% test. The machine learning classifier chosen for this feature set is the same as the previous section for comparison purposes and as it gave good results. A CNN was created using the Keras library [58]. The architecture of the model is shown in Figure 4.5. The input dimension was set to (128, 3), as one segment recording had a feature dimension of 128x3. Parameter tuning was done by trying different parameter combinations shown in Table 5 to see if the performance could be improved.

5.4 Results

5.4.1 Test set

The best parameter combination were found to be: batch_size=128, optimizer=Adam and a dropout_rate=0.2. A CNN network was trained on these parameters and evaluated on the test set. The confusion matrix and confusion report is shown in the upper part and lower part, respectively, in Figure 5.3. For the BS class, the obtained metrics were 85% for precision, 88% for recall and 87% for f1-score. For NBS class, the obtained metrics were 88% for precision, 85% for recall and 86% for f1-score. The ROC-curve in Figure 5.5 of the CNN network bows towards the left upper corner which is good. The CNN network obtained an AUC of 95%. From figure 5.4, it is shown that the validation loss starts to flatten out after epoch 25 and is at its minimum at 43 epochs. The model was therefore saved with 43 epochs to avoid overfitting.

	precision	recall	f1-score	support
NBS	0.88	0.85	0.86	79
BS	0.85	0.88	0.87	78
accuracy			0.87	157
macro avg	0.87	0.87	0.87	157
weighted avg	0.87	0.87	0.87	157

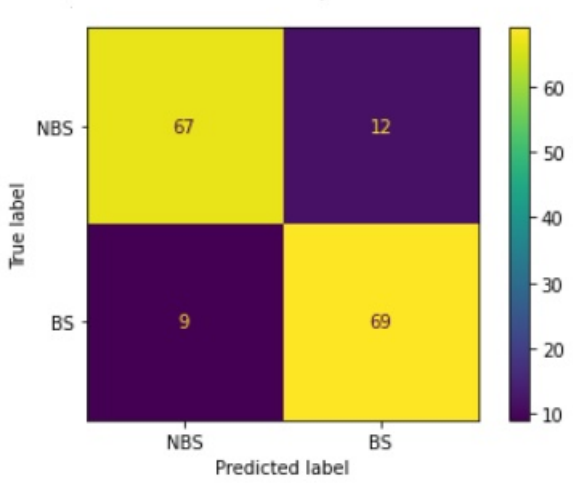


Figure 5.3: Confusion report and confusion matrix.

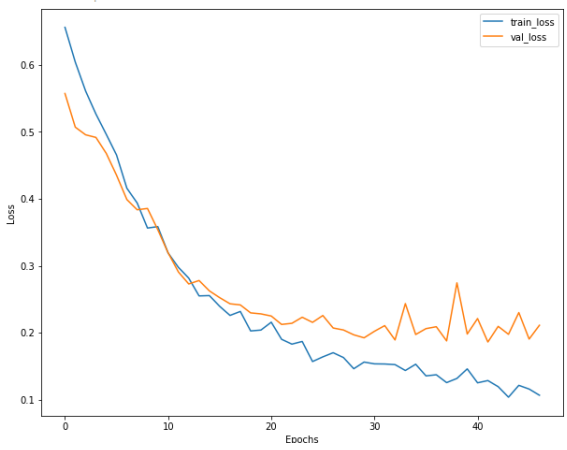


Figure 5.4: Training history for the CNN network showing the validation loss in orange and training loss in blue.

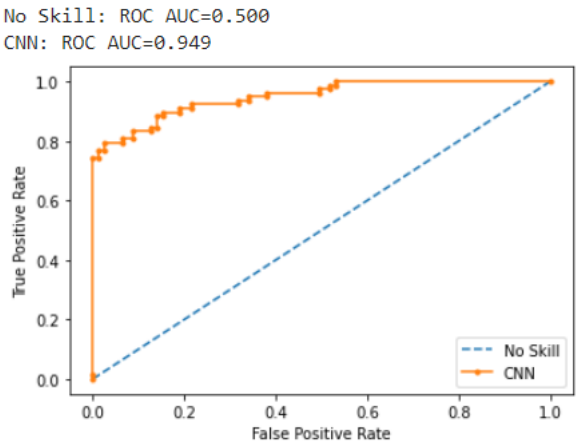


Figure 5.5: ROC-curve for the CNN network. The curves show the true and false positive rates for different thresholds between 0.0 and 1.0. The blue and orange graph represent a ROC-curve of a no-skilled model and the created CNN model, repsectively.

5.4.2 Testing on collected data

The bowel sound detector would be used to identify the bowel sounds on all the collected data. Some segments are in the training set already, however, this is a small part of the recordings and will not have a big impact on the results. The recordings were preprocessed the same way as earlier: each minute in the recording got framed with a window length of 60 ms and a hop length of 50 ms to have some continuity. Each signal piece got then downsampled to 8 kHz, filtered by a second-order Butterworth high-pass filter with a cutoff frequency of 80 Hz, normalized to the [-1, 1], and fed to the model in time-order. Table 8 shows the proportion of detected BS and NBS frames in the different recordings, BS frames did not occur as much as the non-bowel sound. As previously, more bowel sounds were detected from microphone 4 than from microphone 3, for all the subjects.

As the subjects followed the protocol shown in Figure 3.5 it can be of interest to see how the bowel sound activity is during the fasting period, meal period and in digestion period. Figure 5.6, 5.7 and 5.8 shows the occurrence of bowel sounds by subject 1 and 2, 3 and 6, and 7 and 9, respectively. The blue, green and red bar identifies one bowel sound like three, four and five consecutive BS frames, respectively. This was done to see if the number of consecutive frames made any differences. If it is occurring a long bowel sound in the recording, it would be identified as multiple bowel sounds. So, having predicted the frame series, 11111111100001111, will result in a total of 4, 3, and 1 number of bowel sounds for the red, green, and blue bars, respectively. The trend is mostly

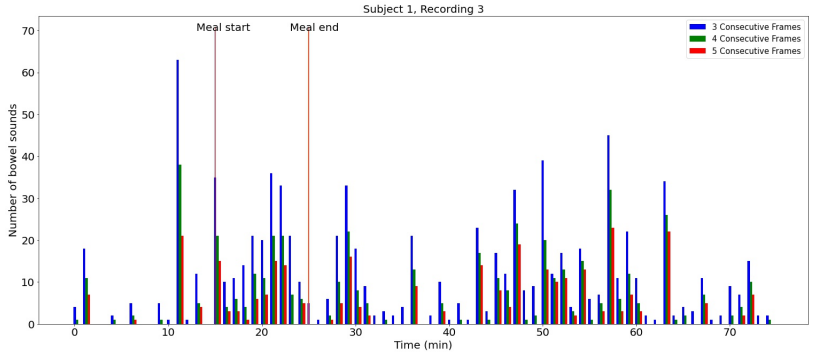
Subject and microphone position	BS	NBS
S1R3	7.79%	92.21%
S1R4	9.94%	90.06%
S2R3	5.8%	94.2%
S2R4	6.4%	93.6%
S3R3	3.89%	96.1%
S3R4	5.32%	94.68%
S6R3	3.2%	96.8%
S6R4	11.1%	88.9%
S7R3	5.00%	95.00%
S7R4	7.60%	92.4%
S9R3	5.14%	94.86%
S9R4	15.4%	84.6%

Table 8: Results of how many BS and NBS were detected for each recording by the model.

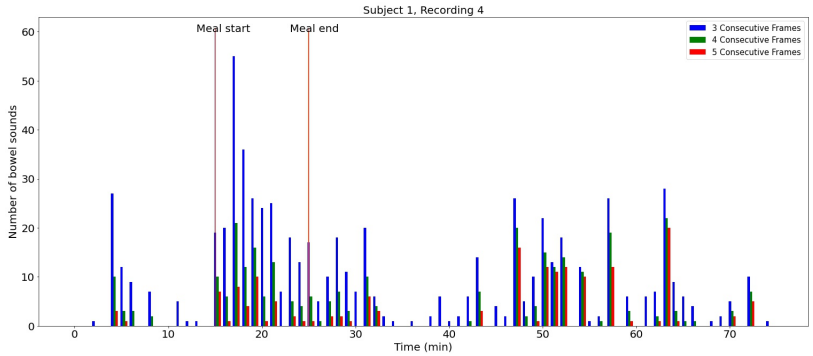
the same for all of the recordings, the number of bowel sounds increases during the meal or digestion.

Analysis of the bar graphs when one bowel sound is identified as fixed number of consecutive BS frames:
Figure 5.6a shows the bar graph of subject 1 and microphone 3. There is registered a high peak of bowel sounds at 11 minutes with detection of around 60 bowel sounds. This is the highest peak in the whole recording. However, besides this peak, there is a second high peak at meal start with detection of around 40 bowel sounds. During the meal period, the detection of bowel sounds starts to increase gradually from 16 minutes to 21 minutes before it starts to gradually decrease again. For the recording collected from LLQ, shown in Figure 5.6b, the highest peak happens during the meal period in 17 minutes with the detection of 60 bowel sounds.
For subject 2, and microphone 3, shown in 5.6c the highest peak is registered at the end of the meal at 25 minutes with detection of over 100 bowel sounds. However, before this peak, during the fasting and meal period, there is not following any patterns of registered bowel sounds. When it comes to microphone 4, shown in Figure 5.6d, the highest peak is also at the meal end.
For subject 3, shown in Figure 5.7a and 5.7b, both collected recordings follows the same pattern. The detection of bowel sounds starts to increase during the meal period with the highest peak at 22 minutes with the detection of 138 bowel sounds and 145 bowel sounds for microphone 3 and 4, respectively.
For subject 6, microphone 3, shown in Figure 5.7c, there is bowel sound activity between 9 minutes and 40 minutes. The highest peak is registered during

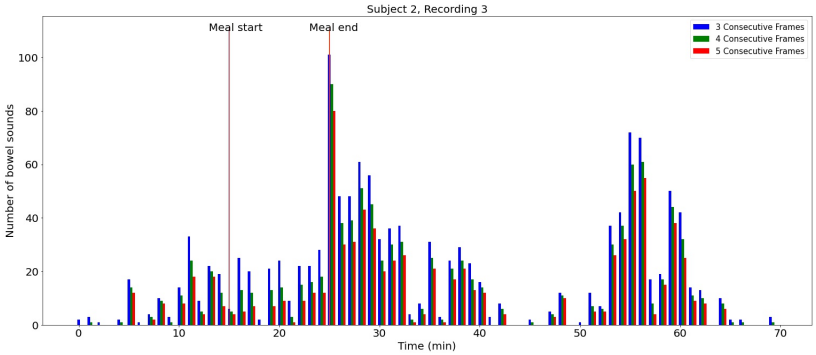
the meal at 22 minutes with the detection of 110 bowel sounds. For recording collected from microphone 4, shown in Figure 5.7d, the bowel sound activity follows a pattern, the registered bowel sounds typically get higher during the meal with a peak of 160 registered bowel sounds in 26 minutes.
For subject 7, microphone 3, shown in Figure 5.8a, there is registered high bowel sound activity in the first minutes of the fasting period. Besides this, there are also registered bowel sounds that follow a trend during the meal. However, the peaks are higher in the fasting period than during the meal. For microphone 4, in Figure 5.8b, the same trend is followed, but the detection of bowel sounds is highest during the meal period.
For subject 9, microphone 3, shown in Figure 5.8c, there is only bowel sound activity during the meal. The highest peak is at 16 minutes with a registered 145 number of bowel sounds. For microphone 4, shown in Figure 5.8d, bowel sounds are detected a little everywhere in the recording, but the highest values occur during the meal period.



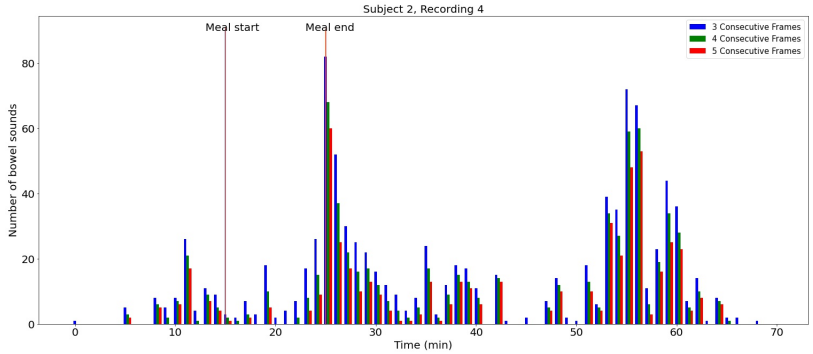
(a) Subject 1, microphone 3.



(b) Subject 1, microphone 4.

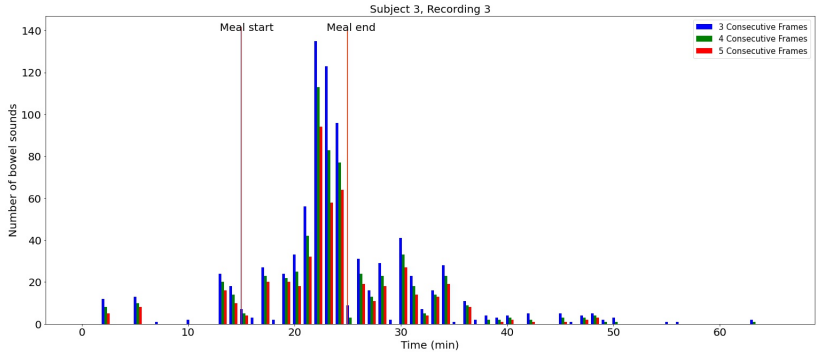


(c) Subject 2, microphone 3.

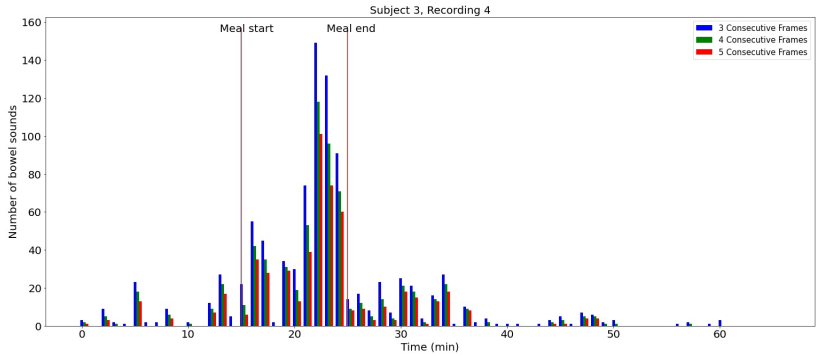


(d) Subject 2, microphone 4.

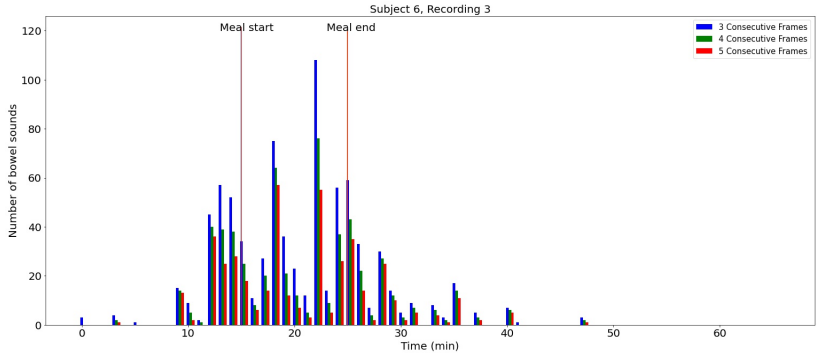
Figure 5.6: Number of bowel sounds in each minute for subject 1 and 2. The y-axis shows the number of bowel sounds in each minute and the x-axis shows the time in minutes. The two red lines show the meal start at 15 minutes and the meal end at 25 minutes. One bowel sound is identified as three, four and five consecutive BS frames in blue, green and red, respectively.



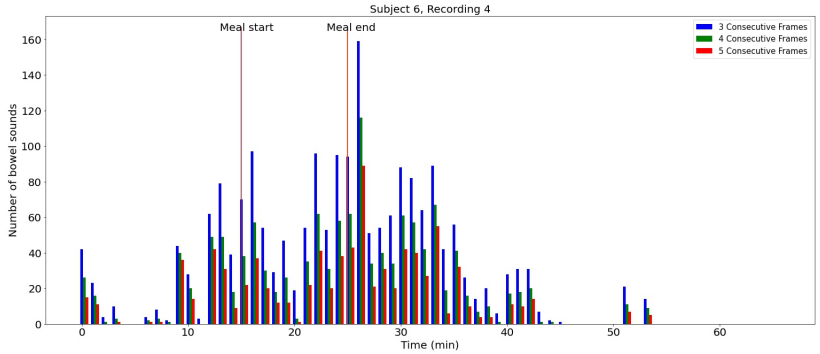
(a) Subject 3, microphone 3.



(b) Subject 3, microphone 4.

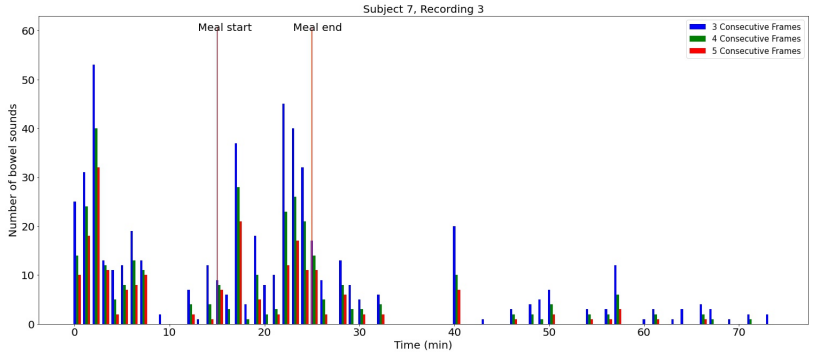


(c) Subject 6, microphone 3.

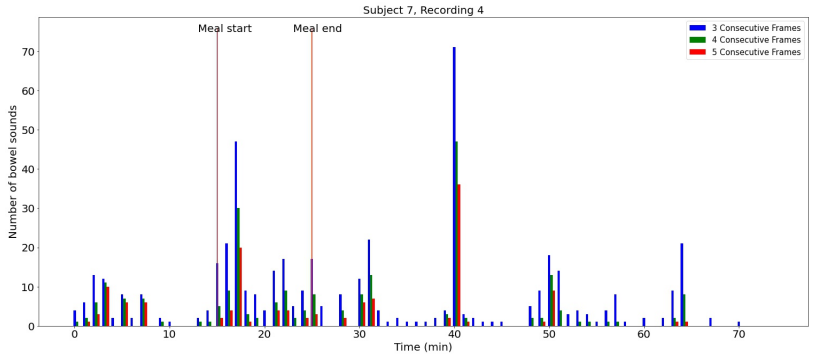


(d) Subject 6, microphone 4.

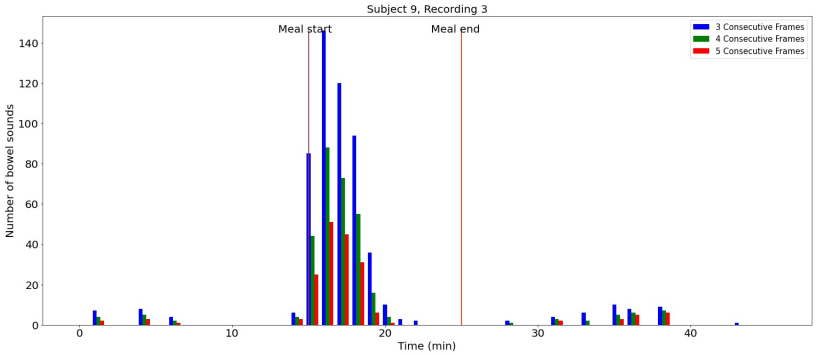
Figure 5.7: Number of bowel sounds in each minute for subject 3 and 6. The y-axis shows the number of bowel sounds in each minute and the x-axis shows the time in minutes. The two red lines show the meal start at 15 minutes and the meal end at 25 minutes. One bowel sound is identified as three, four and five consecutive BS frames in blue, green and red, respectively.



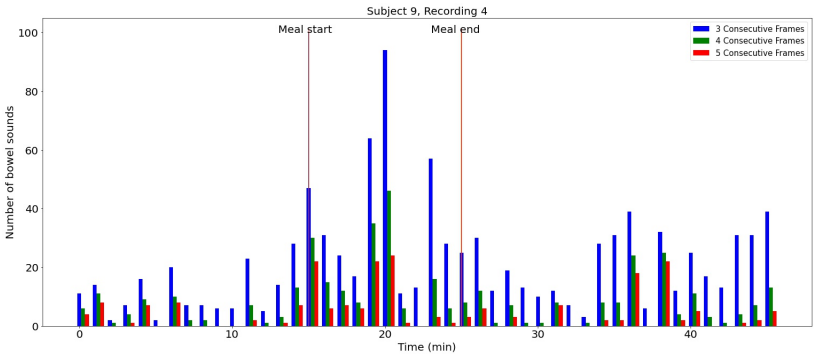
(a) Subject 7, microphone 3.



(b) Subject 7, microphone 4.



(c) Subject 9, microphone 3.

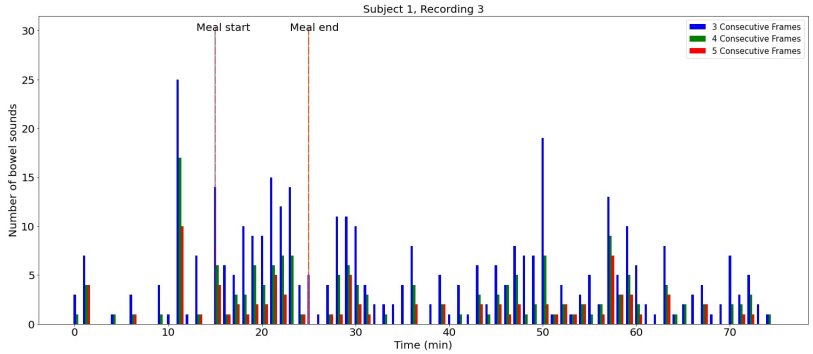


(d) Subject 9, microphone 4.

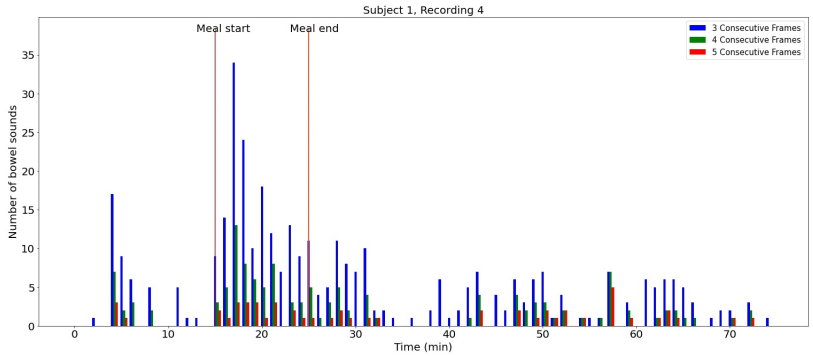
Figure 5.8: Number of bowel sounds in each minute for subject 7 and 9. The y-axis shows the number of bowel sounds in each minute and the x-axis shows the time in minutes. The two red lines show the meal start at 15 minutes and the meal end at 25 minutes. One bowel sound is identified as three, four and five consecutive BS frames in blue, green and red, respectively.

For comparison purposes, it can be interesting to see how many bowel sounds occur when waiting on a non-bowel sound. Figure 5.9, 5.10 and 5.11 shows the bowel sound activity by subject 1 and 2, 3 and 6, and 7 and 9, respectively. One bowel sound is identified as at least three, four and five consecutive BS frames. Meaning there is NBS before and after the identified bowel sound. Having the same prediction series as earlier, 1111111100001111, will result in a total of 2, 2, 1 number of bowel sounds for the red, green and blue bar, respectively. The x-axis shows the time in minutes. Also here, the trend is mostly the same for all of the recordings, the number of identified bowel sounds increases during the meal or digestion. In general, the number of bowel sounds has decreased uniformly as one bowel sound can be identified as one long bowel sound by having many consecutive 1's.

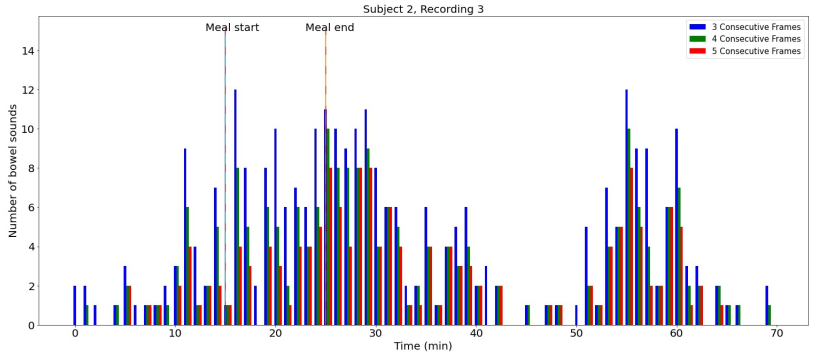
Analysis of the bar graphs when one bowel sound is identified as a series of consecutive BS frames:
For subject 1, recording 3 and 4 in Figure 5.9a and 5.9b follows the same trend as for previous bar graph.
For subject 2, recording 3 shown in Figure 5.9c has a different peak at 16 minutes instead of 25 minutes. Also, this peak does not stand out as much as the others as the top that was found in the second case. However, there are detected more bowel sounds after meal start than during the fasting period. Microphone 4 shown in Figure 5.9d follows the same trend as for previous case.
Subject 3 and 6 for both of the recordings shown in Figure 5.10a, 5.10b, 5.10c and 5.10d follows the same trend as in the previous identification of one bowel sound.
Subject 7 recording 3 shown in Figure 5.11a has the highest peak during the meal period while this occurs at the start of the fasting period for the other case. This means the duration is longer during the meal than in the fasting period. For recording 4 shown in Figure 5.11b, the same trend is followed as in the previous case. Subject 9 for both of the recordings, shown in Figure 5.11c and 5.11d, follows the same trend as earlier.



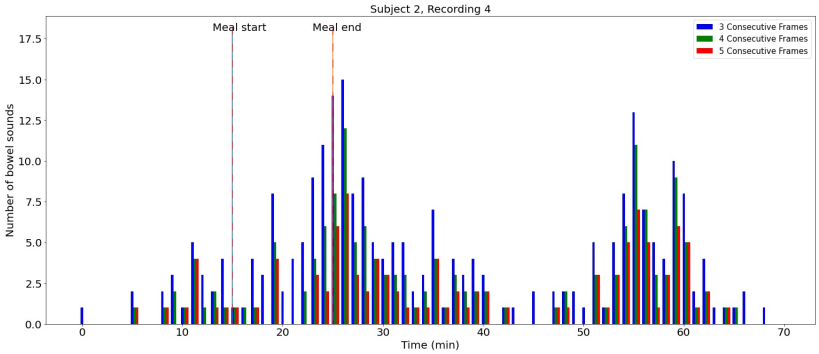
(a) Subject 1, microphone 3.



(b) Subject 1, microphone 4.

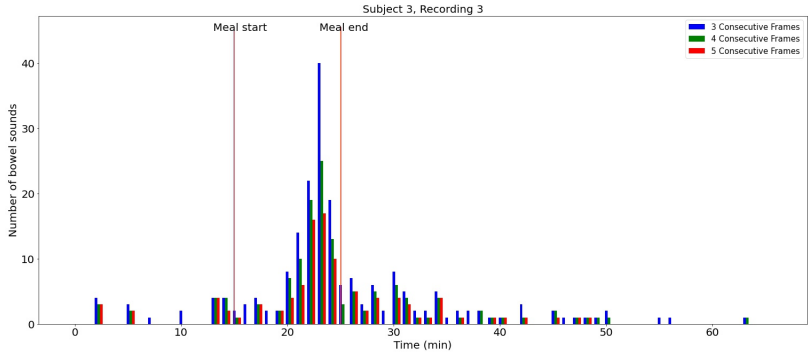


(c) Subject 2, microphone 3.

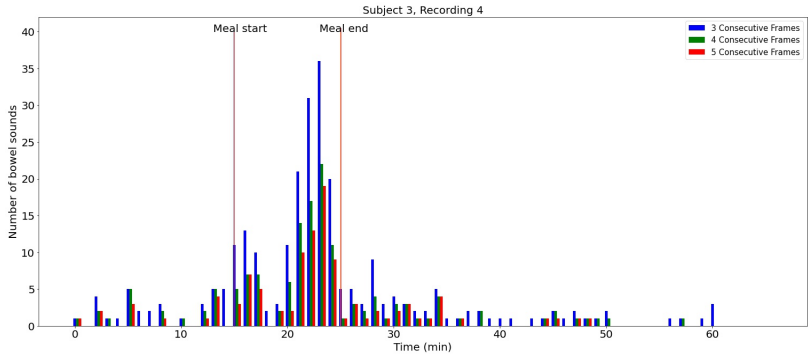


(d) Subject 2, microphone 4.

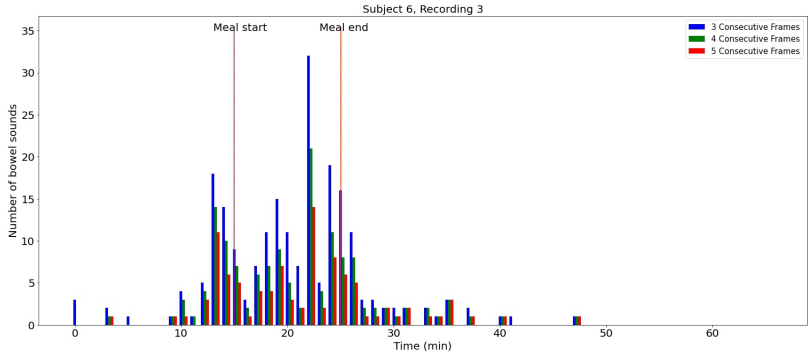
Figure 5.9: Number of bowel sounds in each minute when waiting on a NBS for subject 1 and 2. One bowel sound is identified as at least 3, 4, and 5 consecutive BS frames by a blue, green and red bar, respectively. The two red lines shows the meal start at 15 minutes and the meal end at 25 minutes.



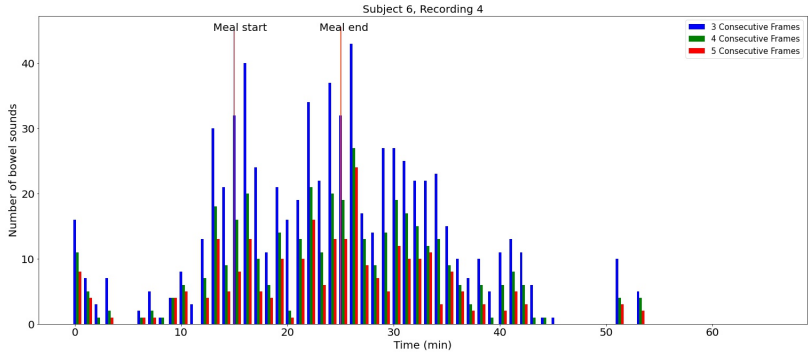
(a) Subject 3, microphone 3.



(b) Subject 3, microphone 4.

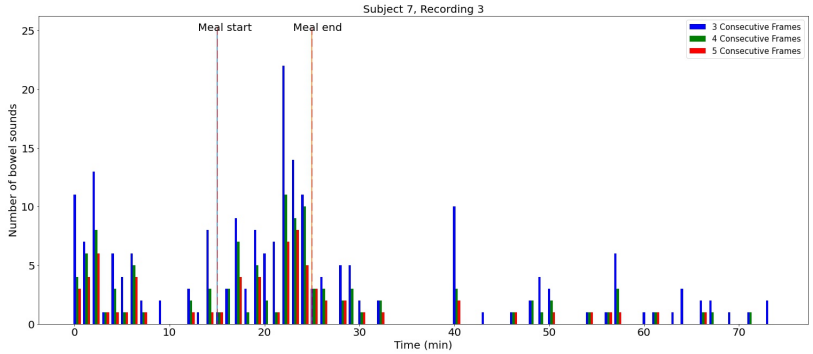


(c) Subject 6, microphone 3.

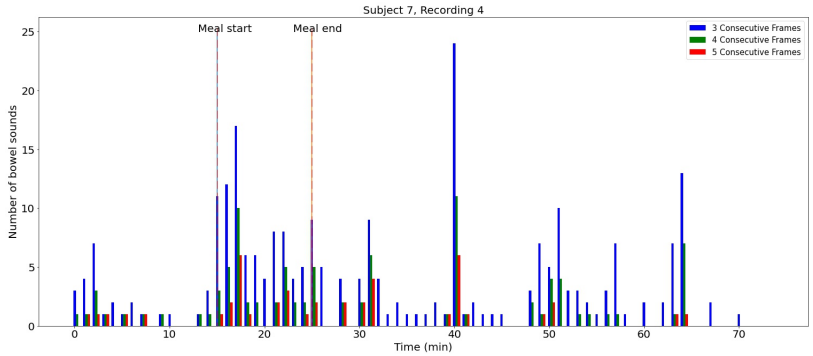


(d) Subject 6, microphone 4.

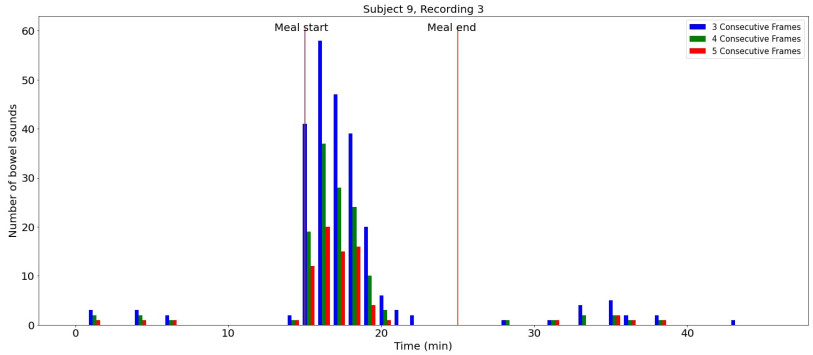
Figure 5.10: Number of bowel sounds in each minute when waiting on a NBS for subject 3 and 6. One bowel sound is identified as at least 3, 4, and 5 consecutive BS frames by a blue, green and red bar, respectively. The two red lines shows the meal start at 15 minutes and the meal end at 25 minutes.



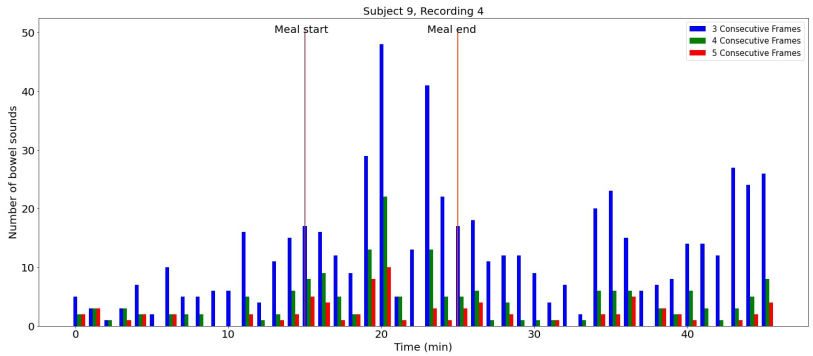
(a) Subject 7, microphone 3.



(b) Subject 7, microphone 4.



(c) Subject 9, microphone 3.



(d) Subject 9, microphone 4.

Figure 5.11: Number of bowel sounds in each minute when waiting on a NBS for subject 7 and 9. One bowel sound is identified as at least 3, 4, and 5 consecutive BS frames by a blue, green and red bar, respectively. The two red lines shows the meal start at 15 minutes and the meal end at 25 minutes.

Due to time constraints, the time windows in the recordings were not labeled as BS/NBS which makes it difficult to verify the predictions from the bar plots. Therefore, another method for verification was done. Everytime the model predicts the first bowel sound (three consecutive frames as BS) in each minute, the current frame number was saved. The various frame numbers got converted to the corresponding time using the function *frame_to_sec(frame)* shown in Listing 2. The function has the input *frame* which is the frame number and returns the time in seconds. All times were converted to a time point in the format *hh: mm: ss* using the function *timedelta* from Datetime package [70]. To evaluate the predictions, the recordings were listened to at the various saved time points. This was done by a non-expert and sometimes, it was difficult to evaluate if the given time was a bowel sound or not. The table 9 shows the time points where the model predicts Bowel sound for subject 9, microphone 3. The red color means the sound at the given time did not sound like a bowel sound. 3 out of 19 were false predictions, meaning 15.8% of the predictions did not sound like bowel sounds. This was done for all of the recordings and is added to Appendix A.1. Table 10 shows the proportion of the true predictions in the different recordings. The true predictions in all of the recordings are over 75%.

```
def frame_to_sec(frames):
  if frames>=1199:
    min_frame=int(frames//1199)
    sec_frames=frame-min_frame*1199
    if sec_frames==0:
      time=min_frame*60+(sec_frames-1)*0.05+0.06
    else:
      time=(frame-1)*0.05+0.06
  return time
```

Listing 2: Python code that converts number of frames to time in seconds.

0:01:17.26	0:04:22.31	0:06:16.81	0:14:05.96
0:15:01.56	0:16:00.06	0:17:01.21	0:18:00.31
0:19:05.96	0:20:05.36	0:21:29.56	0:22:51.96
0:28:52.41	0:31:33.81	0:33:00.46	0:35:11.61
0:36:25.06	0:38:16.46	0:43:43.76	

Table 9: Different time points where the detector has predicted the first bowel sound in each minute for subject 9, recording 3. The red color is the wrong prediction.

Subject	recording	True prediction	recording	True prediction
1	mic 3	98.3%	mic 4	77.6%
2	mic 3	90%	mic 4	89.5%
3	mic 3	88.1%	mic 4	92.3%
6	mic 3	90.0%	mic 4	82.6%
7	mic 3	84.8%	mic 4	81.0%
9	mic 3	84.2%	mic 4	87.0%

Table 10: The proportion of true predictions of the first time point the detector predicts a bowel sound (at least three consecutive frames) in each minute.

5.5 Discussion

The results obtained when testing the model on the test set were good. It got an overall accuracy of 87%, with a recall of 88% and 85% for BS class and NBS class, respectively. The model did also beat a no-skilled model and performed well with an AUC of 94.9%. However, the results on the test set were worse than in the previous case. This may be because NBS segments in this section were carefully selected so it was not a silent period to avoid making a sound detector. The disadvantage of doing this is the CRS can be labeled as NBS and therefore cause worse results than found in the research [12] which got over 90% accuracy when using the same classifier setup. Therefore, the results may get improved if a clinical expert verifies the segments that are unsure.
When testing the model on the collected data, the occurrence of BS frames was much less than NBS frames, as expected. This is an improvement from the previous case as the detector also detected the noise/sound as BS. The occurrence of bowel sounds was also generally low in the fasting period than during the meal which is promising as there may be a possibility of detecting when the food intake happens. This observation supports the results achieved in [13] and [71]. However, this did not apply to all recordings. Some of them, as the subject 7, recording 3, had the highest detection bowel sound activity at the start of the recording when one bowel sound was identified as three consecutive BS frames. The disadvantage of choosing this identification is a long bowel sound will be counted as multiple bowel sounds. By changing the identification to a bowel sound is identified as at least three consecutive frames and has both NBS frames in front and behind the prediction series, the results changed a bit as a whole series of 1's will only be counted as one bowel sound. When this identification was applied, the highest occurrence of bowel sounds for the subject 7 recording 3 was during the meal in 22 minutes. The different number of consecutive BS frames to identify one bowel sound did not make any big changes in the results.
It is shown from the verification method used to evaluate the detections of the bowel sounds that at least 75% were true predictions. Regardless, in some

cases, it was difficult to say whether the given time point, contained a bowel sound or noise as mentioned earlier. Therefore, the evaluation should be done by a clinical expert to get a more precise result. Also, another limitation is only the first bowel sound that occurred in the minute was evaluated, there are plenty of bowel sounds that can happen after the saved time point. Lastly, the predictions of non-bowel sounds were not evaluated which can lead to many missing bowel sounds.

6 Feasibility of early meal detection

The results from the previous section showed that the bowel sound activity usually followed a pattern during the given protocol. The number of bowel sounds increased during or after food ingestion but did not apply to all of the recordings. Therefore, in this section, different acoustic features will be calculated on the detected bowel sounds to see if there are any differences in fasting state and during the meal. The idea is to have support for a meal detector. The implemented detector from the previous section will be used to detect the bowel sounds.

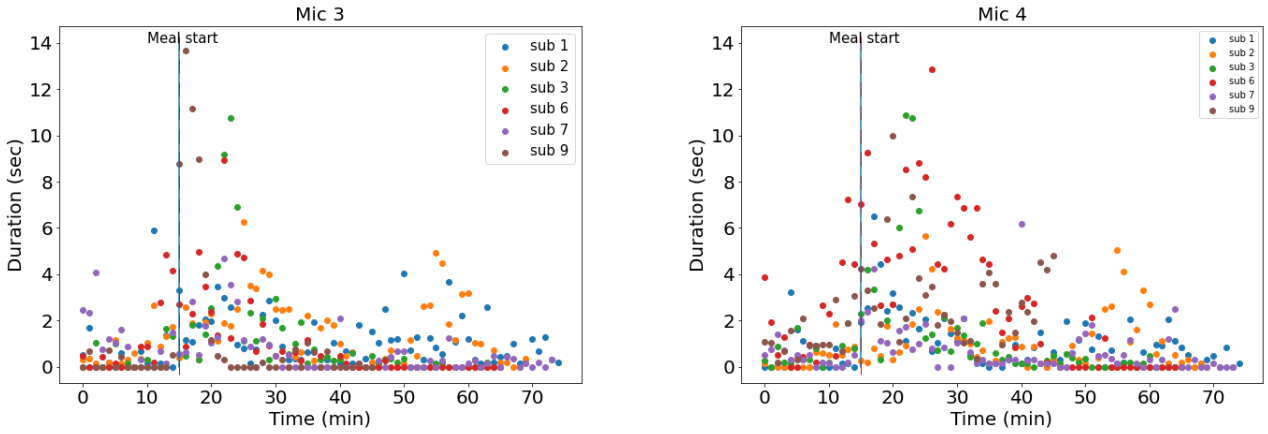
6.1 Method

The same 12 recordings from the previous section were included further in this analysis. The recordings were quantized, segmented, preprocessed and extracted features as earlier and fed in time series to the detector. For each time the model produced at least three consecutive frames as BS, the SC, SBW and duration were calculated and extracted. The SC and SBW were calculated using the equations 6 and 7, respectively. The duration was calculated by converting the series of 1's to seconds using the function *frame_to_sec* presented in Listing 2. Also, the total duration of the bowel sounds in each minute was calculated by adding the extracted durations in each minute.

6.2 Results

Figure 6.1a and 6.1b shows a scatter plot of the total duration of bowel sounds in each minute for microphone 3 and 4, respectively. The colors blue, orange, green, red, purple and brown corresponds to the subject 1, 2, 3, 6, 7 and 9, respectively. The total duration in each minute starts to increase at meal start at 15 minutes for almost all of the recordings. From the recordings collected by microphone 3, the longest duration is found from subject 9 at 18 minutes with a duration of approximately 14 seconds, followed by subject 3 with the highest peak at 24 minutes with a duration of approximately 10 seconds. During the fasting period, subject 7 has the highest peak at 3 minutes with a duration of approximately 4 seconds, during the meal for the same subject, the highest peak is at 22 minutes with a duration of approximately 4.5 seconds. From the recordings collected by microphone 4, the longest duration is found from subject 6 at 26 minutes with a duration of approximately 12.5 seconds, followed by subject 3 with a duration of approximately 11 seconds at 25 minutes. During the fasting period, subject 6 has the highest peak in the first minute with a duration of approximately 4 seconds. However, during the meal, the highest duration is approximately 9 seconds at 17 minutes. Also, a plot of the duration

of each detected bowel sound is added to Appendix A.2. Almost all the subjects are following the same pattern: the duration of many of the predicted bowel sounds right before or during the meal are longer than the ones in the fasting period except for subject 7 which has a long duration in the fasting period.



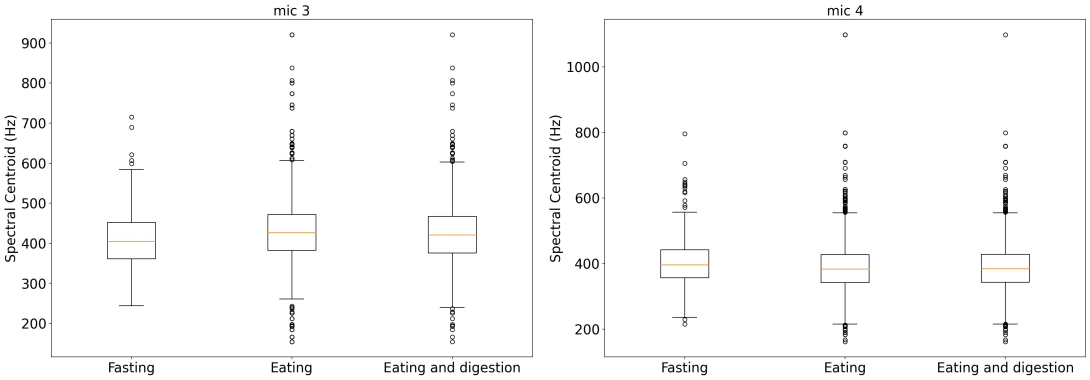
(a) Microphone 3.

(b) Microphone 4.

Figure 6.1: Total duration of detected bowel sounds in each minute. The y-axis shows the total duration in seconds and x-axis shows the time in minutes.

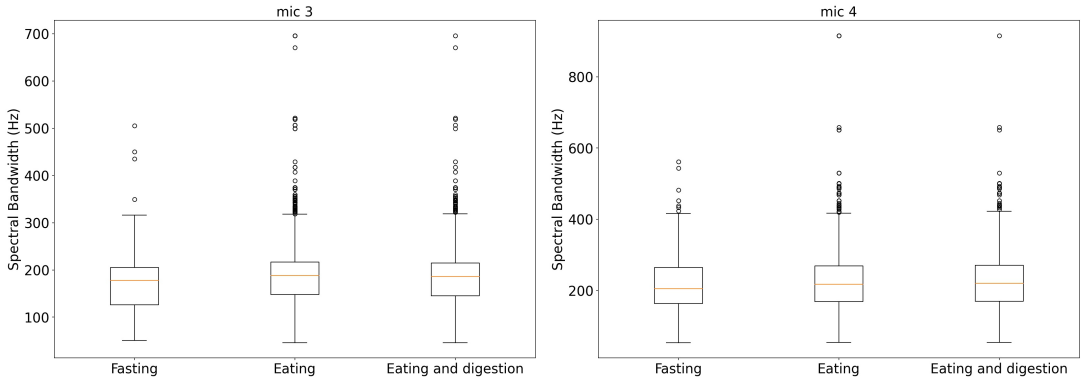
Figure 6.2a and 6.2b shows a box plot of SC during the different states in the protocol for microphone 3 and 4, respectively. The calculated SC from all of the subjects during the fasting period (<15 minutes), only eating (15 to 25 minutes), and eating and digestion (15 minutes>) were put together to see if there were any differences. From the box plots, it does not look like it is a difference between the fasting period and the eating period, the interquartile range seems to be quite the same, 360 Hz-460 HZ. Also, many outliers have been found during and after food intake.
In addition, a plot was created for each subject and microphone showing the calculated SC for each detected bowel sound which is added to Appendix A.3.

Figure 6.3a and 6.3b shows a box plot of SBW during different states of the protocol for microphone 3 and 4, respectively. The calculated SBW from all of the subjects during the fasting period (<15 minutes), only eating (15 to 25 minutes), and eating and digestion (15 minutes>) were put together to see if it is possible to distinguish between these states. Also, a plot was created for each collected recording showing the calculated SBW for each detected bowel sound. This is added to Appendix A.4.



(a) Microphone 3. **(b)** Microphone 4.

Figure 6.2: Box plot of calculated SC during fasting, eating and digestion.



(a) Microphone 3. **(b)** Microphone 4.

Figure 6.3: Box plot of calculated SBW during fasting, eating and digestion.

6.3 Discussion

The total duration of the detected bowel sounds in each minute increased close to the meal or right after food intake for almost all of the subjects. The reason for the former can be when the subjects are presented a nice food of their own choice close to the food intake, they get increased saliva in their mouth as explained in Pavlov's Dogs Study [72], the stomach prepares to digest the food and begins to make some sounds [73]. This also applies to the duration of each bowel sound that is detected - the duration increases close to the meal or right after food intake. However, there is an individual who stands out, the duration of each detected bowel sound for subject 7 is high at the start of the recording from both of the microphones, compared to the other subjects. This may be because the individual is making noise such as movement at the start of the recording session as this occurs from both of the microphones and the

detector has not been tested on such sounds.
There was not found any particular trend of the SC and SBW as found in the research [12]. Most of the bowel sounds, had a SC and SBW between 300-550 Hz and 100-300 Hz, respectively. During the fasting period, most of the values are in a specific range, but when it comes to the meal period, some of the values are both higher and lower values are found. This may be because there are many more bowel sounds occurring during the meal which increases the probability of finding different types of bowel sounds and false predictions which can have different values in SBW and SC.

7 Feasibility of distinguish between meal type

For diabetes patients, it is important to know how much insulin should be injected. This is dependent on what type of meal is eaten and how much glucose the food contains. In this section, a hard meal and a soft meal has been eaten by two subjects. The goal is to see if there are any possibilities to differentiate between these meals.

7.1 Method

Two different subjects followed the protocol described in Section 3.2.3. They did the audio experiment twice, where they ate different meals, a soft meal and a hard meal containing oatmeal and salat, respectively. This gives a total of 4 recordings for each meal since two microphones were placed in different positions on the abdomen, as earlier. The subjects are named after the recordings ID which are 1528, 1647, 1601, and 1726. Of these, are subject ID 1528 and 1647 eating a soft meal, while subject ID 1601 and 1726 eating a hard meal. Both of the subjects did the audio experiment earlier and the bowel sound detector has been trained on their data.

The recordings were quantized, segmented, preprocessed and feature extracted as earlier and fed in time series to the detector to produce a time series of BS and NBS predictions. The proportion of BS frames in the different collected recordings was extracted to see if the trend is the same as earlier. Also, the number of bowel sounds has been plotted through time to see if the bowel sounds get more frequent during the meal, as previously found. The acoustic features, SBW and SC has been calculated whenever a bowel sound (at least three consecutive BS frames) has been detected during the meal period to see if it is possible to distinguish between the salad and oatmeal. The duration has also been calculated using the function *frame_to_sec* shown in Listing 2 everytime time a bowel sound has been found during the meal period.

7.2 Results

The proportion of BS frames detected in the different recordings are shown in Table 11. The results are relatively low compared to what was previously found. Recordings collected from microphone 3 had higher percentages of BS frames than microphone 4, which are the opposite case from earlier. The highest and lowest proportion of detected bowel sounds have been found from the subject 1647, recording 3 and subject 1601, recording 4, respectively.

Figure 7.1 and 7.2 show bar plots of the number of predicted bowel sounds

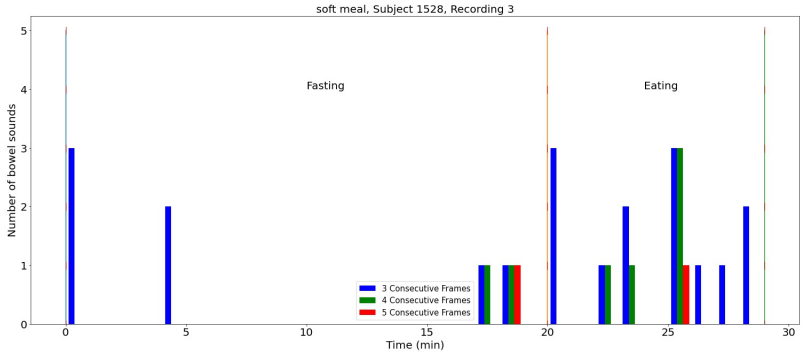
Recording title	BS	NBS
SOFT 1528 mic 3	1.9%	98.1%
SOFT 1528 mic 4	1.3%	98.6%
SOFT 1647 mic 3	5.5%	94.5%
SOFT 1647 mic 4	4.7%	95.3%
HARD 1601 mic 3	2.9%	97.1%
HARD 1601 mic 4	1.0%	99.0%
HARD 1726 mic 3	3.2%	96.8%
HARD 1726 mic 4	2.1%	97.9%

Table 11: Proportion of BS frames found in the different recordings.

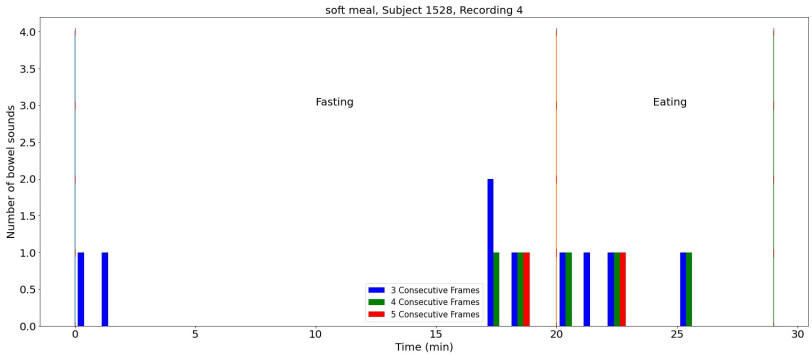
in each minute when subjects are eating a soft meal and a hard meal, respectively. The blue, green and red bar identifies one bowel sound as at least three, four and five consecutive BS frames, respectively. Figure 7.1a and 7.1b show the resulting bar plot by subject 1528 recording 3 and recording 4, respectively. There are more frequent detections during the meal for both of the recordings. Regardless, the number of bowel sounds in each minute does not increase during the meal, the peaks are equally high as in the fasting period. Figure 7.1c and 7.1d show the resulting bar plot by subject 1647 recording 3 and recording 4, respectively. In recording 4, there is an increase in the number of bowel sounds at the start of the meal period and are also more frequent compared to the fasting period. However, in recording 3, this is not the case, the occurrence of the bowel sounds is less during the meal. Figure 7.2a and 7.2b shows the resulting bar plot by subject 1601 recording 3 and recording 4, respectively. In recording 4, there are not detected bowel sounds during the meal. On the other hand, recording 3 has detected bowel sounds during the meal but not so much as in the first minute of the recording. Figure 7.2c and 7.2d shows the bar plot by subject 1726 recording 3 and recording 4, respectively. Both of the recordings has more frequent and higher number of detected bowel sounds during the meal than during the fasting period.

Figure 7.3a and 7.3b shows a scatter plot of the total duration in each minute for microphone 3 and 4, respectively. The colors blue, orange, green and red correspond to the subject 1601, 1726, 1528 and 1647, respectively. There are more variations between the durations and not so strong trend as earlier. A plot of the duration of each detected bowel sound is added to Appendix B. The trend is more clearly now: the duration of each detected bowel sound during the meal tends to be longer than in the fasting period, as previously found.

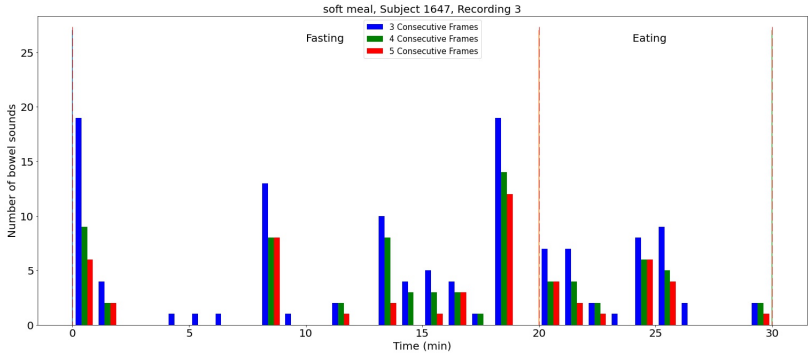
Figure 7.4 shows different box plots of the total duration in a minute in 7.4a, the SC in 7.4b and SBW in 7.4c of the detected bowel sounds during the meal



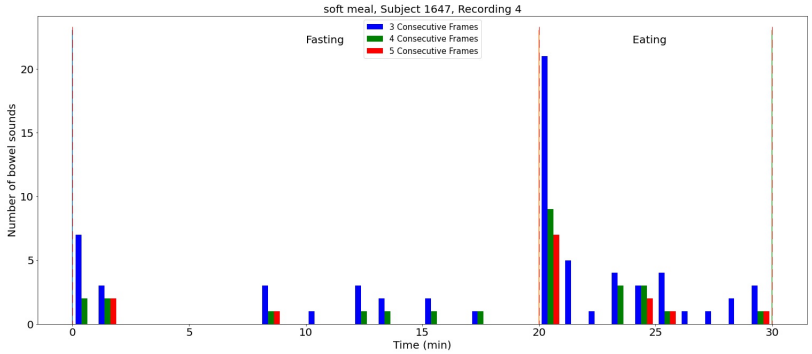
(a) Soft meal, Subject 1528, microphone 3.



(b) Soft meal, Subject 1528, microphone 4.

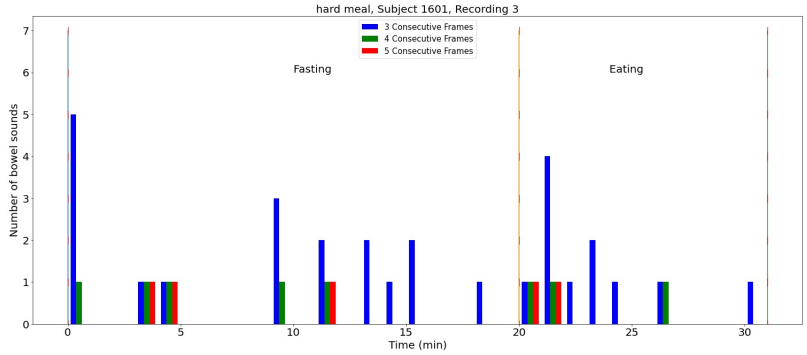


(c) Soft meal, Subject 1647, microphone 3.

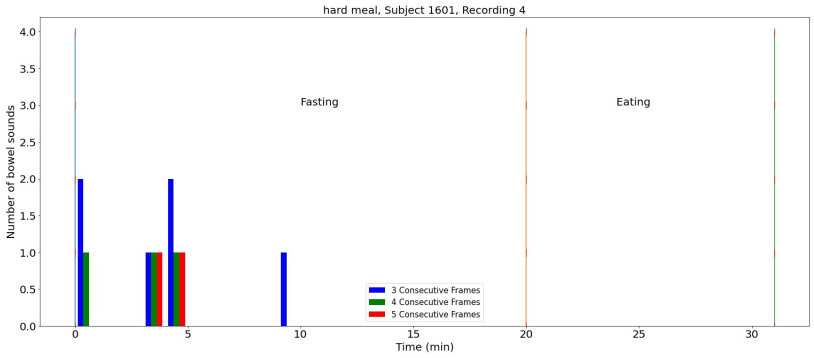


(d) Soft meal, Subject 1647, microphone 4.

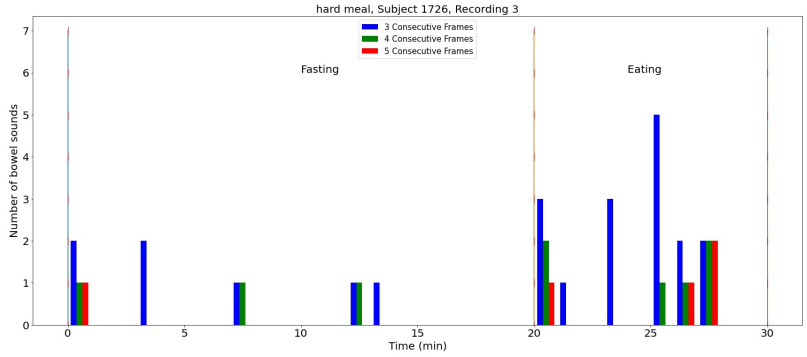
Figure 7.1: Distribution of total detected number of bowel sounds in each minute. One bowel sound is identified as at least three, four and five consecutive BS frames in blue, green and red bar, respectively. The red line at 20 minutes indicates the meal start and the red followed by, indicates the end of the session.



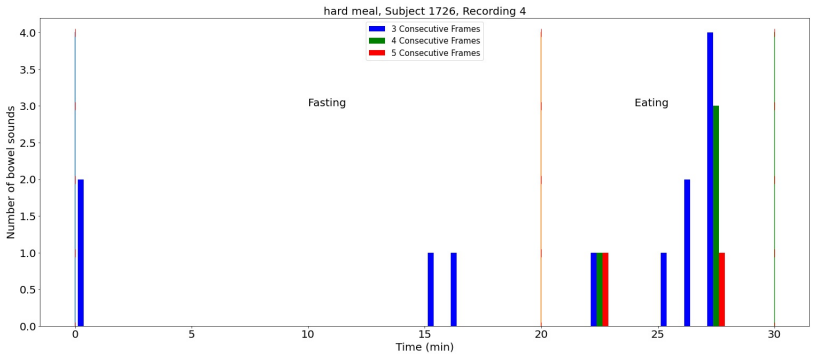
(a) Hard meal, Subject 1601, microphone 3.



(b) Hard meal, Subject 1601, microphone 4.

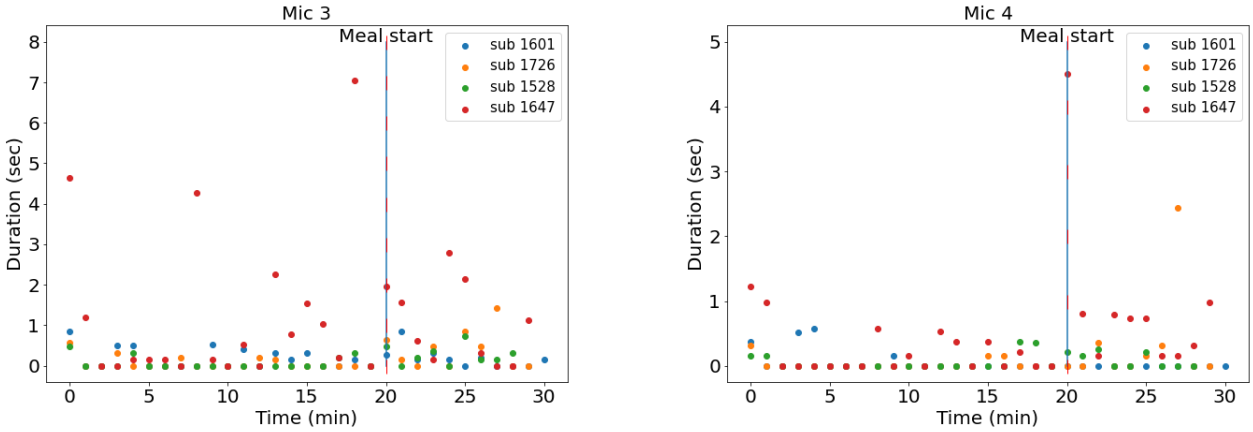


(c) Hard meal, Subject 1726, microphone 3.



(d) Hard meal, Subject 1726, microphone 4.

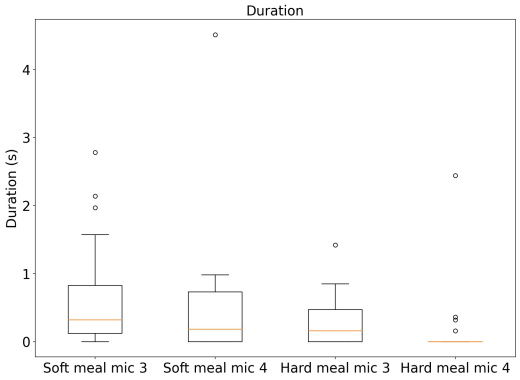
Figure 7.2: Distribution of total detected number of bowel sounds in each minute. One bowel sound is identified as at least three, four and five consecutive BS frames in blue, green and red bar, respectively. The red line at 20 minutes indicates the meal start and the red followed by, indicates the end of the session.



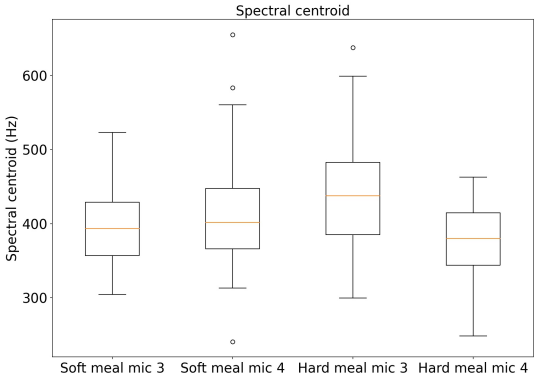
(a) Microphone 3. **(b)** Microphone 4.

Figure 7.3: Total duration of detected bowel sound in each minute.

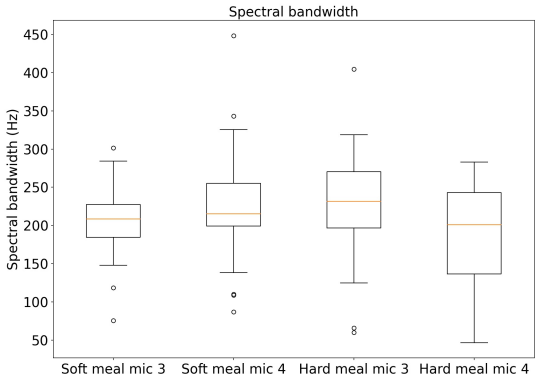
from the different recordings. The total duration in each minute is longer in the soft meal than in the hard meal for both of the recordings. For the hard meal, the SC is higher for the recordings collected from microphone 3 and lower when it is collected from microphone 4. For the SBW, the values are much lower in the hard meal for recordings collected from microphone 4.



(a) Total duration of bowel sounds in each minute.



(b) SC of whenever the model detects a BS.



(c) SBW of whenever the model detects a BS.

Figure 7.4: Box plots of different acoustic features of the detected bowel sounds during the meal.

7.3 Discussion

There was a smaller proportion of BS frames found in the recordings compared to the previous case. This may be because of the shorter meal and digestion period as the BS frames increases after the meal start. However, in some cases, there was minimal detection of bowel sounds during the meal, but the duration of the detected bowel sound during the meal period was often longer than the ones during the fasting period which is the same observation as earlier. Both of the subjects are included in the training data of the model which makes the former results difficult to understand. It may be that something is different from the previous case such as different background noise and different positions of the microphones, making it difficult for the model to recognize the bowel sounds. It has also been confirmed from the subjects afterward that they were unsure if they fasted 3 hours before the experiment. This makes the results difficult to compare to the earlier case. Regardless, the detector should be more

investigated and more subjects should do the audio experiment carefully to see what the reason can be.

When it comes to differentiating the type of meals, there were typical trends in all of the acoustic features during the soft- and hard meal. The duration of the detected bowel sounds per minute during the meal was longer from both microphones when the subjects ate oatmeal than when they ate salad. The SC and SBW were higher and lower when the subjects ate salad and collected from microphones 3 and 4, respectively. This indicates it may be possible to distinguish whenever a person has salad or oatmeal as a meal. However, a limitation is that the results are only based on 10 minutes meal period from two recordings which is minimal. Also, it was not detected any bowel sounds during the meal for one of the subjects which makes this statement weaker.

8 Meal Simulation

A meal simulation can be used to test the feasibility of early meal detection using the implemented bowel sound detector. Earlier, it has been shown the number of bowel sounds usually increases during the meal, but it is of interest to see how the detector behaves when different subjects watch a food video. The bowel sounds may occur frequently due to chemical digestion. The idea is to see if increased bowel sound activity occurs during a food video, and if so, are there any differences between the bowel sounds happening during a food video and the ones during a meal.

8.1 Method

7 different subjects did the audio session by following the protocol described in Section 3.2.4. For the record, the subjects are named after the session ID which is 1108, 1247, 1205, 1452, 1240, 1300 and 1302. The recording session contains the states: 15 minutes fasting, 15 minutes of watching a food video of own choice, 15 min doing nothing and under 15 minutes of eating a meal of own choice. The microphones were placed at the same positions as earlier, in RLQ and LLQ for microphones 1 and 2, respectively. This produces a total of 14 recordings.

All of the recordings were preprocessed and segmented as earlier and fed to the detector in time series which produces predictions of BS/NBS in the corresponding time series. The method of this analysis is the same as in the previous section: the proportion of predicted BS frames got calculated and the number of detected bowel sounds got plotted through the session time with different identification of one bowel sound. Also, each time the detector predicted bowel sound, the total duration in each minute, SBW, and SC got calculated the same way as earlier.

8.2 Results

Table 12 shows the proportion of detected BS and NBS frames. More bowel sounds have been found from the recordings collected from microphone 1 than the recordings collected from microphone 2 except for subject 1452. Subject 1300 found the highest number of BS frames, a proportion of 33.7% from microphone 1, while the lowest amount was found by subject 1452, microphone 1 with a proportion of 1.9%.

Subject	Microphone	BS	NBS	Microphone	BS	NBS
1108	mic 1	14.1%	85.9%	mic 2	7.8%	92.2%
1247	mic 1	3.0%	97.0%	mic 2	2.8%	97.2%
1205	mic 1	3.4%	96.6%	mic 2	2.5%	97.5%
1452	mic 1	1.9%	98.1%	mic 2	2.7%	97.3%
1240	mic 1	8.9%	91.1%	mic 2	5.5%	94.5%
1300	mic 1	33.7%	66.3%	mic 2	28.1%	71.9%
1302	mic 1	30.3%	70.0%	mic 2	24.2%	75.8%

Table 12: The proportion of predicted BS frames in the different recordings.

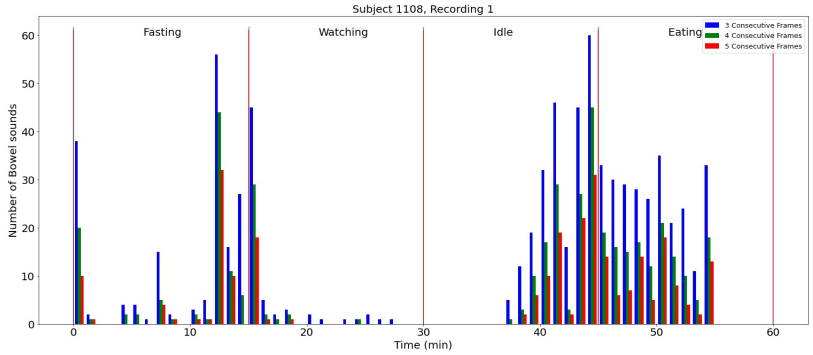
Analysis of the bar graphs when one bowel sound is identified as a series of consecutive BS frames:
Figure 8.1, 8.2, 8.3 and 8.4 shows the number of bowel sounds per minute through the recording session by subject 1108 and 1247, 1205 and 1452, 1240 and 1300 and 1302, respectively.
Figure 8.1a and 8.1b shows a bar plot of subject 1108 microphone 1 and 2, respectively. The number of bowel sounds increases just before the subject watches a food video and eats the meal.
Figure 8.1c and 8.1d shows a bar plot of subject 1247 microphone 1 and 2, respectively. The recording collected from microphone 1 has a peak in the meal period, while the recording collected from microphone 2 has a peak in the fasting period.
Figure 8.2a and 8.2b shows a bar plot of subject 1205 microphone 1 and 2, respectively. Both of the recordings have high peaks during the meal period.
Figure 8.2c and 8.2d shows the bar plot of subject 1452 microphone 1 and 2, respectively. Both of the recordings have frequent relatively high peaks during and after the watching of the food video.
Figure 8.3a and 8.3b shows a bar plot of subject 1452 microphone 1 and 2, respectively. The recording collected from microphone 1 has approximately equal peaks throughout the recording session, but these peaks are more frequent during the meal. The recording collected from microphone 2 has more frequently high peaks that stand out from the rest of the recording during the meal.
Figure 8.3c and 8.3d shows a bar plot of subject 1300 microphone 1 and 2, respectively. Both of the recordings have high peaks throughout the recording session, but the peaks during the meal from recording 1 have the highest peaks during the meal.

Figure 8.4a and 8.4b shows a bar plot of subject 1302 microphone 1 and 2, respectively. As with the previous subject, the recordings have high peaks throughout the session. However, the highest peaks are found during the meal

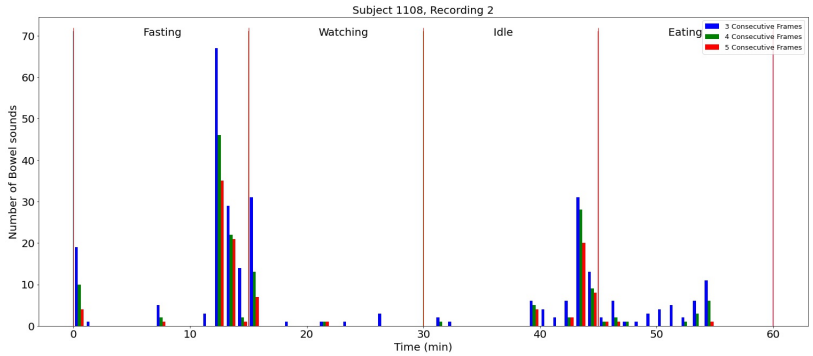
for both of the recordings. It was noticed that the recordings from subjects 1300 and 1302 contained noise such as coughing and movements by using Audacity.

Figure 8.5a and 8.5b shows a scatter plot of the total duration of the detected bowel sounds in each minute for microphone 1 and 2, respectively. The colors blue, orange, green, red, purple, brown, and pink corresponds to the subject 1108, 1247, 1205, 1452, 1240, 1300, and 1302, respectively. There are a lot of spreads in the plot, but one can see that the total duration of the detected bowel sounds during the period when the subjects are having a meal is high. However, it is difficult to say how some of the subjects are performing as subject 1247, 1205, 1452 and 1240. A plot that shows the duration of each detected bowel sound was made separately for each subject and is added to the Appendix C. This shows more clearly that most of the subjects have an increase in the duration of bowel sounds during the meal. For some of the subjects, e.g. subjects 1108 and 1452 have an increase in duration right before or when the subject is watching a food video.

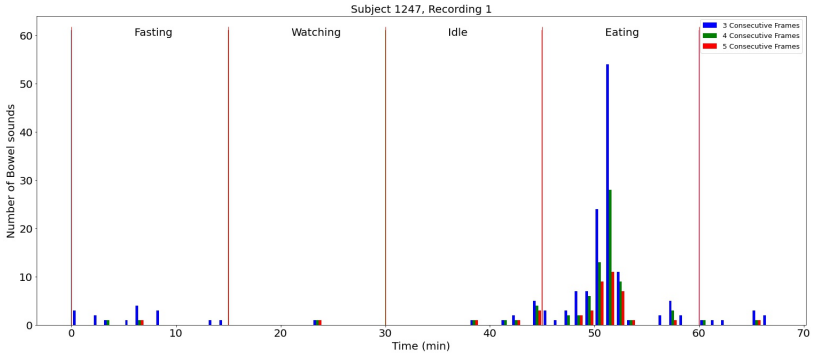
Figure 8.6 shows a box plot of SC and SBW of the detected bowel sounds in the different states of the protocol. It does not seem there is a difference between the bowel sounds occurring in the different phases, the interquartile range is the same for all of the states.



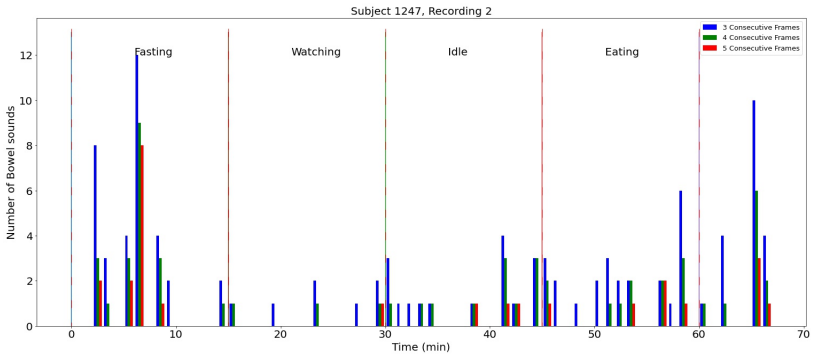
(a) Subject 1108, microphone 1.



(b) Subject 1108, microphone 2.

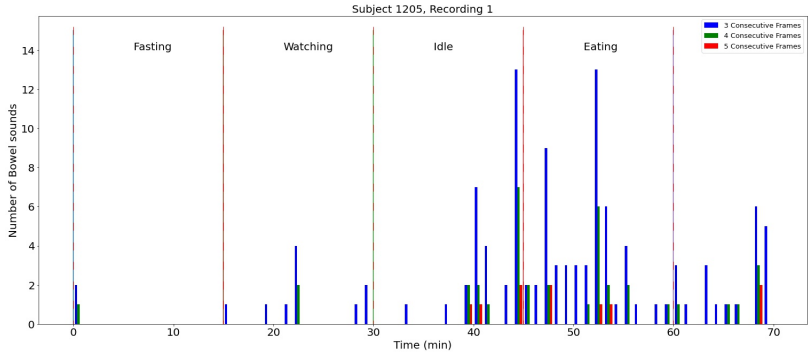


(c) Subject 1247, microphone 1.

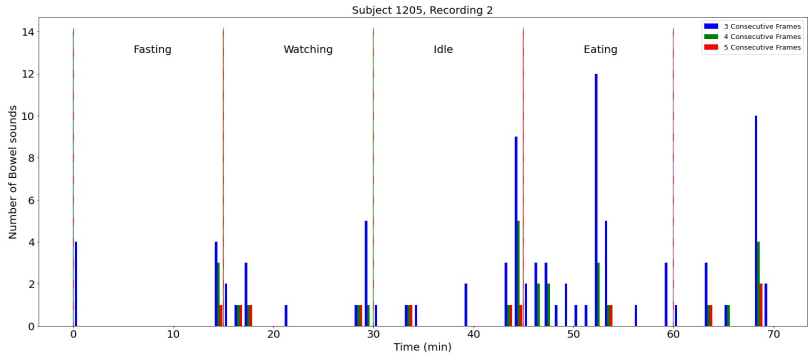


(d) Subject 1247, microphone 2.

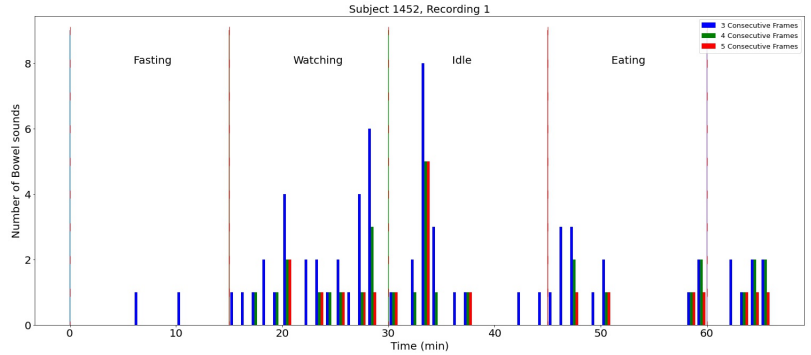
Figure 8.1: Number of bowel sounds in each minute. One bowel sound is identified as three, four and five consecutive BS frames in blue, green and red, respectivly. The red lines at 0 minutes, 15 minutes, 30 minutes, 45 minutes and 60 minutes shows the different phases of the recording session.



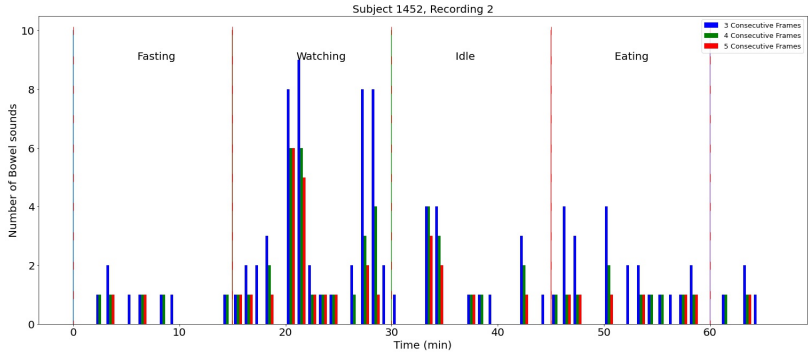
(a) Subject 1205, microphone 1.



(b) Subject 1205, microphone 2.

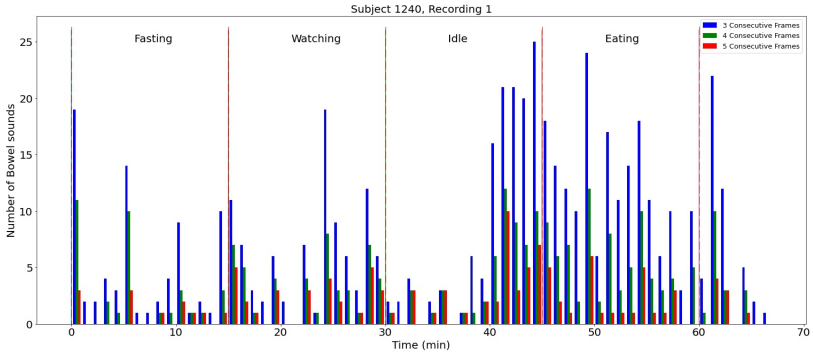


(c) Subject 1452, microphone 1.

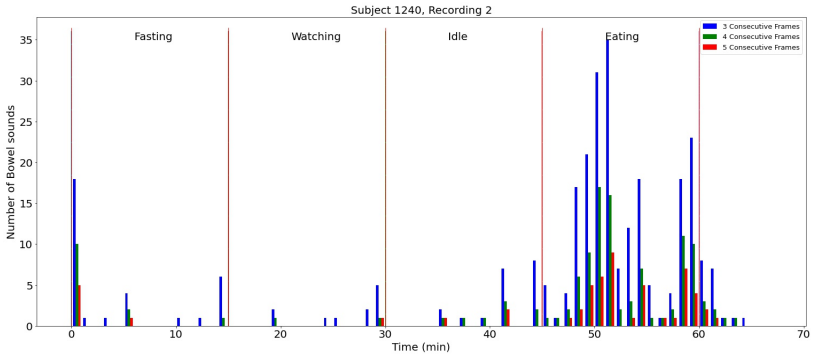


(d) Subject 1452, microphone 2.

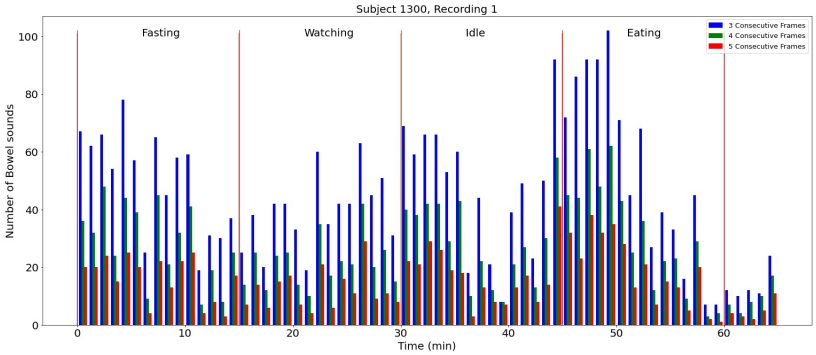
Figure 8.2: Number of bowel sounds in each minute. One bowel sound is identified as three, four and five consecutive BS frames in blue, green and red, respectivly. The red lines at 0 minutes, 15 minutes, 30 minutes, 45 minutes and 60 minutes shows the different phases of the recording session.



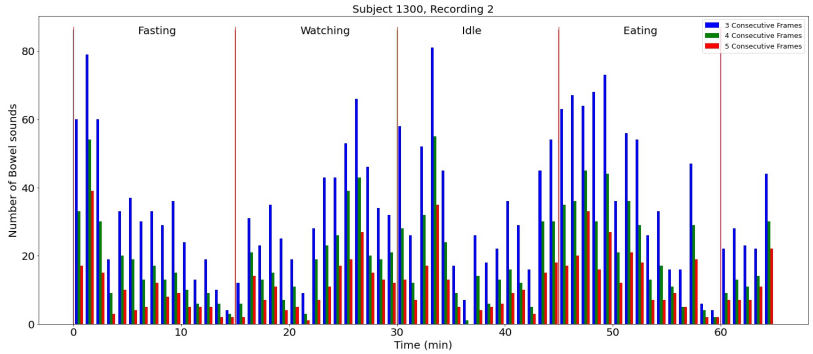
(a) Subject 1240, microphone 1.



(b) Subject 1240, microphone 2.

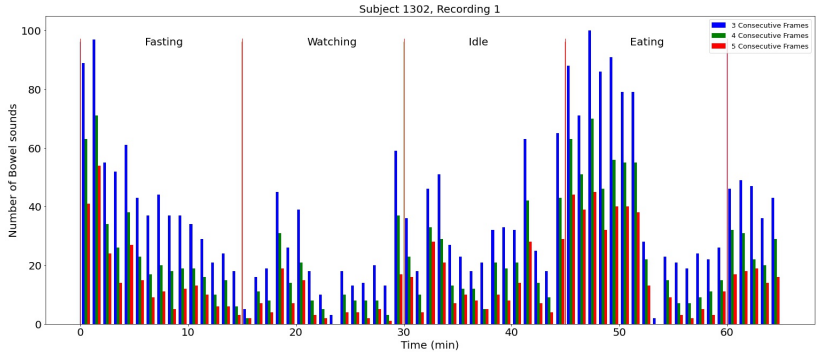


(c) Subject 1300, microphone 1.

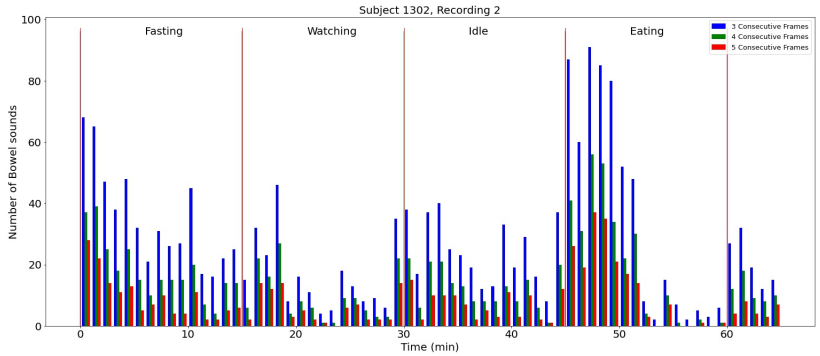


(d) Subject 1300, microphone 2.

Figure 8.3: Number of bowel sounds in each minute. One bowel sound is identified as three, four and five consecutive BS frames in blue, green and red, respectivly. The red lines at 0 minutes, 15 minutes, 30 minutes, 45 minutes and 60 minutes shows the different phases of the recording session.



(a) Subject 1302, microphone 1.



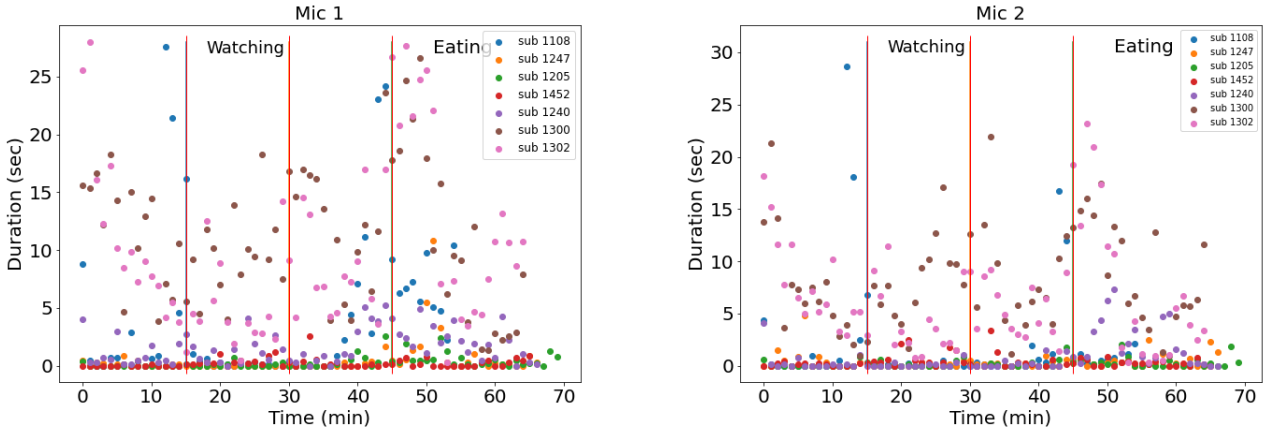
(b) Subject 1302, microphone 2.

Figure 8.4: Number of bowel sounds in each minute. One bowel sound is identified as three, four and five consecutive BS frames in blue, green and red, respectivly. The red lines at 0 minutes, 15 minutes, 30 minutes, 45 minutes and 60 minutes shows the different phases of the recording session.

This may be because the recordings contain noise such as coughing and movement.

8.3 Discussion

The proportion of detected BS frames differed between the subjects. The highest incidence is found from the recordings collected by subject 1300 and subject 1302 with a proportion of over 30%. This may be because the recordings contain noise such as coughing and movements. The bar graphs of these recordings show a more even distribution of the detected number of bowels than earlier results, which makes the statement stronger. The detector should be tested on contaminated recordings to see how sensitive it is to noise. However, the highest peaks are found during the meal for both of the microphones. For the rest of the subjects most followed, the same trend found earlier: the number of bowel sounds increases during the meal. In addition, some of them, such as subject



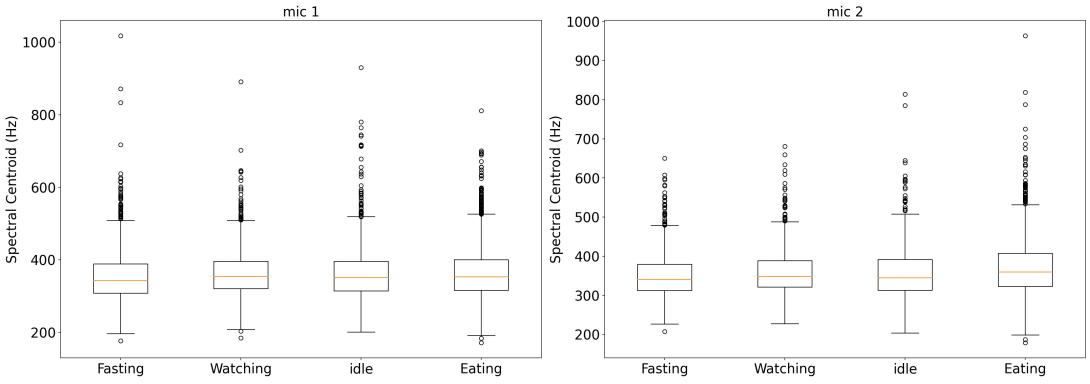
(a) Microphone 1.

(b) Microphone 2.

Figure 8.5: Total duration of the detected bowel sounds in each minute.

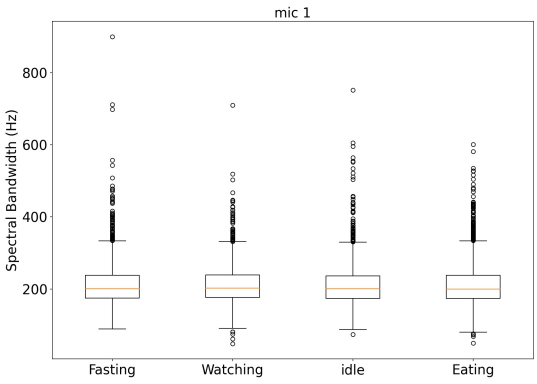
1452 and 1240 has an increase when watching food video.
The duration of each detected bowel sound increased after meal start for most of the subjects, which is the same observation as before. One subject had in a addition, a clearly increase in duration when watching the food video. The reason may be because of physiological reasons, the person gets surprised by the food presented in the video so the chemical digestion starts. The saliva increases in the mouth and stomach are preparing for digesting the meal and therefore starts to make some sounds. This is the same phenomenon which occurs when the subjects are presented for nice food right before meal start which are discussed earlier.

The acoustic features, SC and SBW did not show any particular trend to distinguish between the states in the protocol.

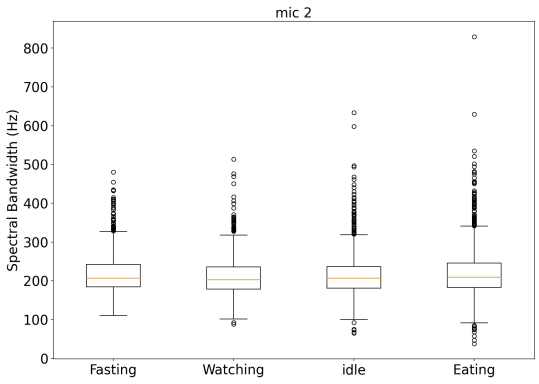


(a) SC microphone 1.

(b) SC microphone 2.



(c) SBW microphone 1.



(d) SBW microphone 2.

Figure 8.6: Acoustic features.

9 Discussion

This chapter summarizes the various discussions from Section 4.5, Section 5.5, Section 6.3, Section 7.3, and Section 8.3. A variety of limitations of the implemented detector, the protocols, and the analysis are also presented as well.

9.1 The implemented BS detector

The best results when training on the test set was achieved by the first model with overall accuracy, precision, and recall of 100% for both of the classes. However, when tested on the collected data, the proportion of the BS frames from the recordings collected from microphone 4 was over 90% which indicates there is a big chance that the model predicts noise as BS. Although, when segments from the acquired data are included in the training set, the retrained model performs poorer in the test set with an overall accuracy of 87%, precision, and recall of 85% and 88% for the BS class, respectively. On the other hand, the proportion of BS frames on the different recordings is much lower, under 15% which supports human physiology. The evaluation of the model on the collected recordings showed over 80% were true predictions. However, a limitation of the chosen evaluation method is that only the first predicted bowel sound in each minute was listened to, meaning the bowel sounds occurring after the first prediction in each minute were not evaluated. The predicted non-bowel sounds were also not evaluated meaning missing bowel sounds may have happened.

9.2 The analysis of meal detection

Feeding the detector frames in time series showed that one specific trend was common, bowel sounds occur more frequently right before or during the meal, which supports human physiology. Also, the duration of each detected bowel sound increased right before or during the meal in mostly all of the collected recordings when the different protocols were followed. This means it may be possible to say if a subject is eating or not. However, the model should be tested in a noisy environment and when the subjects are living normally daily life. Also, the results from the meal simulation showed some of the subjects had an increase in duration and bowel activity right before/during the meal, which may be a limitation.

The results on SC and SBW showed there was not any particular trend between the fasting and eating states. However, during the hard meal, the SBW and SC values tend to be higher than when eating a soft meal for microphone 3. The opposite case was found from microphone 4. Also, the total duration

of the detected bowel sounds in each minute showed it increased when eating a hard meal for both microphones. This shows that it may be possible to distinguish between meals, but several investigations must be done as not many bowel sounds were detected during the meals.

9.3 Limitations of the thesis

Due to time constraints, it was not possible to investigate the detector further. Listening to the recordings, labeling segments as BS/NBS, and evaluating the model by listening to many time points were time-consuming. Also, the labeling and evaluation were not done by a clinical expert which makes the result not so sure. The results were neither analysed by medical professionals.

10 Conclusion

This thesis aimed to implement and design a bowel sound detector and analyze different recordings when different protocols were followed. The final implementation of the detector was done by using data from Youtube and provided by APT. The acquired recordings were collected from two different places in the abdomen, RLQ and LLQ. Each minute got framed with a frame length of 60 ms and a hop length of 50 ms. These were fed to the detector which produced predictions in time series. The goal was to see if early meal detection is feasible. The analysis showed that most, bowel sounds occur more frequently and had a longer duration right before or during the meal for most of the subjects when different protocols were followed. However, the detector is not tested against noise and should be focused on in further work. Also, the collected recordings should be collected in a more noisy environment and when the subjects are living their normal life.

11 Suggestions for future work

This section presents some suggestions for further work and investigations to make the detector more robust and do more analysis to see if it is possible to implement an early meal detector.

11.1 Collection of more data

As discussed, a considerable limitation of the present study is the data set. Due to time constraints, it was not possible to label as many segments and these were neither verified by an expert. In order to make the detector more generalized, more data should be collected from different persons of different ages and health conditions. Also, the labeled segments and the evaluation of the model should be done by a clinical expert. A way of doing this is to include different labeled open-source data sets which are available such as the Bowel sounds data set [74]. A possiblity to label the collected recordings is to use a method developed by [75] which filters bowel sounds using multivariate empirical mode decomposition. The proposed method is tested on a contaminated data set which shows promising performance, nearly 100% of the manually labeled bowel sounds are identified.

11.2 Preprocessing

Each labeled segment got normalized with the highest peak in the signal so it is in the range [-1,1]. The signal of a noise segment in a given period of time contains of random values, meaning the highest peak can be anywhere. Applying peak normalisation can therefore lead to distortions of the NBS signal. Therefore another scaling, as root mean square (RMS) normalisation should be tried out. The peaks in the signal will be adjusted based on its perceived loudness.

11.3 Features

The features in this thesis were focused on Mel-scaled spectrograms, due to the fact it achieved good results from the research of designing a bowel sound detector [12]. However, other studies such as [76] used the features Mel-frequency cepstral coefficients (MFCCs) and power-normalized cepstral coefficients (PNCCs) on a ANN and achieved an accuracy of approximately 90%. Also an interesting approach could be trying to train the data with other networks such as Recurrent Neural Network (RNN), LSTM and Resnet as they are good with features varying in time domain [77].

11.4 Implementation of an meal detector

The results from the analysis of this thesis showed that the duration and the occurrence of the detected bowel sound tend to be longer and more frequent during the meal. A meal detector can be implemented by using these features along with other features such as MFCC (Mel-frequency cepstral coefficients) to see if it is possible to distinguish between meal and non-meal bowel sound segments. An improved bowel sound detector can be used to find the bowel sounds.

Bibliography

[1] Johns Hopkins. The digestive process: What is the role of your pancreas in digestion? https://www.hopkinsmedicine.org/health/conditions-and-diseases/the-digestive-process-what-is-the-role-of-your-pancreas-in-digestion. Accessed: 2022-04-07.

[2] Steven Dowshen. When blood sugar is too high. https://kidshealth.org/en/teens/high-blood-sugar.html. Accessed: 2022-06-08.

[3] NHS. Gestational diabetes. https://www.nhs.uk/conditions/gestational-diabetes/. Accessed: 2022-04-07.

[4] Familydoctor. Insulin therapy. https://familydoctor.org/insulin-therapy/. Accessed: 2022-03-21.

[5] Michael Dansinger. What are insulin pumps? https://www.webmd.com/diabetes/insulin-pump. Accessed: 2022-06-08.

[6] National Institute of Diabetes, Digestive, and Kidney Diseases. Artificial pancreas. https://www.niddk.nih.gov/health-information/diabetes/overview/managing-diabetes/artificial-pancreas. Accessed: 2022-03-21.

[7] Simon O'Connor Nicholas J. Talley. *Clinical examination: a systematic guide to physical diagnosis*. Churchill Livingstone Australia; 7th edition, 2009. Norwegian University of Science and Technology, Trondheim, Norway.

[8] Hanjun Jiang, Shulin Feng, Juzheng Liu, Ping Chen, Binjie Zhu, and Zhihua Wang. Bowel sound recognition using svm classification in a wearable health monitoring system. *Science China Information Sciences*, 61, 08 2018.

[9] Ryunosuke Sato, Takahiro Emoto, Yuki Gojima, and Masatake Akutagawa. Automatic bowel motility evaluation technique for noncontact sound recordings. *Applied Sciences*, 8(6), 2018.

[10] Juzheng Liu, Yue Yin, Hanjun Jiang, Huili Kan, Zongwang Zhang, Ping Chen, Binjie Zhu, and Zhihua Wang. Bowel sound detection based on mfcc feature and lstm neural network. In *2018 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pages 1–4, 2018.

[11] Kang Zhao, Hanjun Jiang, Tao Yuan, Chun Zhang, Wen Jia, and Zhihua Wang. A cnn based human bowel sound segment recognition algorithm with reduced computation complexity for wearable healthcare system. In *2020 IEEE International Symposium on Circuits and Systems (ISCAS)*, pages 1–5, 2020.

[12] Marshall B Ning Wang. Development of a bowel sound detector adapted to demonstrate the effect of food intake. *Science China Information Sciences*, 21, 01 2022.

[13] Konstanze Kölle, Anders Lyngvi Fougner, Reinold Ellingsen, Sven Magnus Carlsen, and Øyvind Stavdahl. Feasibility of early meal detection based on abdominal sound. *IEEE Journal of Translational Engineering in Health and Medicine*, 7:1–12, 2019.

[14] APT. Artificial pancreas trondheim. `https://www.apt-norway.com/`. Accessed: 2021-12-15.

[15] ADAM Health Solutions. Abdominal sounds. `https://medlineplus.gov/ency/article/003137.htm`. Accessed: 2022-04-12.

[16] ADAM Health Solutions. Abdominal sounds. `https://medlineplus.gov/ency/article/003137.htm`. Accessed: 2022-05-10.

[17] Medical News Today. Stomach noises after eating: Causes and treatment. `https://www.medicalnewstoday.com/articles/why-does-my-stomach-make-noise-after-i-eat`. Accessed: 2022-04-16.

[18] Sonia Martell. Liver pain | current health advice, health blog articles and tips. `https://co.pinterest.com/pin/315674255128796850/`. Accessed: 2022-04-12.

[19] Webberley KM Osseiran A Marshall BJ Du X, Allwood G. Bowel sounds identification and migrating motor complex detection with low-cost piezoelectric acoustic sensing device. *IEEE Journal of Translational Engineering in Health and Medicine*, 2018.

[20] Charalampos Dimoulas, George Papanikolaou, and V. PETRIDIS. Pattern classification and audiovisual content management techniques using hybrid expert systems: A video-assisted bioacoustics application in abdominal sounds pattern analysis. *Expert Syst. Appl.*, 38:13082–13093, 09 2011.

[21] Vagisha Gupta, Shelly Sachdeva, and Neha Dohare. *Trends in Deep Learning Methodologies*. Academic press, 2021. P. 183-206.

[22] Robert Triggs. What you think you know about bit-depth is probably wrong. https://www.soundguys.com/audio-bit-depth-explained-23706/. Accessed: 2022-05-20.

[23] Nyquist sampling theorem. http://musicweb.ucsd.edu/~trsmyth/digitalAudio171/Nyquist_Sampling_Theorem.html. Accessed: 2021-10-30.

[24] Dewesoft. Signal filtering, signal suppression, signal processing. https://training.dewesoft.com/online/course/filters. Accessed: 2022-03-03.

[25] Electrical4U. Butterworth filter: What is it? (design applications). https://www.electrical4u.com/butterworth-filter/. Accessed: 2022-06-05.

[26] Xiao Zhang. Normalize data. https://docs.microsoft.com/en-us/azure/machine-learning/studio-module-reference/normalize-data. Accessed: 2021-12-07.

[27] SensiML. Feature (machine learning). https://sensiml.com/entity/feature-machine-learning/. Accessed: 2022-06-08.

[28] Ali N.Akansu Richard A.Haddad. *Multiresolution Signal Decomposition (Second Edition)*. New Jersey Institute of Technology, Newark, NJ, 2001. P. 331-390.

[29] Dewesoft. What is a spectrogram? https://vibrationresearch.com/blog/what-is-a-spectrogram/. Accessed: 2022-03-07.

[30] Nasser Kehtarnavaz. Short-time fourier transform. https://www.sciencedirect.com/topics/engineering/short-time-fourier-transform. Accessed: 2022-06-08.

[31] Music Information Retrieva. Mel frequency cepstral coefficients (mfccs). https://musicinformationretrieval.com/mfcc.html. Accessed: 2021-12-07.

[32] Ketan Doshi. Audio deep learning made simple (part 2): Why mel spectrograms perform better. 2021. Towards Data Science.

[33] librosa development team. librosa. https://librosa.org/doc/latest/index.html. Accessed: 2022-03-31.

[34] librosa development team. librosa.feature.melspectrogram. https://librosa.org/doc/main/generated/librosa.feature.melspectrogram.html. Accessed: 2022-05-20.

[35] A. Eronen W. Theimer, I. Vatolkin. *Definitions of Audio Features for Music Content Description*. Technische universitat dortmund, Dortmund, Germany, 2008. Algorithm Engineering Report.

[36] An introduction to machine learning. https://www.digitalocean.com/community/tutorials/an-introduction-to-machine-learning. Accessed: 2022-04-13.

[37] Daniel Johnson. Unsupervised machine learning: Algorithms, types with example. https://www.guru99.com/unsupervised-machine-learning.html. Accessed: 2022-04-13.

[38] E. Bashier M. Mohammed, M. Khan. *Machine Learning: Algorithms and Applications*. CRC Press, Florida, USA, 2016.

[39] IBM Cloud Education. Supervised learning. https://www.ibm.com/cloud/learn/supervised-learning. Accessed: 2022-06-06.

[40] K. Laajab. *Algorithm design and testing of COVID-19 detection system based on Machine Learning*. Norwegian University of Science and Technology, Trondheim, Norway, 2022. Term project, Norwegian University of Science and Technology, Trondheim, Norway.

[41] Fahmi Nurfikri. An illustrated guide to artificial neural networks. https://towardsdatascience.com/an-illustrated-guide-to-artificial-neural-networks-f149a549ba74. Accessed: 2021-12-15.

[42] B. Marr. What are artificial neural networks - a simple explanation for absolutely anyone. 2018. Forbes.

[43] Analythics Vidhya. Crash course on multi-layer perceptron neural networks. https://machinelearningmastery.com/neural-networks-crash-course/. Accessed: 2021-12-08.

[44] Analythics Vidhya. Understanding neural networks. https://towardsdatascience.com/understanding-neural-networks-677a1b01e371. Accessed: 2021-12-14.

[45] Nishit Jain. An overview of the gradient descent algorithm. https://towardsdatascience.com/an-overview-of-the-gradient-descent-algorithm-8645c9e4de1e. Accessed: 2022-06-06.

[46] Analythics Vidhya. Gentle introduction to the adam optimization algorithm for deep learning. https://machinelearningmastery.com/adam-optimization-algorithm-for-deep-learning/. Accessed: 2021-12-14.

[47] Sarah Schuhegger. Creative commons attribution-sharealike 4.0 international. https://www.researchgate.net/figure/Example-of-a-CNN-for-predicting-the-vehicle-visible-in-an-image-At-the-begi fig3_355391296. Accessed: 2022-06-06.

[48] Mayank Mishra. Convolutional neural networks, explained. 2020. Towards Data Science.

[49] Richard Kinh Gian Do Kaori Togashi Rikiya Yamashita, Mizuho Nishio. Convolutional neural networks: an overview and application in radiology. 2018. Insights into Imaging.

[50] Pooja Mahajan. Max pooling. 2020. medium.

[51] Jason Brownlee. Overfitting and underfitting with machine learning algorithms. https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning-algorithms/. Accessed: 2021-12-19.

[52] Hackernoon. 7 effective ways to deal with a small dataset. https://hackernoon.com/7-effective-ways-to-deal-with-a-small-dataset-2gyl407s. Accessed: 2021-12-17.

[53] Deep AI. What is an epoch? https://deepai.org/machine-learning-glossary-and-terms/epoch. Accessed: 2021-12-19.

[54] Abhinav Sagar. 5 techniques to prevent overfitting in neural networks. https://www.kdnuggets.com/2019/12/5-techniques-prevent-overfitting-neural-networks.html. Accessed: 2022-06-06.

[55] Python. What is python? executive summary. https://www.python.org/doc/essays/blurb/. Accessed: 2021-12-17.

[56] Scipy tutorial: What is python scipy and how to use it? https://www.edureka.co/blog/scipy-tutorial/. Accessed: 2021-12-18.

[57] scikit-learn. https://scikit-learn.org/stable/. Accessed: 2021-12-18.

[58] Keras. https://keras.io/. Accessed: 2021-12-18.

[59] librosa.load. https://librosa.org/doc/main/generated/librosa.load.html. Accessed: 2021-12-18.

[60] Dinkar Kamat. Audacity. https://audacity.en.softonic.com/?utm_source=SEM&utm_medium=paid&utm_campaign=EN_desktop_RegionA_conversions_DSA&gclid=CjwKCAjwopWSBhB6EiwAjxmqDQlBELO5_ojV-0alSlKewZyqkR98IKG2-RU82eTRofkXikmalGjzlRoCBZMQAvD_BwE. Accessed: 2022-03-31.

[61] Saul McLeod. What does a box plot tell you? https://www.simplypsychology.org/boxplots.html. Accessed: 2022-06-05.

[62] Michael Galarnyk. Understanding boxplots. https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51. Accessed: 2022-06-05.

[63] stomach and intestines sound. stomach and intestines sound. https://www.youtube.com/channel/UCUXDY4AzKt3OM_il3zeo7Yw. Accessed: 2022-06-07.

[64] Sven Magnus Carlsen Konstanze Kölle. Protocol for the pilot study: Analysis of bowel sounds related to meal onset. 2017.

[65] Roland Corporation. Octa-capture. https://www.roland.com/global/products/octa-capture/. Accessed: 2022-06-07.

[66] ADAM Health Solutions. Welcome to colab! https://colab.research.google.com/?utm_source=scs-index. Accessed: 2022-05-11.

[67] Scikit learn. sklearn.preprocessing.standardscaler. https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html. Accessed: 2021-12-17.

[68] Jason Brownlee. Save and load machine learning models in python with scikit-learn. https://machinelearningmastery.com/save-load-machine-learning-models-python-scikit-learn/. Accessed: 2022-05-31.

[69] Jennifer Lapum. Abdomen – auscultation. https://pressbooks.library.ryerson.ca/assessmentnursing/chapter/abdomen-auscultation/. Accessed: 2022-05-30.

[70] Python.org. datetime — basic date and time types. https://docs.python.org/3/library/datetime.html. Accessed: 2022-05-16.

[71] Radu Ranta, Valérie Louis-Dorr, Christian Heinrich, Didier Wolf, and F. Guillemin. Digestive activity evaluation by multichannel abdominal sounds analysis. *IEEE transactions on bio-medical engineering*, 57:1507–19, 02 2010.

[72] Saul McLeod. Pavlov's dogs study and pavlovian conditioning explained. 2022. SimplyPsychology.

[73] Ken DeVault Carol Ann Rinzler. The human digestion process (or, what happens after you eat food). 2019. dummies.

[74] Rober Nowak. Bowel sounds. `https://www.kaggle.com/datasets/robertnowak/bowel-sounds`. Accessed: 2022-05-31.

[75] Leif Erik Andersson Anders Lyngvi Fougner Øyvind Stavdahl Konstanze Kölle, Muhammad Faisal Aftab. Data driven filtering of bowel sounds using multivariate empirical mode decomposition. 2019.

[76] Ryunosuke Sato, Takahiro Emoto, Yuki Gojima, and Masatake Akutagawa. Automatic bowel motility evaluation technique for noncontact sound recordings. *Applied Sciences*, 8(6), 2018.

[77] Analythics Vidhya. Explaining recurrent neural networks. `https://www.bouvet.no/bouvet-deler/explaining-recurrent-neural-networks`. Accessed: 2021-12-14.

A Final implemented detector and further analysis

A.1 Evaluation of the detector

Table 13, 14, 15, 16, 17, 18, 19, 20, 21, 22 and 23 shows the first time point in each minute the detector predicts a bowel sound (at least three consecutive BS frames) from subject 1 recording 3, subject 1 recording 4, subject 2 recording 3, subject 2 recording 4, subject 3 recording 3, subject 6 recording 3, subject 6 recording 4, subject 7 recording 3, subject 7 recording 4 and subject 9 recording 4, respectively. All the times are listened to to identify the whether it is a bowel sound or not. The red ones are false predictions, meaning the time in the recording does not sound like s bowel sound.

0:00:02.91	0:01:03.46	0:04:44.36	0:06:07.66	0:09:12.96	0:10:36.41
0:11:04.51	0:12:25.11	0:13:06.11	0:15:01.66	0:16:00.66	0:17:11.56
0:18:04.21	0:19:06.41	0:20:15.51	0:21:12.91	0:22:01.51	0:23:00.96
0:24:15.41	0:25:05.86	0:26:00.51	0:27:05.01	0:28:01.31	0:29:04.96
0:30:01.61	0:31:16.51	0:32:11.16	0:33:53.51	0:34:35.96	0:35:39.56
0:36:09.51	0:38:46.51	0:39:27.46	0:40:41.06	0:41:15.01	0:42:36.56
0:43:12.06	0:44:51.66	0:45:06.21	0:46:19.56	0:47:04.81	0:48:04.46
0:49:00.41	0:50:05.86	0:51:48.26	0:52:01.11	0:53:08.26	0:54:51.21
0:55:17.81	0:56:37.56	0:57:23.66	0:58:12.26	0:59:06.01	1:00:08.81
1:01:32.16	1:02:16.36	1:03:01.46	1:04:30.81	1:05:12.76	

Table 13: Different time points where the detector has predicted the first bowel sound in each minute for subject 1, recording 3.

0:02:29.91	0:04:20.56	0:05:09.11	0:06:01.41	0:08:26.81	0:11:00.01
0:12:49.86	0:13:31.41	0:15:15.86	0:16:00.21	0:17:06.51	0:18:00.01
0:19:11.41	0:20:04.41	0:21:15.96	0:22:05.71	0:23:14.51	0:24:02.76
0:25:04.81	0:26:04.76	0:27:04.76	0:28:00.06	0:29:07.26	0:30:01.46
0:31:16.51	0:32:27.21	0:33:31.16	0:34:53.96	0:36:39.11	0:38:00.86
0:39:06.51	0:40:55.86	0:41:23.16	0:42:11.16	0:43:14.06	0:45:34.41
0:46:06.11	0:47:20.86	0:48:01.81	0:49:00.41	0:50:17.21	0:51:48.26
0:52:01.11	0:54:51.26	0:55:27.81	0:56:58.41	0:57:36.06	0:59:19.31
1:01:19.81	1:02:02.01	1:03:01.46	1:04:14.96	1:05:09.71	

Table 14: Different time points where the detector has predicted the first bowel sound in each minute for subject 1, recording 4. The red color is wrong prediction.

0:00:16.91	0:01:35.71	0:02:05.91	0:04:43.41	0:05:09.46	0:06:13.76
0:07:43.86	0:08:15.06	0:09:44.21	0:10:04.36	0:11:17.16	0:12:13.96
0:13:29.16	0:14:04.01	0:15:01.16	0:16:10.11	0:17:01.41	0:18:27.31
0:19:00.46	0:20:04.41	0:21:06.86	0:22:07.36	0:23:26.46	0:24:01.31
0:25:03.16	0:26:02.86	0:27:01.56	0:28:01.76	0:29:16.26	0:30:01.26
0:31:03.41	0:32:00.26	0:33:01.81	0:34:11.81	0:35:26.31	0:36:26.11
0:37:09.46	0:38:00.01	0:39:07.81	0:40:25.66	0:41:30.21	0:42:44.76
0:45:05.41	0:47:56.26	0:48:47.81	0:50:02.96	0:51:11.66	0:52:48.91
0:53:03.56	0:54:00.06	0:55:02.56	0:56:04.06	0:57:01.61	0:58:17.31
0:59:02.31	1:00:07.56	1:01:10.91	1:02:05.71	1:04:03.76	1:05:34.01

Table 15: Different time points where the detector has predicted the first bowel sound in each minute for subject 2, recording 3. The red color is wrong prediction.

0:00:16.91	0:05:09.51	0:08:15.0	0:09:09.76	0:10:20.01	0:11:17.16
0:12:15.76	0:13:29.46	0:14:17.56	0:15:13.71	0:16:53.46	0:17:17.71
0:18:16.76	0:19:04.46	0:20:30.91	0:21:34.86	0:22:06.01	0:23:11.31
0:24:01.36	0:25:00.01	0:26:02.86	0:27:01.51	0:28:01.71	0:29:16.26
0:30:00.36	0:31:03.66	0:32:00.21	0:33:01.81	0:34:11.81	0:35:01.71
0:36:26.11	0:37:09.46	0:38:11.56	0:39:03.41	0:40:25.66	0:42:44.76
0:43:13.11	0:45:05.41	0:47:45.31	0:48:03.71	0:49:50.61	0:50:27.21
0:51:11.66	0:52:48.96	0:53:03.56	0:54:00.06	0:55:02.56	0:56:04.26
0:57:06.01	0:58:13.16	0:59:02.36	1:00:07.46	1:01:10.91	1:02:05.71
1:03:45.36	1:04:03.76	1:05:34.01			

Table 16: Different time points where the detector has predicted the first bowel sound in each minute for subject 2, recording 4. The red color is wrong prediction.

0:02:25.46	0:05:15.16	0:07:45.21	0:10:30.46	0:13:09.56	0:14:19.71
0:15:13.31	0:16:09.81	0:17:18.11	0:18:30.66	0:19:08.36	0:20:21.21
0:21:26.46	0:22:01.06	0:23:03.21	0:24:00.16	0:25:04.36	0:26:01.61
0:27:46.31	0:28:02.76	0:29:11.71	0:30:03.06	0:31:04.11	0:32:03.36
0:33:35.21	0:34:12.11	0:35:18.46	0:36:13.71	0:37:11.26	0:38:17.06
0:39:57.66	0:40:25.61	0:42:12.16	0:45:18.41	0:46:29.31	0:47:39.61
0:48:05.21	0:49:44.41	0:50:43.46	0:55:52.51	0:56:09.81	1:03:34.91

Table 17: Different time points where the detector has predicted the first bowel sound in each minute for subject 3, recording 3. The red color is wrong prediction.

0:00:02.66	0:02:25.46	0:03:02.01	0:04:01.41	0:05:15.01	0:06:01.76
0:07:45.21	0:08:00.66	0:10:47.01	0:12:10.66	0:13:08.81	0:14:21.11
0:15:00.86	0:16:09.26	0:17:11.51	0:18:05.61	0:19:07.06	0:20:04.56
0:21:13.16	0:22:00.71	0:23:04.06	0:24:00.21	0:25:04.61	0:26:01.61
0:27:47.41	0:28:02.76	0:29:11.71	0:30:03.51	0:31:04.11	0:32:03.36
0:33:35.21	0:34:12.11	0:35:18.46	0:36:13.71	0:37:11.26	0:38:17.06
0:39:48.56	0:40:25.71	0:41:52.91	0:43:45.71	0:44:05.81	0:45:18.36
0:46:06.51	0:47:39.56	0:47:40.06	0:48:05.16	0:49:44.41	0:50:43.46
0:56:09.81	0:57:46.91	0:59:12.61	1:00:45.16		

Table 18: Different time points where the detector has predicted the first bowel sound in each minute for subject 3, recording 4. The red color is wrong prediction.

0:00:02.76	0:03:38.61	0:05:40.01	0:09:26.56	0:10:01.41	0:11:00.91
0:12:44.81	0:13:03.76	0:14:09.26	0:15:04.46	0:16:25.36	0:17:31.66
0:18:16.61	0:19:00.11	0:20:07.96	0:21:17.21	0:22:03.26	0:23:00.76
0:24:01.66	0:25:01.91	0:26:15.81	0:27:02.66	0:28:03.06	0:29:09.41
0:30:55.46	0:31:33.86	0:33:27.16	0:34:17.41	0:35:33.31	0:37:22.31
0:40:13.11	0:41:09.71	0:47:11.46			

Table 19: Different time points where the detector has predicted the first bowel sound in each minute for subject 6, recording 3. The red color is wrong prediction.

0:00:00.31	0:01:29.41	0:02:02.81	0:03:13.41	0:06:30.16	0:07:25.56
0:08:06.26	0:09:20.56	0:10:01.41	0:11:00.96	0:12:03.11	0:13:01.71
0:14:03.16	0:15:04.46	0:16:00.46	0:17:00.31	0:18:07.91	0:19:00.11
0:20:02.01	0:21:03.81	0:22:03.36	0:23:00.26	0:24:00.11	0:25:00.81
0:26:06.96	0:27:00.01	0:28:00.11	0:29:01.26	0:30:01.26	0:31:00.81
0:32:00.71	0:33:11.96	0:34:01.61	0:35:10.21	0:36:39.16	0:37:08.86
0:38:17.86	0:39:12.26	0:40:12.31	0:41:00.26	0:42:10.61	0:43:12.11
0:44:12.46	0:45:15.71	0:51:00.06	0:53:45.01		

Table 20: Different time points where the detector has predicted the first bowel sound in each minute for subject 6, recording 4. The red color is wrong prediction.

0:00:14.41	0:01:21.86	0:02:02.76	0:03:29.56	0:04:00.66	0:05:11.11
0:06:12.36	0:07:45.51	0:09:01.01	0:12:25.06	0:13:30.21	0:14:01.46
0:15:27.86	0:16:17.71	0:17:09.01	0:18:00.51	0:19:37.86	0:20:13.21
0:21:23.01	0:22:00.86	0:23:00.01	0:24:01.46	0:25:03.51	0:26:03.21
0:28:08.46	0:29:00.46	0:30:56.71	0:32:23.06	0:40:30.81	0:43:39.86
0:46:44.41	0:48:55.46	0:49:24.81	0:50:41.56	0:54:12.96	0:56:32.21
0:57:30.41	1:00:25.81	1:01:11.66	1:03:06.21	1:04:15.71	1:06:19.86
1:07:02.76	1:09:00.61	1:11:36.56	1:13:18.76		

Table 21: Different time points where the detector has predicted the first bowel sound in each minute for subject 7, recording 3. The red color is wrong prediction.

0:00:22.31	0:01:28.56	0:02:36.01	0:03:29.56	0:04:00.66	0:05:45.36
0:06:23.01	0:07:56.31	0:09:00.96	0:10:52.31	0:13:28.56	0:14:03.21
0:15:18.16	0:16:04.56	0:17:02.26	0:18:08.11	0:19:01.56	0:20:35.91
0:21:04.76	0:22:23.56	0:23:01.46	0:24:00.11	0:25:06.61	0:26:00.31
0:28:07.31	0:30:57.86	0:31:00.26	0:32:08.86	0:33:41.51	0:34:23.16
0:35:59.31	0:36:31.41	0:37:04.36	0:38:08.76	0:39:30.31	0:40:05.86
0:41:27.71	0:42:41.91	0:43:50.46	0:44:07.51	0:45:00.31	0:48:55.26
0:49:23.11	0:50:12.11	0:51:01.81	0:52:35.11	0:53:22.56	0:54:07.61
0:55:16.46	0:56:50.41	0:57:05.61	0:58:30.26	1:00:26.56	1:02:03.61
1:03:05.81	1:04:13.21	01:07:02.81	1:10:14.81		

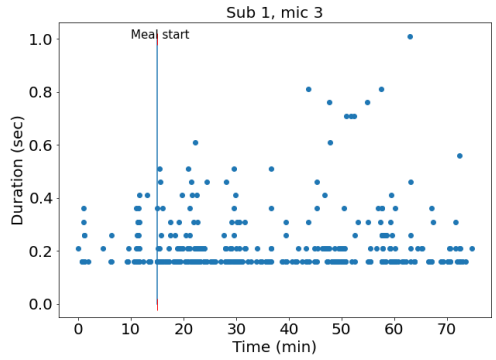
Table 22: Different time points where the detector has predicted the first bowel sound in each minute for subject 7, recording 4. The red color is wrong prediction.

0:00:01.16	0:01:17.26	0:02:25.21	0:03:14.51	0:04:22.31	0:05:31.11
0:06:13.21	0:07:20.71	0:08:10.01	0:09:16.56	0:10:06.61	0:11:19.76
0:12:25.56	0:13:03.16	0:14:05.86	0:15:00.86	0:16:04.81	0:17:03.21
0:18:00.66	0:19:02.96	0:20:00.61	0:21:17.51	0:22:26.86	0:23:00.81
0:24:00.31	0:25:02.21	0:26:00.86	0:27:17.26	0:28:06.51	0:29:03.76
0:30:04.16	0:31:18.81	0:32:16.11	0:33:20.36	0:34:00.21	0:35:00.76
0:36:04.41	0:37:23.36	0:38:01.86	0:39:05.71	0:40:07.21	0:41:01.11
0:42:08.06	0:43:00.06	0:44:00.21	0:45:00.66		

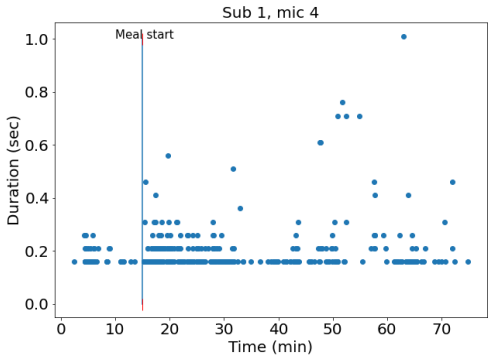
Table 23: Different time points where the detector has predicted the first bowel sound in each minute for subject 9, recording 4. The red color is wrong prediction.

A.2 Duration of each detected bowel sound

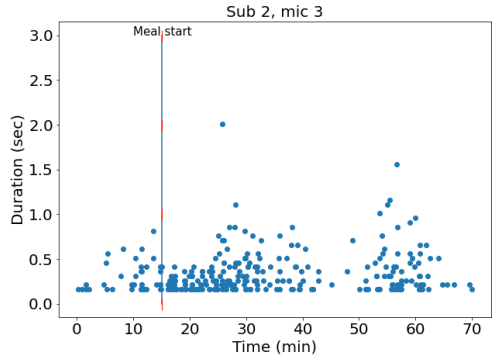
Figure A.1 and A.2 shows the durations of each detected bowel sound when the subjects are following the protocol described in Section 3.2.2.



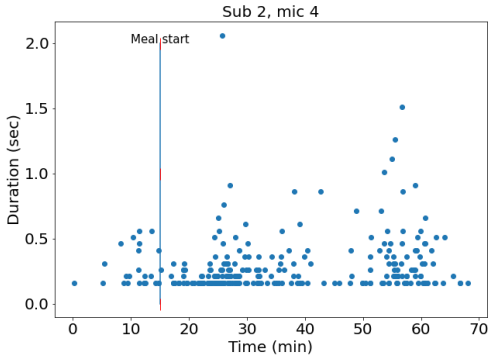
(a) Subject 1, microphone 3.



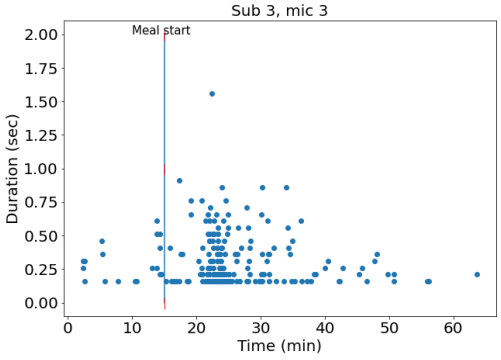
(b) Subject 1, microphone 4.



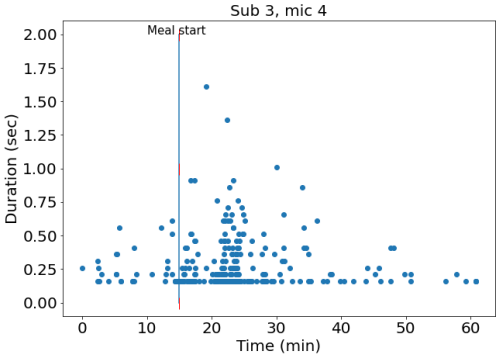
(c) Subject 2, microphone 3.



(d) Subject 2, microphone 4.

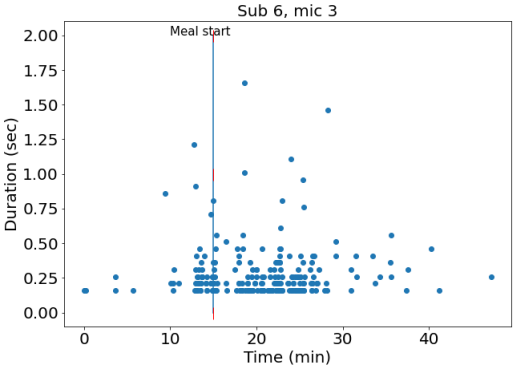


(e) Subject 3, microphone 3.

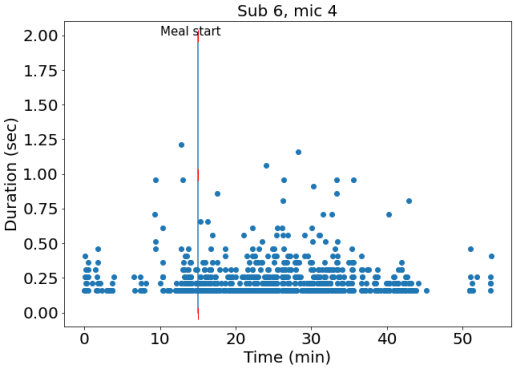


(f) Subject 3, microphone 4.

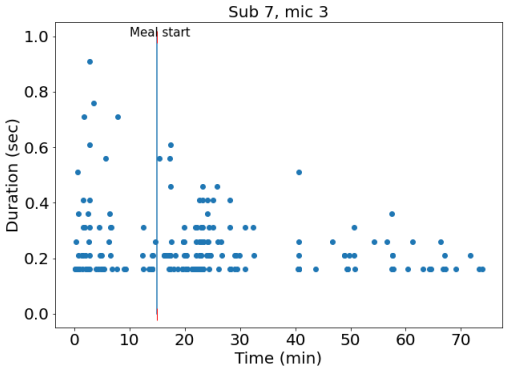
Figure A.1: Duration of each detected BS for subjects 1, 2, 3.



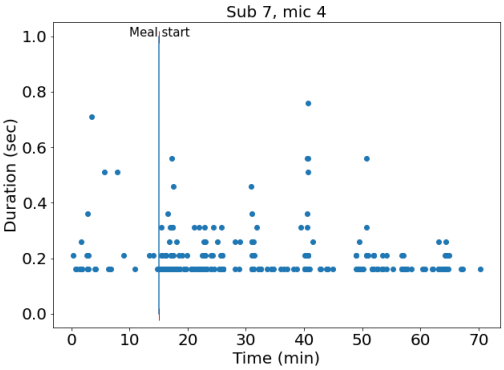
(a) Subject 6, microphone 3.



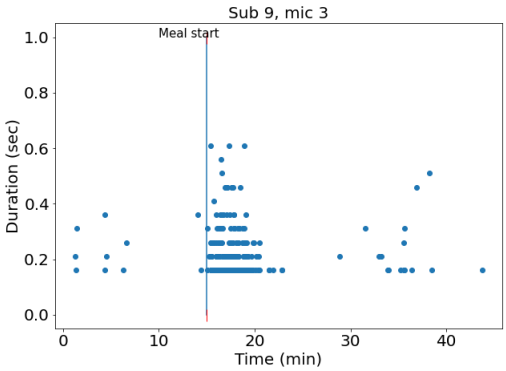
(b) Subject 6, microphone 4.



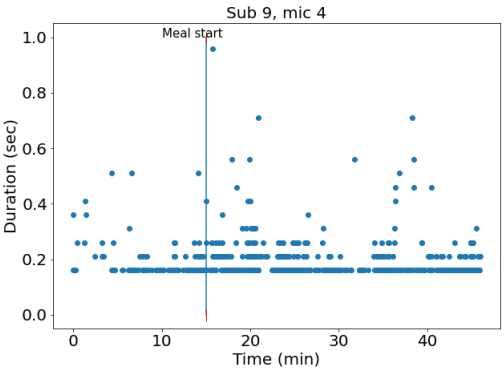
(c) Subject 7, microphone 3.



(d) Subject 7, microphone 4.



(e) Subject 9, microphone 3.

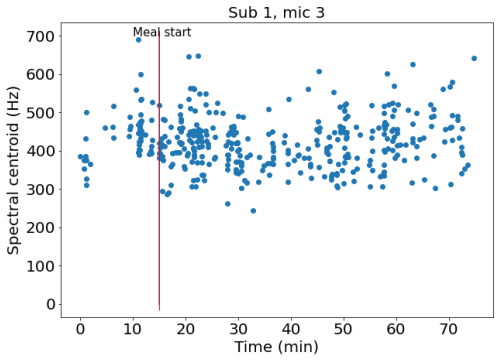


(f) Subject 9, microphone 4.

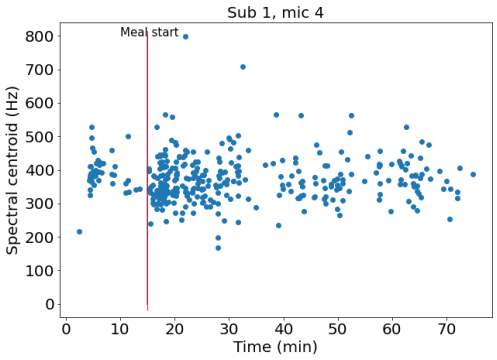
Figure A.2: Duration of each detected bowel sound for subjects 6, 7, 9. One bowel sound is identified of at least three consecutive BS frames.

A.3 SC in each detected bowel sound

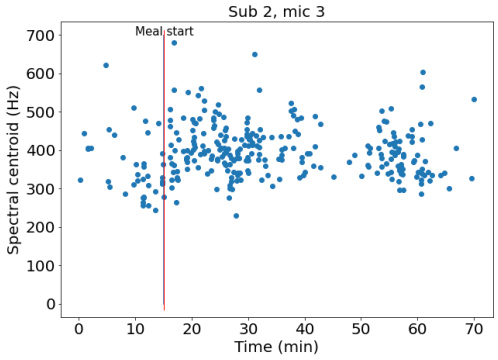
Figure A.3 and A.4 shows the SC of each detected bowel sound when the subjects are following the protocol described in Section 3.2.2.



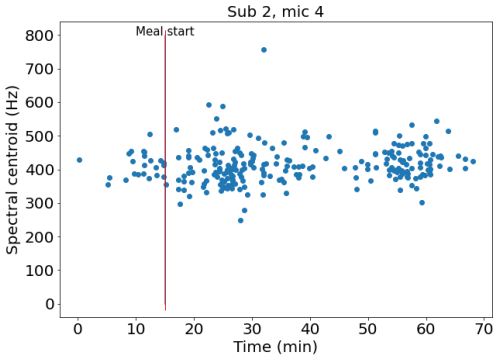
(a) Subject 1, microphone 3.



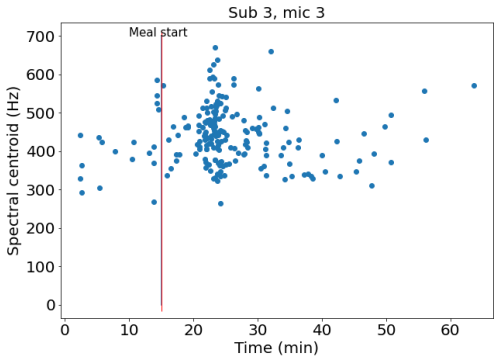
(b) Subject 1, microphone 4.



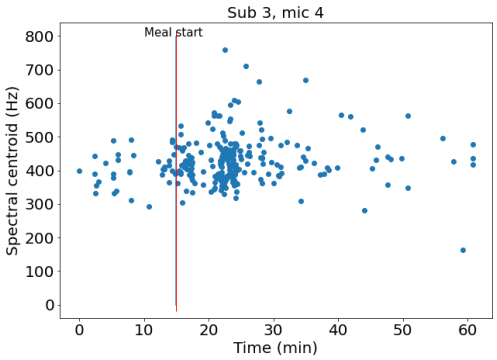
(c) Subject 2, microphone 3.



(d) Subject 2, microphone 4.

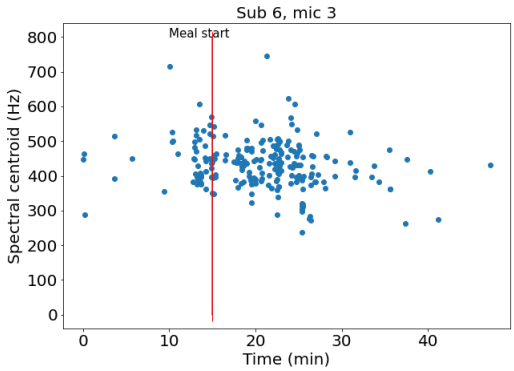


(e) Subject 3, microphone 3.

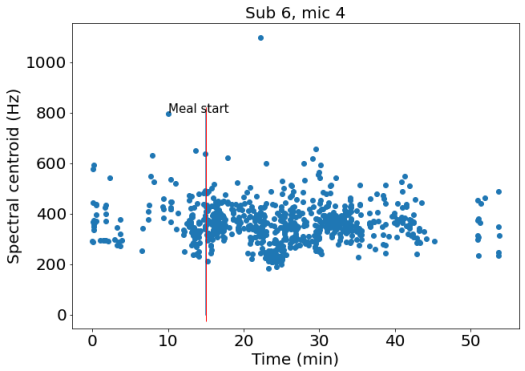


(f) Subject 3, microphone 4.

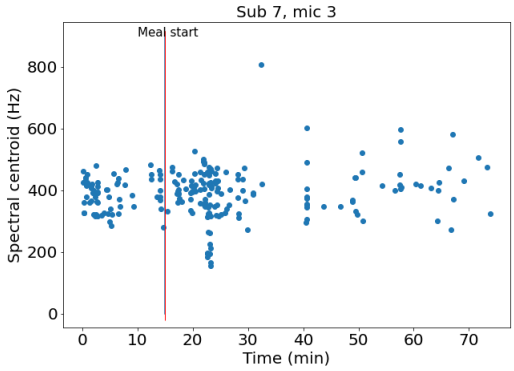
Figure A.3: SC of detected bowel sound for different subjects and microphones.



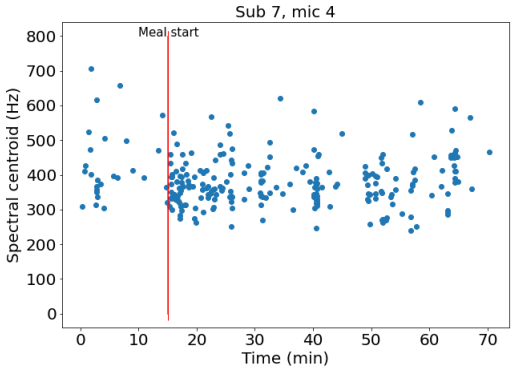
(a) Subject 6, microphone 3.



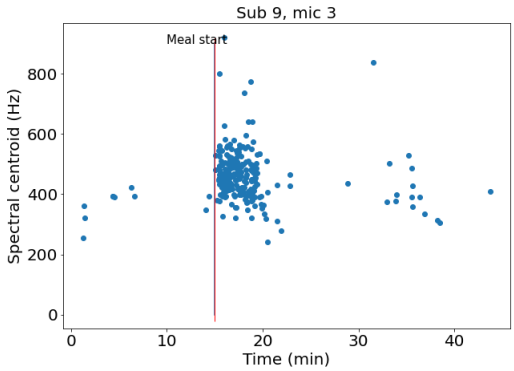
(b) Subject 6, microphone 4.



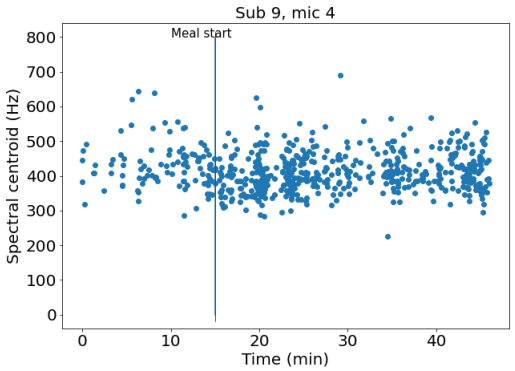
(c) Subject 7, microphone 3.



(d) Subject 7, microphone 4.



(e) Subject 9, microphone 3.

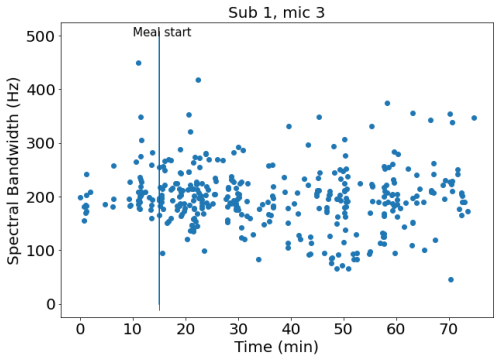


(f) Subject 9, microphone 4.

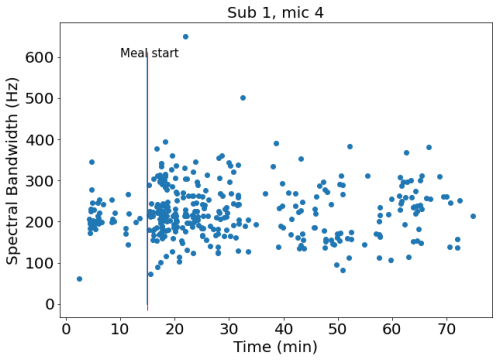
Figure A.4: SC of detected bowel sound for different subjects and microphones.

A.4 SBW in each detected bowel sound

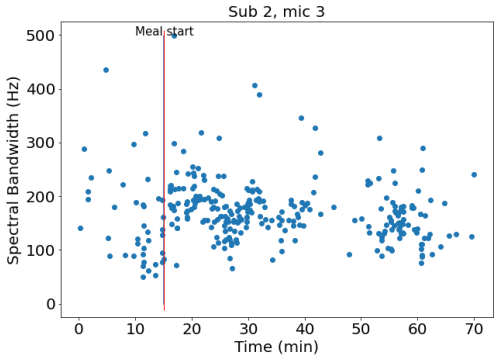
Figure A.5 and A.6 shows the SC of each detected bowel sound when the subjects are following the protocol described in Section 3.2.2.



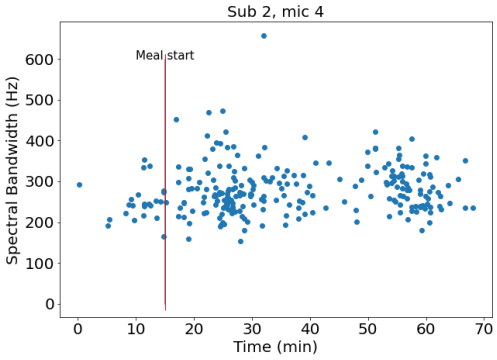
(a) Subject 1, microphone 3.



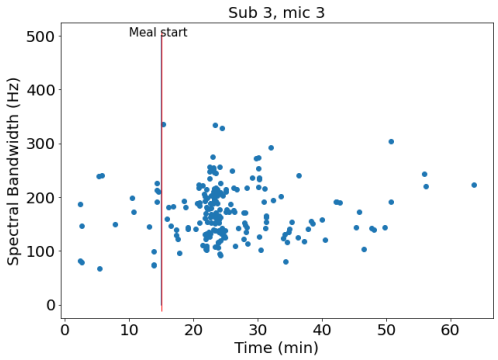
(b) Subject 1, microphone 4.



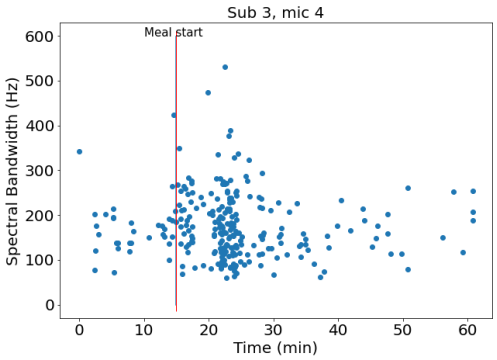
(c) Subject 2, microphone 3.



(d) Subject 2, microphone 4.

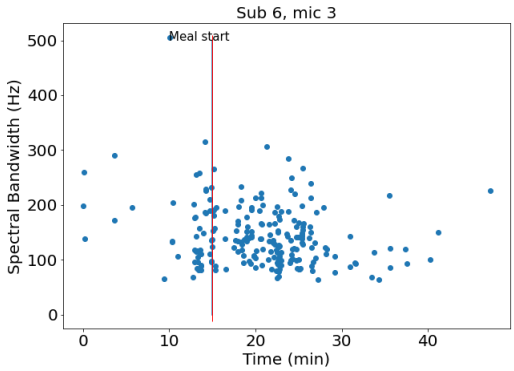


(e) Subject 3, microphone 3.

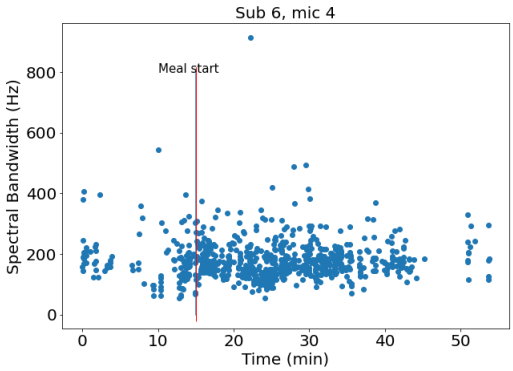


(f) Subject 3, microphone 4.

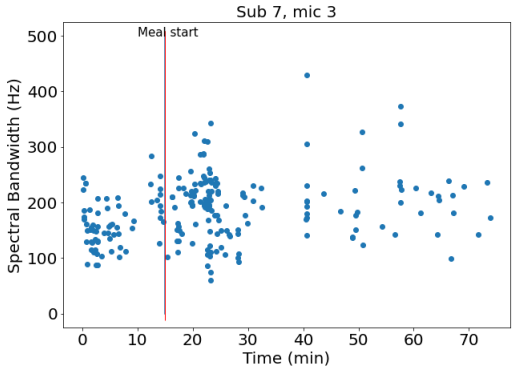
Figure A.5: SBW of detected bowel sound for different subjects and microphones.



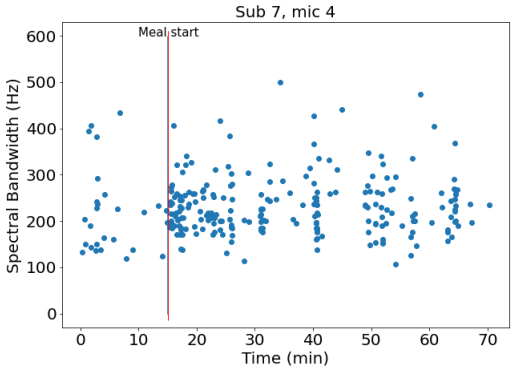
(a) Subject 6, microphone 3.



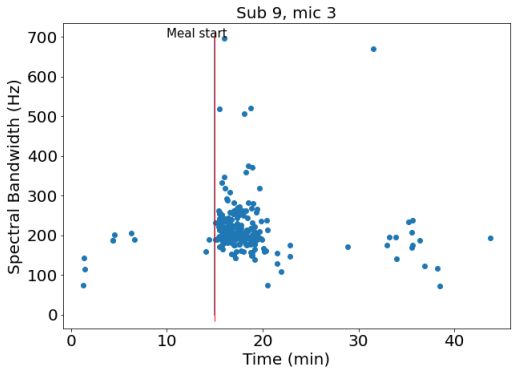
(b) Subject 6, microphone 4.



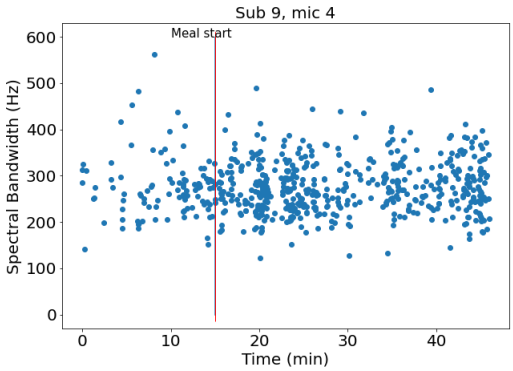
(c) Subject 7, microphone 3.



(d) Subject 7, microphone 4.



(e) Subject 9, microphone 3.

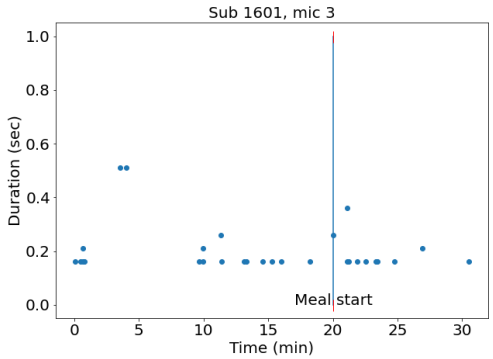


(f) Subject 9, microphone 4.

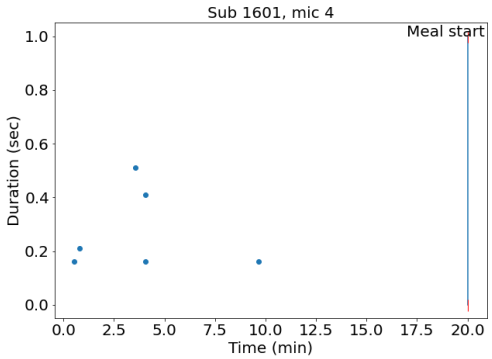
Figure A.6: SBW of detected bowel sound for different subjects and microphones.

B Distinguish between meals

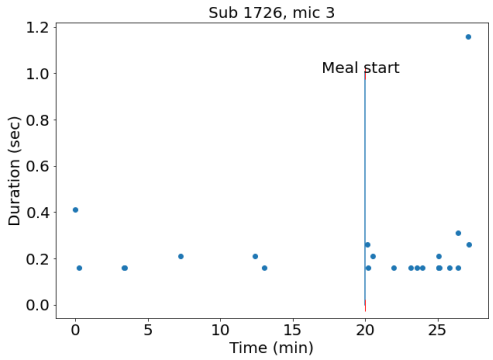
Figure B.1 shows the duration of each detected bowel sound when the subjects are following the protocol described in Section 3.2.3.



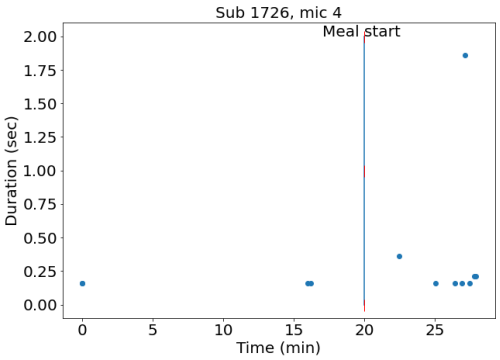
(a) Subject 1601, microphone 3.



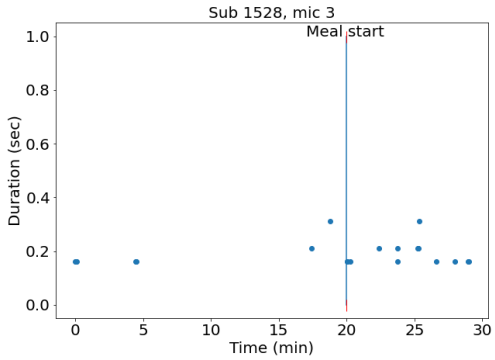
(b) Subject 1601, microphone 4.



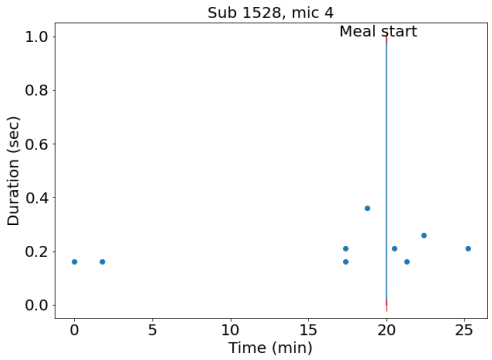
(c) Subject 1726, microphone 3.



(d) Subject 1726, microphone 4.



(e) Subject 1528, microphone 3.

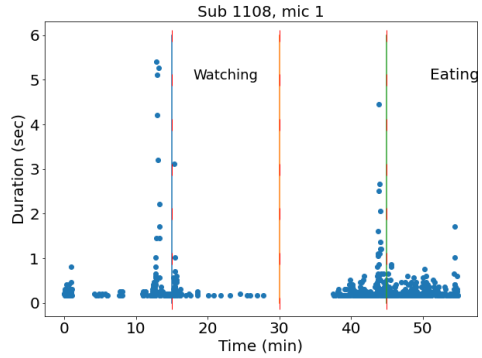


(f) Subject 1528, microphone 4.

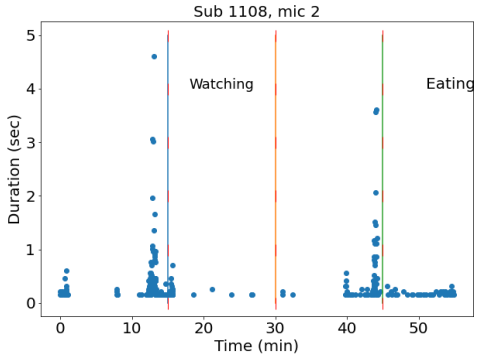
Figure B.1: Duration of each detected bowel sound.

C Meal simulation

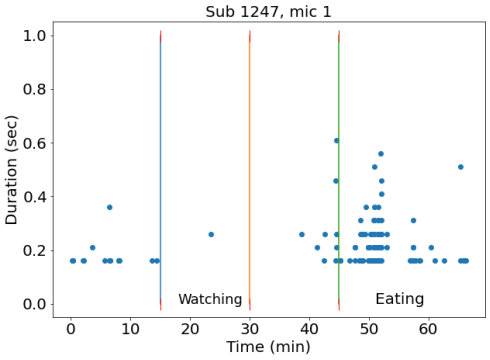
Figure C.1, C.2, and C.3 shows the duration of each detected bowel sound when the subjects are following the protocol described in Section 3.2.4.



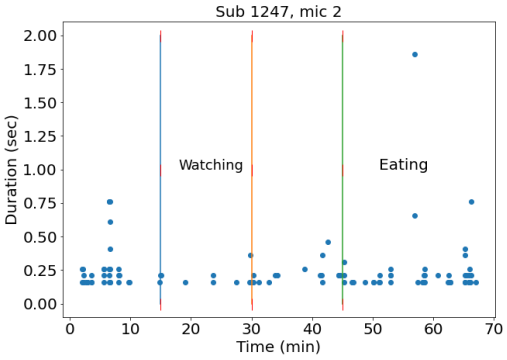
(a) Subject 1108, microphone 1.



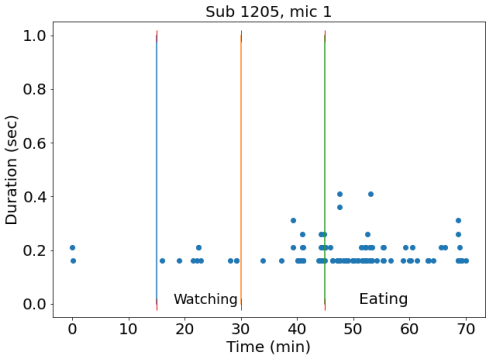
(b) Subject 1108, microphone 2.



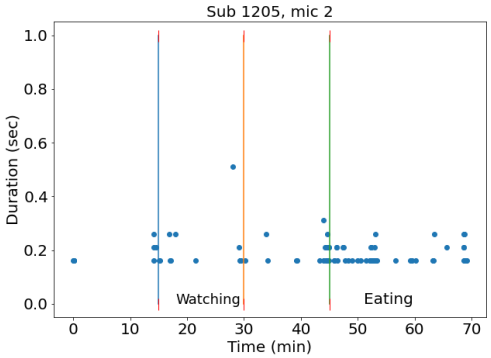
(c) Subject 1247, microphone 1.



(d) Subject 1247, microphone 2.

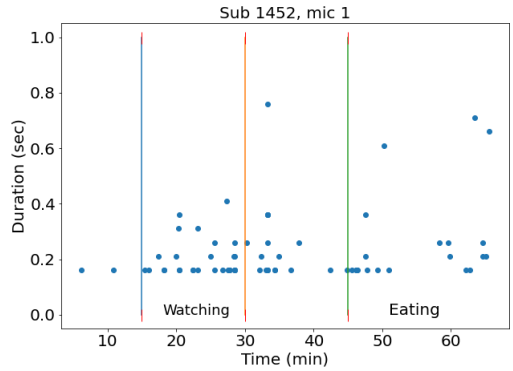


(e) Subject 1205, microphone 1.

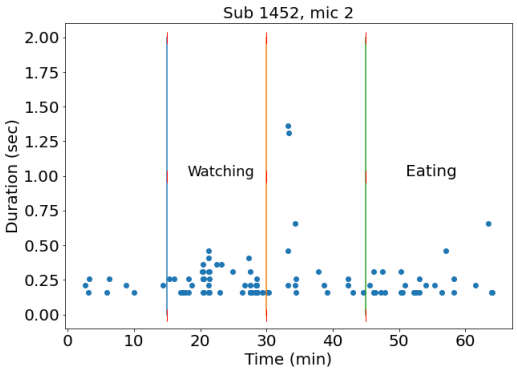


(f) Subject 1205, microphone 2.

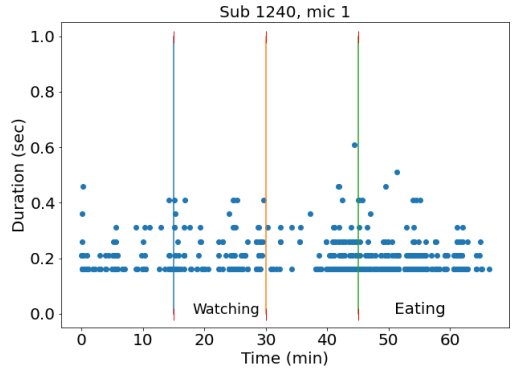
Figure C.1: Duration of each detected bowel sound.



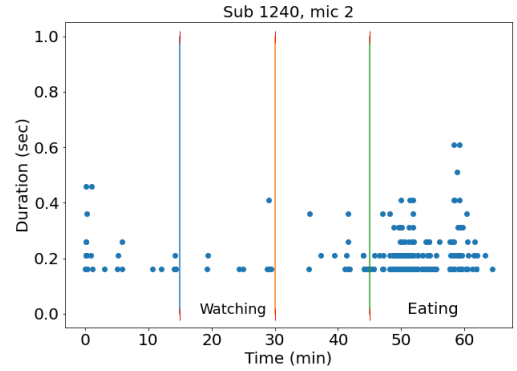
(a) Subject 1452, microphone 1.



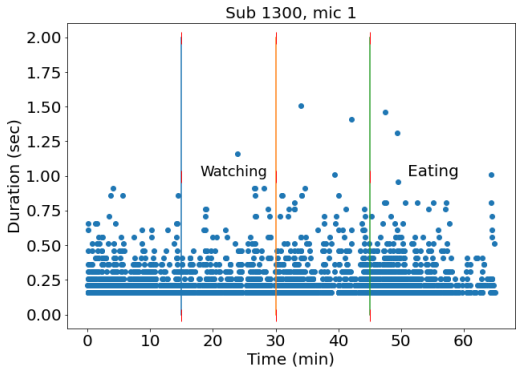
(b) Subject 1452, microphone 2.



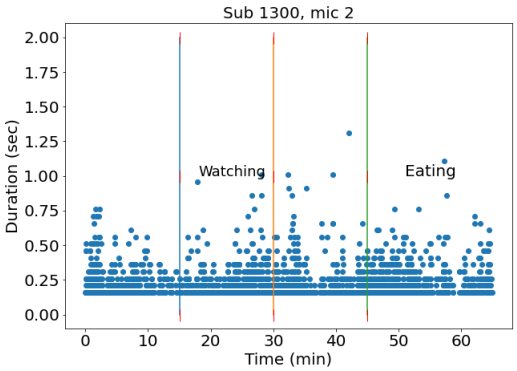
(c) Subject 1240, microphone 1.



(d) Subject 1240, microphone 2.

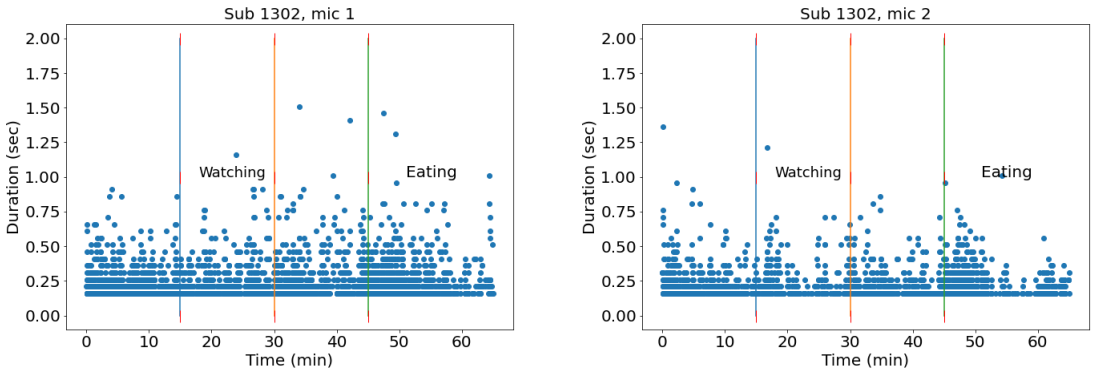


(e) Subject 1300, microphone 1.



(f) Subject 1300, microphone 2.

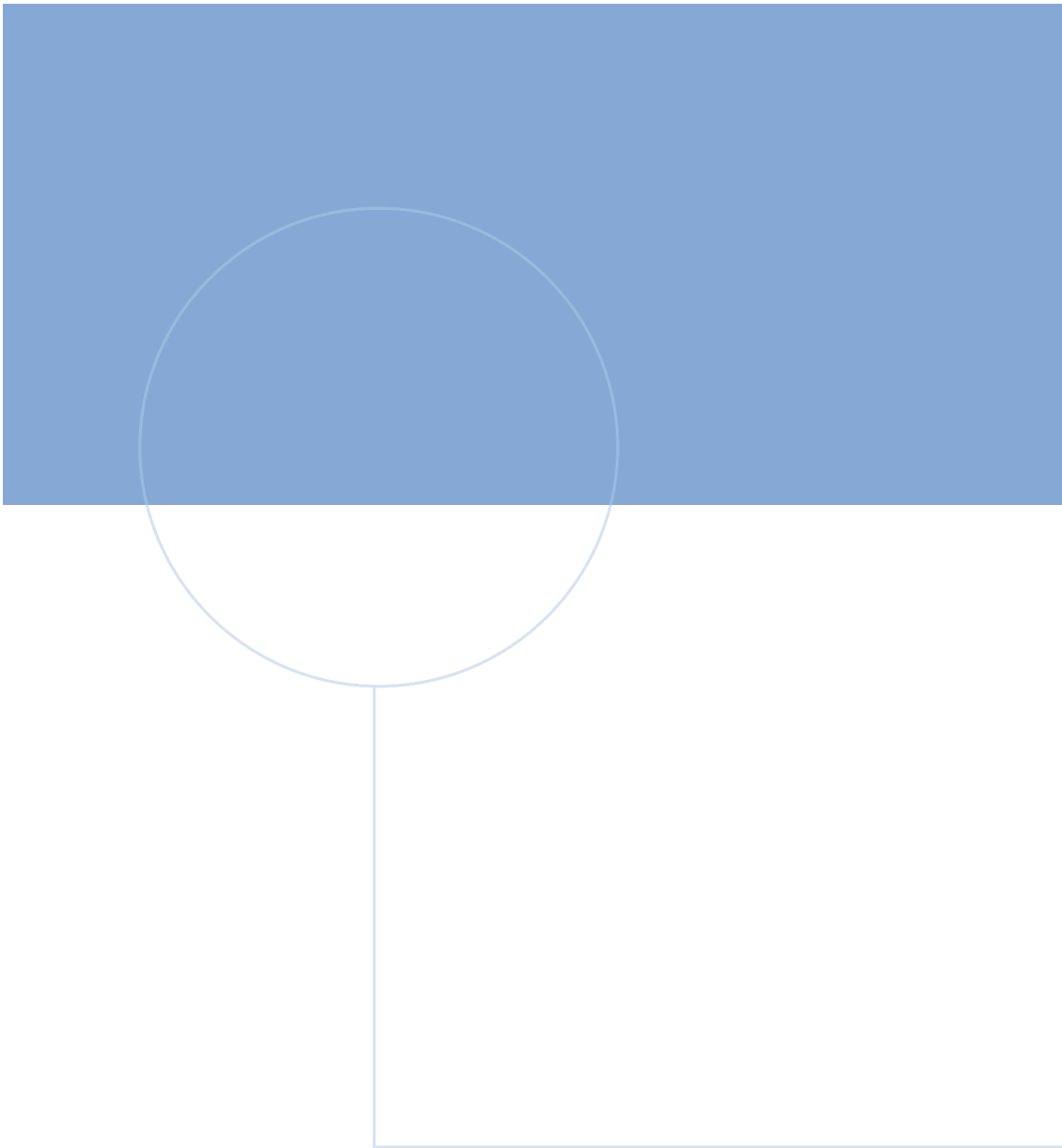
Figure C.2: Duration of each detected bowel sound.



(a) Subject 1302, microphone 1.

(b) Subject 1302, microphone 2.

Figure C.3: Duration of each detected bowel sound.



NTNU

Norwegian University of
Science and Technology