*Term project*
*TFE4595 Electronic Systems Design and Innovation*

# *Combining bowel and swallowing sounds for improved meal detection*

*Ahmed Isifan*

*Advisors:*
*Salman Ijaz Siddiqui*
*Anders Lyngvi Fougner*

*Supervisor:*
*Dag Roar Hjelme*

*Trondheim, December 19th,2021*

NTNU
Norwegian University of
Science and Technology

Faculty of Information Technology, and Electrical Engineering
DEPARTMENT OF ELECTRONIC SYSTEMS

# Abstract

Early meal detection can help improve the performance of continuous glucose monitoring systems (CGM). Modern diabetes solutions such as artificial pancreas rely on CGM systems to monitor the glucose level in the blood and based on the sugar level in the blood, insulin is dosed. These CGM systems are however not ideal and are subject to time delays of 30-40 min from meal onset until the meal is detected. Earlier studies have shown promising results in meal detection by using recorded sounds of bowel movement during and after a meal onset. Such a method could be used to improve CGM systems by reducing time delays, however, the method suffered from a high number of false positives (FP). In this study, both bowel and swallowing sound recordings were used to reduce the number of FP's. Results showed that both the number of FP's and the meal detection time were reduced. The average meal detection time for such a system is 1-2 min.

For this project, a total of 10 meal recordings were obtained, where each recording gathered data from four microphones simultaneously. Each one of the four microphones gathered data from a specific location, two of the microphones were placed at the right and left side of the lower part of the abdomen, one was placed right above the collar bone, and the last one was placed right below the right ear. The microphones captured bowel, swallowing, and chewing sounds, and these were used for training a support vector machine classifier using frequency spectrum features.

# Preface

This thesis is submitted as a part of the term project in the subject TFE4595 Electronic Systems Design and Innovation in the Department of Information Technology and Electrical Engineering. The project was provided by the Artificial Pancreas Trondheim (APT) group for the autumn semester.

First of all, I would like to thank my advisor Salman Ijaz Siddiqui, who provided me with guidance and support throughout the whole research process. I would also like to thank Anders lyngvi Fougner for his great support. A special thanks must also go to my friends, Davis klavins and Erlend Løland Gundersen, and everyone who motivated and supported me. Most importantly I would also like to thank my family who was always there when I needed support.

# Contents

# List of Figures

# List of Tables

# Nomenclature

| | |
|---|---|
| $CGM$ | Continues glucose monitoring |
| $FP$ | False positive |
| $FPR$ | False positive rate |
| $PSD$ | Power spectral density |
| $TN$ | True negative |
| $TP$ | True positive |
| $TPR$ | True positive rate |
| ADC | Analog to digital converter |
| AP | Artificial pancreas |
| LOOCV | leave-one-out-cross-validation |
| MI | Mutual information |
| ML | Machine learning |
| MSE | Mean squared error |
| RBF | Radial basis function |
| SNR | Signal-to-noise ratio |
| SVM | Support vector machines |

# 1 Introduction

## 1.1 Background

### 1.1.1 Diabetes

Diabetes is a metabolic disease that is caused by the reduction or absence of the production of the hormone insulin. Most of the food eaten is broken down into sugar (glucose) and released into the bloodstream. When the blood sugar goes up, the body signals the pancreas to release insulin, which acts like a key to let the blood sugar into the body cells for use as energy [2].

Diabetes patients have inadequate or no production of insulin, which may result in high blood glucose levels, also known as hyperglycemia. Symptoms of hyperglycemia develop slowly over several days or weeks. The longer blood sugar levels stay high, the more serious the symptoms become. Hyperglycemia may lead to shortness of breath, weakness, confusion, abdominal pain, and also coma [3].

There are three main types of diabetes, which are type 1 diabetes, type 2 diabetes, and gastrointestinal diabetes. Type 1 diabetes is caused by an autoimmune reaction, where the body attacks itself by mistake, this stops the body's insulin production. Approximately $5 - 10\%$ of diabetes patients have diabetes type 1 [2]. While for type 2 diabetes, the body produces insulin, but however it does not use it well, and it can not keep the blood sugar at normal levels. About $90 - 95\%$ of people who have diabetes have type 2 diabetes [2]. Gastrointestinal diabetes is due to a high blood glucose level that develops during pregnancy and usually disappears after [4].

### 1.1.2 Diabetes treatment

Type 2 diabetes can be prevented or even delayed with a healthier lifestyle change, such as being more active and eating healthier. That is however not the case for type 1 diabetes patients, as they need to take insulin on a daily basis to survive [2].

The insulin dosage for type 1 diabetes patients depends on the measured glucose level in the blood, thus a finger stick blood test is often required before administrating the insulin dose. The patient can take an insulin shot using a syringe, an insulin pen, or an insulin pump. An insulin pump is a small machine

that delivers steady insulin doses throughout the day, although patients might require to take an extra dose of insulin at mealtime through the pump.

More advanced diabetes treatments such as artificial pancreas (AP) relies on continues glucose monitoring (CGM) systems. These types of systems monitor the glucose level in the bloodstream periodically using sensor technologies. However, a major problem with this type of technology is that it needs calibration from time to time, thus twice a day the patient is required to test a drop of blood on a standard glucose meter[5].

## 1.2 Motivation

A major problem with these advanced diabetes treatments is that the patient's involvement in the therapy is still vital, the treatment affects the daily life of the patient. Another thing about CGM systems is that they are subject to time delays of 30-40 minutes from meal onset until the meal is detected [6]. The CGM systems are subject to time delays and slow dynamics due to the latency of interstitial fluid. As a consequence, patients are required to announce the meal intake, this is required by clinically tested systems for glucose control.

An automatic meal detection system could help mitigate some of the problems of CGM systems by reducing the time required for the administration of insulin, and it may also cut down the need for meal announcements. Such an early meal detection system using bowel sound recordings was attempted by Konstanze Kölle [6]. His study showed promising results, as the average meal detection time was reduced to 10 minutes. Even though the study showed promising results, it was still lacking as the accuracy and recall of the system were low, and the system produced a lot of false positives. Other systems considered the possibility of detecting a meal intake based on swallowing sounds, such systems had high accuracy, as high as 75% [7] [8] [9][10].

Both types of systems showed significant results, however, as a standalone system for meal detection these systems are unreliable since any false positive can trigger a wrong insulin dose, and that can result in transient and serious hyperglycemia [11]. A system that combines both bowel and swallowing sounds could help increase the reliability of meal detection. If such an automatic meal detection system is possible, then when combined with a CGM system, the delay between-meal onset and detection could be drastically reduced, improving the overall performance of CGM systems.

## 1.3 Objective

This thesis takes the work done by Konstanze a step further by relying on both bowel and swallowing sound recordings for building a meal detection system. The same data acquisition protocol used by Konstanze will be followed in this thesis, for both swallowing and bowel sound recordings. All the recordings used in this project were captured by four channels (microphones), the microphones were used to capture bowel, swallowing, and chewing sounds. The chewing sounds were recorded but not used, because another master student needed them for his project.

## 1.4 Thesis outline

This thesis will be organized as such. First, some important concepts for understanding the different parts of the system will be explained in detail in the theory section. Then, in the subsequent two sections, the protocol for data acquisition as well as the method for implementing the system will be introduced. These sections will also describe how the classifier is trained and validated, and how the performance of the system is assessed. Followed by that, a section will deal with presenting the results found in this project. These results will then be discussed, and the most important findings in this project will be summarized. Lastly, the feasibility of such a system would be discussed in addition to future work.

# 2 Theory

## 2.1 Introduction

The diagram in Figure 1 shows the general steps for data analysis in any machine learning system, and these steps yield also for this project. During this section, these steps will be presented in more detail.

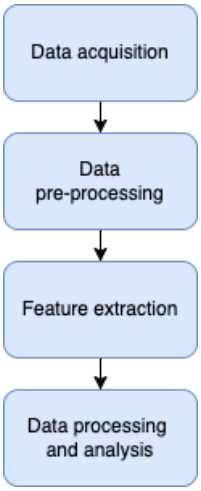


**Figure 1:** General steps of bowel and swallowing sound analysis.

In the first step, the data is acquired for this project, data is collected through four microphones that record simultaneously. These microphones measure the variation in air pressure and convert the variations into an electrical signal via an analog-to-digital converter (ADC).

In the second step, the recordings will be filtered to remove unwanted information in the signal. This includes high-frequency content where there is no information of interest. After that, the signal is then normalized before the feature calculation step.

In the third step, the filtered and normalized data are used to extract the most important and nonredundant information. This data is then used in the fourth step, where it's processed and analyzed.

## 2.2 Signal processing

### 2.2.1 Quantization

The number of bits of information per sample is the bit depth. When quantization is performed, the bit depth of the signal is reduced, this process leads to constraining the large set of values to a smaller one. Quantization leads to rounding of the values in the original signal, which in turn 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 peaks [12]. The frequency response of the signal is however unaffected by this operation since it's only constrained by the sampling rate of the signal.

### 2.2.2 Decimation

Downsampling is the operation where the sampling rate is reduced by keeping every M'th sample, where M is the downsampling factor. Lowpass filtering is not involved in the operation. Decimation is the operation where the sampling rate is reduced, but before that lowpass filtering is applied to avoid aliasing. Aliasing is caused by downsampling with a sampling rate below the Nyquist rate. Decimation leads to a reduction in the power of the signal since the high-frequency content of the signal is attenuated [13].

### 2.2.3 Normalization

The goal of normalization is to use a common scale for the data without distorting the differences in the range of values. This helps with increasing the training speed of the classifier as the features used are in a similar range and values [14]. Linear normalization was used for this project, where the range of the values after normalization was between 0 and 1. Normalization is given by

$$x' = \frac{x - x_{min}}{x_{max} - xmin} \tag{1}$$

where x' is the normalized signal, x is the signal amplitude at a given time, and $x_{min}$, $x_{max}$ is the minimum and maximum amplitude of the signal respectively.

### 2.2.4 Median filtering

Median is a term describing the element in the middle of a sorted data set. Median filtering of length N will iterate through the data using a window of length N, where for each iteration the elements in the window are sorted and only the median is kept. This operation leads to a smoothing of the data, as outliers are removed.

## 2.3 Features

Training a classifier can be an intensive operation, as it requires a lot of processing power. Training requires a large amount of data, and usually, there is a lot of redundancy in the data, this could be exploited. By removing the redundant data, not only is the processing time reduced but also the prediction ability of the classifier is improved. The set of selected data for training the classifier is known as features.

### 2.3.1 Feature extraction

The frequency spectrum shows the amplitude of the frequency content of the signal. For this project, the features extracted are power frequency-based, for both bowel and swallowing sounds.

Earlier studies confirmed that most of the signal power spectrum density for bowel sounds is concentrated below 1000 Hz [15][16]. The largest power spectrum density of abdominal sounds is located between 100-500 Hz [15]. This was also confirmed by Konstanze Kölle, who noted that during a meal onset most of the power increase is for frequencies below 1000 Hz [6].

Regarding the frequency spectrum of swallowing sounds, the active frequency content of the signal during a meal intake seems to be in the region 400-1000 Hz [17]. However, information about the meal content, such as the type of food or liquid is located in the higher frequency range. This information is located in the frequency range up to 3600 Hz [8].

For this project, as it was not important to detect the type of food that is ingested, only meal onset, a frequency range of 0-1000 Hz sufficed as most meal onset information were located at these frequencies for both swallowing and bowel sounds. To ensure that no information is lost, and allow for the possibility of accessing additional information, later on, a frequency range of 0-2000

Hz was used. This allowed also for more relaxed filter constraints during decimation (lowpass filtering).

### 2.3.2 Feature calculation

Using the frequency spectrum of the signal, features such as power spectral density (PSD) could be calculated. PSD shows the energy of the signal as a function of frequency. The PSD is obtained using the Fourier Transform and is given by

$$\hat{P}_s(f) = \frac{\Delta t}{N} |\sum_{n=0}^{N-1} x_{s,n} e^{-i2\pi fn}|^2 \tag{2}$$

Where $\Delta t$ is the sampling interval and N is the number of samples in time segment s [18].

## 2.4 Machine learning

Machine learning (ML) is a subfield of artificial intelligence that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. ML algorithms use historical data as input to predict new output values [19]. There are two main types of ML, which are supervised learning and unsupervised learning. This project uses supervised learning.

### 2.4.1 Supervised learning

In supervised ML, the algorithm is supplied with training data that is labeled. The algorithm trains using the labeled data and then try to learn how to map the new data, $X_{new}$, to an output, $Y_{predicted}$, by using the known input and output pair (X,Y) [20]. The performance of the mapping is measured using a loss function, the loss function used for this project is the mean squared error (MSE), and is given by (3), where N is the number of training samples.

$$L(Y, Y_{predicted}) = \frac{1}{N} \sum_{i=0}^{N} (Y - Y_{predicted})^2 \tag{3}$$

Supervised learning can be separated into two types of problems, which are classification and regression. This project focuses on classification problems. In

classification problems, an algorithm is used to assign data to a specific class. It recognizes specific entities within the data set and attempts to draw some conclusions on how those entities should be labeled or defined [20].

### 2.4.2 Training, validation, and test data

All ML systems require data to work, as predictions depend on the data fed into the system. The most common three types of data used to build a ML system are training, validation, and testing. Training data is used to train the classifier and estimate the parameters of the model. Validation data is used to validate and tune the hyperparameters of the model by providing an unbiased data set. The test data is used to test the overall performance of the system, this data must not be used during training.

### 2.4.3 Support vector machines

Support vector machines (SVM) is the classifier used for this project. The classifier performs the classification by finding the optimal hyperplane, which is the optimal plane in the feature space that separates the training data. The hyperplane is chosen based on the highest possible margin. Margin is the distance from the hyperplane to the nearest data point [21]. A major advantage of SVM's is that it only relies on a small subset of data called support vectors for classification, for that reason it does not require a large data set.

Assuming the data is linearly separable, a simple linear model can be used, like the one shown in Figure 2 (a). However, if the data is nonlinear, a kernel can be used, where a nonlinear mapping of the data occurs in higher dimensions, making it easier to find the hyperplane. An illustration of such a hyperplane is shown in Figure 2 (b), where the data is mapped from 2 to 3 dimensions to find a linear hyperplane. This decision boundary can then be projected back into its original space.

### 2.4.4 Kernel function

Kernel functions transform low-dimensional input into a higher dimensional, converting a non-separable problem into a separable one. The kernel used in this project is the radial basis function (RBF). RBF has the property that each basis function depends on the radial distance from a center $\mu_j$ so that

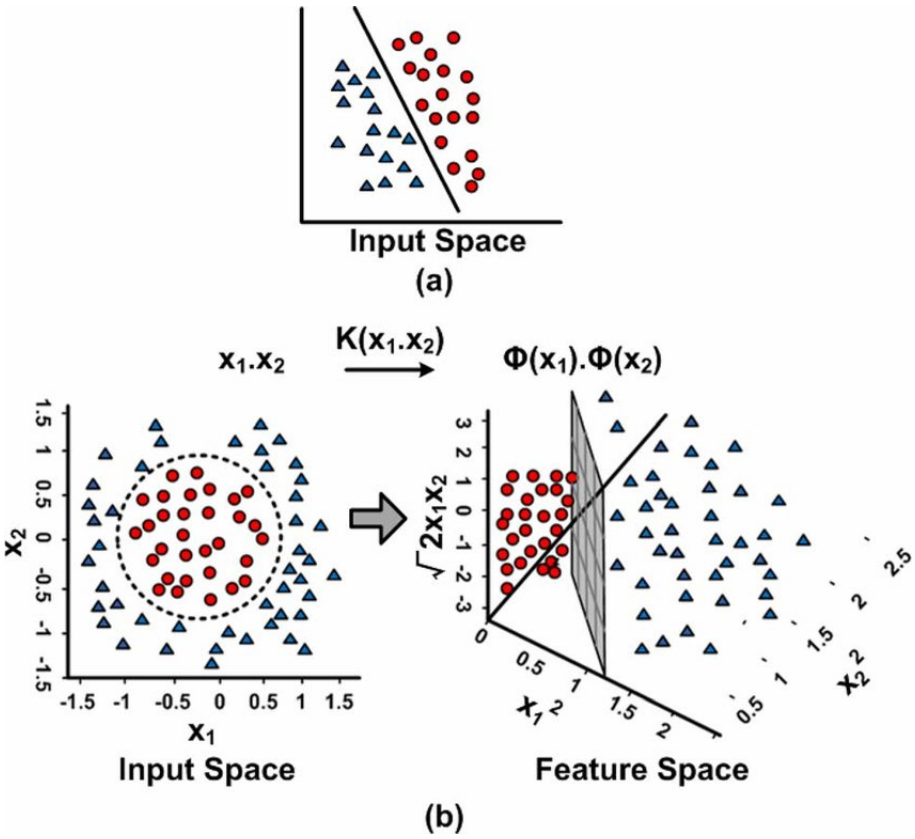


**Figure 2:** Illustration of a linear SVM (a) and a nonlinear SVM (b) [1].

$\phi_j(x) = h(||x - \mu_j||)$. The function uses a similarity measure between the input vector $\{x_1, ...., x_N\}$ and the target vector $\{t_1, ...., t_N\}$ using the Euclidean distance, $||x - \mu_j||$. The goal is to find a smooth function $f(x)$ that fits every target value exactly, so that $f(x_n) = t_n$ for $n = 1, ....N$. To achieve this, $f(x)$ is expressed as a linear combination of RBF's, one centered on every data point. The function is given by

$$f(x) = \sum_{n=1}^{N} \omega_n h(||x - x_n||) \tag{4}$$

Where the values of the coefficients $\{\omega_n\}$ are found by least squares. Since there is the same number of coefficients as there are constraints, the result is a function that fits every target value [21].

### 2.4.5 Selection of hyperparameters

SVM's implemented in the scikit-learn library has two hyperparameters, C and $\gamma$. The selection of the hyperplane depends on the margin, and the margin is influenced by the hyperparameters of the classifier. The kernel coefficient for RBF is $\gamma$, the higher the value of $\gamma$ the higher the generalization error. The classifier tends to overfit for large $\gamma$, as a result, the predictive quality of the classifier becomes bad. C is the penalty parameter of the error term, it controls the trade-off between smooth decision boundary and classifying the training points correctly. Low C encourages a larger margin, therefore allowing for more errors. C can be seen as a regularization parameter.

Both the values of $\gamma$ and C should be balanced, in such a way that the classifier performs well on the test data. To see the effects of the hyperparameters on the classifier, a loss function such as MSE can be used, as shown in (3).

### 2.4.6 Feature selection

Feature selection is an important step in every ML system. This step helps the classifier reduce the number of variables in the data set by using only the relevant and nonredundant ones. In this project mutual information (MI) was used for feature selection. MI measures the statistical dependence between data, which involves detecting any sort of relationship between the data. Everything from mean, variance, or even higher moments. Given two random variables X and Y, MI is given by

$$I(X,Y) = \int_X \int_Y p(x,y) log \frac{p(x,y)}{p(x)(p(y))} dxdy \tag{5}$$

Where p(x) and p(y) are the marginal density functions, and p(x,y) is the joint probability density function of X and Y [22]. Assuming X, and Y are uncorrelated then the value of MI will be zero as the joint probability density function is equal to $p(x,y) = p(x)p(y)$, consequently, $I(X,Y) = 0$. If X and Y are correlated, then $I(X,Y) > 0$.

To find the best features, the mutual information should be maximized with respect to the selected features, $X_s$, and the target variable y.

$$\tilde{S} = \arg\max_S I(X_s; y) \tag{6}$$

Such that $|S| = k$, where k is the number of features we want to select [23]. For this project, only the three best features were used for training, $k = 3$.

### 2.4.7 Performance assessment

A common metric to measure the accuracy of classifiers is the ratio between the true positive (TP), false positive (FP) and true negative (TN). Where TP is the outcome when the classifier predicts a meal onset correctly and FP is the outcome when the classifier predicts a meal onset falsely (no meal is detected as a meal onset). TN is the outcome when the classifier correctly detects no meal onset.

The true positive rate (TPR), also called recall, is defined as

$$TPR = \frac{TP}{TP + FP} \tag{7}$$

The recall is a measure of the classifier's ability to find all the predicted positives. It aims at measuring the proportion of the samples classified as positive which really belong to the positive class.

False positive rate (FPR) is defined as

$$FPR = \frac{FP}{FP + TN} \tag{8}$$

FPR is the probability that the classifier gets a positive value when the true value is negative.

It is common to plot the TPR with respect to the FPR on a graph to represent the performance of a classifier. The better the classifier performance, the higher TPR should be, and the lower FPR should be, and vice versa. A perfect classifier would have a TPR of one and an FPR of zero.

# 3 Equipment and protocol for data acquisition

## 3.1 Introduction

The data recording required a protocol that was approved, that is why the same protocol for the pilot study of "Analysis of bowel sounds related to meal onset" by Konstanze Kölle was used [24]. The following section describes the protocol and equipment used for data gathering.

## 3.2 Recording equipment

A diagram of the equipment used for recording is shown in Figure 3. The sound is picked up using four SPM0687LR5H-1 microphones. The microphone has a uniform frequency response and a high SNR [25]. The microphones were placed in a stethoscope-shaped (disc-shaped) microphone holder to capture the sounds made by the body more clearly. Before placing the microphone on the skin, a double-ended tape (ring-shaped) was attached to the microphone to hold it in place during recording.

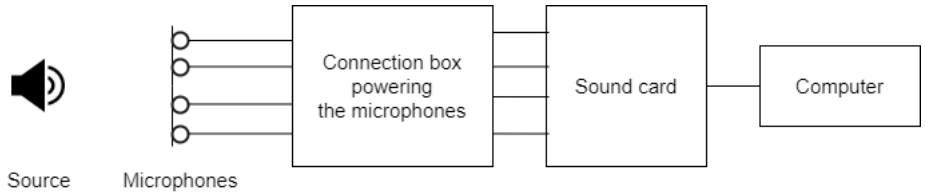


**Figure 3:** Representation of the equipment layout during the data collection.

The four microphones are connected to a box, where they are powered via a power source (battery). The microphone signal is forwarded to a sound card. The sound card used for the project is Roland Octa-Capture, with a bit resolution of 24-bits and a sampling frequency of 48 kHz audio [26]. The output of the sound card is connected to a computer, where the recordings are saved.

## 3.3 Protocol for recording data without noise

For each audio recording, there was a total of four microphones recording simultaneously. Each one of the microphones was attached to a specific location. The location of the microphones can be seen in Figure 4.

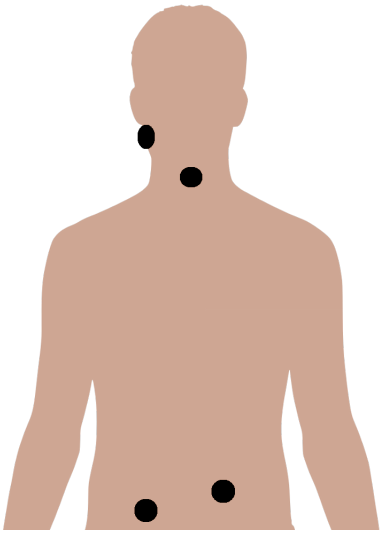

**Figure 4:** Locations of the four microphones.

The first microphone was placed right under the right ear and was used for recording chewing sounds, this recording was not used in this project. The second microphone was placed just over the collar bone on the neck to record swallowing sounds. Different microphone placements were tested in the paper "Automatic detection and recognition of swallowing sounds" [8]. Based on the findings in the paper, the best placement appeared to be just above the collar bone, since these recordings had the highest power. This location was, however, not the most comfortable with regard to head movements. The third and fourth microphones were placed on the lower abdominal region, the right and left part of the abdomen, to capture bowel sounds.

Before the recording sessions, the subjects were asked to fast for a minimum of three hours. The first 15 min of the recording was used as a reference, as the subjects were asked to continue fasting. After that, the subjects could then eat a meal of their choice, often the meal consisted of a slice of bread with cheese and a glass of water. The meal duration had to be less than 15 min. After the meal, the recording continued for another 45 min to monitor the digestive behavior.

During the recordings, the subjects were asked to move as little as possible to avoid disturbing the recordings, since movement created friction noise in the recordings. The subjects were allowed to read a book or use their phone to avoid falling asleep after the meal onset, after 30 min of recording. For this protocol, there was a total of seven recordings, most were 65-75 min long. Nevertheless, only five of them were included in this project, due to problems with two of the recordings. The protocol is illustrated in Figure 5.

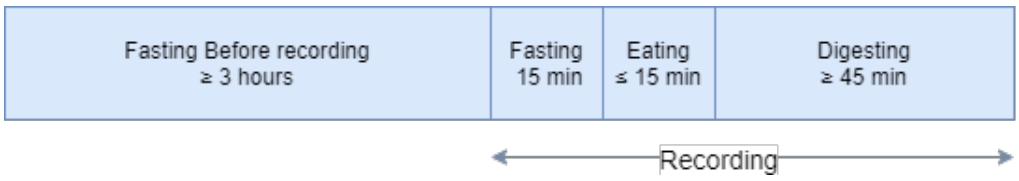


**Figure 5:** Protocol for recording data without noise.

## 3.4 Protocol for recording data with noise

To test the robustness of the classification system, five new recordings were acquired. These new recordings were intended for testing how noise in the swallowing recordings affected the system, and to check how well the system performed in a more realistic environment. The new recordings followed a newly proposed protocol, Figure 6.


**Figure 6:** Protocol for recording data with noise.

In these new recordings, only swallowing and chewing sounds were recorded, hence no fasting was required before recording. The first 5 min of recording were used as a reference, similarly as in the previous protocol. For the next 10 min, the subject was asked to read out loud from a book or a newspaper. After that, the subject could eat a meal of their choice for up to 15 min. After the meal, the subject had to be quiet for 5 min, and then read for another 5 min. Two out of those five recordings had also some reading during the meal onset, the subject was asked to eat for 25-30s and then read for at least 10-15 seconds, and then repeat. This part was added to see if noise during the meal onset could be picked up by the system.

# 4 Method description and implementation

## 4.1 Introduction

The main task of the project was implementing a system for automatic meal detection using bowel and swallowing sounds. Konstanze Kölle's system and research [6] served as a baseline for this project. This section describes how the system was implemented.

## 4.2 Data acquisition

The first seven recordings were acquired using the protocol in Figure 5. These recordings were taken by two subjects (master students). During the recordings the subjects were sitting on a chair, while the microphones were taped on the four locations as shown in Figure 4, the microphones on the abdomen region were covered by clothes for comfort reasons. Out of the seven recordings, only five recordings were used, as two of the recordings were corrupted.

In addition to the first five recordings, five new recordings of data augmented with noise were acquired using the protocol in Figure 6. These new recordings were needed to test the robustness of the system. These five recordings were also taken by the same two subjects. Two out of these five recordings had also noise during the meal onset (see Section 3.4).

In total there were only 10 recordings that were used for this project, and due to lack of time no more recordings were taken. The recordings were uploaded on google drive, in order to make accessing them from google colaboratory easier. Google colaboratory was used as a python compiler, as it provided free access to computing resources including GPUs.

## 4.3 Data Pre-processing

The recorded audio files had a bit rate of 24 bits, while python only worked for 8- and 16-bit rates. Thus it was necessary to quantize the data, in order to convert the bit rate of the files from 24 bits to 16 bits. This process was first performed in python, however, it was later on changed. Due to convenience purposes, the quantization was performed on a digital audio editor, Audacity, because the audio files were too large to handle on google drive.

After quantization, the data was decimated. In order to decimate, the audio files were lowpass filtered to remove the frequency content above 2 kHz. The lowpass, anti-aliasing filter used was the 8'th order Chebyshev type 1 filter. This filter was used due to its flat frequency response in the frequency range of interest, furthermore, the magnitude of the filter at cutoff was equal to -3 dB (0.5). After filtering the data was downsampled to 4 kHz using a downsampling factor of 12. This lead to a reduction in the total file sizes by a factor equal to the downsampling factor.

The final step in the pre-processing was normalization. Each recording was normalized individually. The data was normalized by a linear normalization, as in (1)

## 4.4 Feature extraction

In the feature extraction part of the project, frequency spectrum features were calculated. Before the feature calculation, the recordings were first segmented using four-time segments of length, 10, 20, 30, and 60 seconds. To allow for continuity between the time segments, an overlap of $50\%$ was carried out. The segmentation was used to see whether the segment length had an influence on the performance of the system. After segmentation power spectrum features were extracted. The total power, the power in 100 Hz bands from 0-2000 Hz, and also the power fraction in these frequency bands were calculated. A detailed list of the extracted features is shown in Table 1.

| Features | Number of features |
|---|---|
| Total power | 1 |
| Power in 100 Hz frequency bands from 0 Hz to 2000 Hz | 2-21 |
| Power fraction in 100 Hz frequency bands from 0 Hz to 2000 Hz | 22-40 |

**Table 1:** Description of the extracted features.

These extracted features were then median filtered and used to build a feature matrix for each recording. Each recording from a single microphone had a total of four feature files, one for each time segment. Each feature matrix was also normalized such that all values were between 0 and 1. A typical feature matrix shape is shown in Figure 7, where each row corresponds to a feature at a given time segment, and n is the number of time segments.

$$\begin{vmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,41} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,41} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,41} \end{vmatrix}$$

**Figure 7:** Feature matrix, where rows correspond to a feature at a given time segment, and n is the number of time segments.

Since this project used features from both swallowing and bowel sound recordings, the feature matrix had to include all these features. A new feature matrix with 123 features was built, one for each meal recording. This was built by combining the feature matrices from each meal recording. An example of such a feature matrix is shown in Figure 8, where the first 41 features were from swallowing sounds, the next 41 were from right bowel sounds, and the last 41 were from left bowel sounds.

$$\begin{vmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,41} & b_{1,1} & b_{1,2} & \cdots & b_{1,41} & c_{1,1} & c_{1,2} & \cdots & c_{1,41} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,41} & b_{2,1} & b_{2,2} & \cdots & b_{2,41} & c_{2,1} & c_{2,2} & \cdots & c_{2,41} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,41} & b_{n,1} & b_{n,2} & \cdots & b_{n,41} & c_{n,1} & c_{n,2} & \cdots & c_{n,41} \end{vmatrix}$$

**Figure 8:** Feature matrix, where rows correspond to a feature at a given time segment, and n is the number of time segments, and a, b, c are swallowing, right bowel, and left bowel sound features respectively.

Different feature matrices were also built throughout this project. Some of the built feature matrices included a matrix with only bowel sound features (82 features), and a matrix with only swallowing sound features (41 features). All these matrices were used to train and build different classification systems.

## 4.5 Data processing and analysis

### 4.5.1 Splitting of the data

After pre-processing the data was split into training, validation, and test set. There were in total 10 meal recordings, six of these meals were used as part of the training and validation set. The splitting between the training set and the validation set was done via leave-one-out-cross-validation (LOOCV). Where each one of the six meals was assigned to the validation set, one at a time, via

an iterative procedure, and the remaining meals were assigned to the training set. This maximized the performance of the classifier, given the small data set that was available. The last four recordings were used as part of the test set. Since two of the meals augmented with noise had also noise during the meal onset, one of the meals was always part of the training set and the other was always part of the test set.

Splitting was sometimes performed separately for the meals with and without noise. In the early stages of the project when there were no meal recordings with noise, only five meals were available, thus three of the meals were used for training and validation, while the other two were used for testing.

The data were either distributed at random between the training/validation and test set or distributed according to the subject of the recording. The latter splitting was executed by assigning the meal recordings from one subject to the training/validation set while assigning the meal recordings from another subject to the test set.

### 4.5.2 Training and Classification

The training of the classifier can be divided into three steps. The procedure for training the classifiers is shown in Figure 10, where the arrows indicate how the output of each step builds the basis for the next step. The training starts after feature extraction.

In the **first step**, the data is split, and the response vector is created. The response vector is the labeling vector that will be used to train the classifier, it could also be called the target vector. A visualization of how the response vector should look like is shown in Figure 9. Response delay is the parameter used to account for the delayed onset of audible bowel activity after the meal started, and the response duration is the parameter that is used to find the interval length best suited for early meal detection. Each element of the response vector was either assigned the class "0" or "1". The response vector is assigned the class "0" as default and "1" as a meal indicator [6].

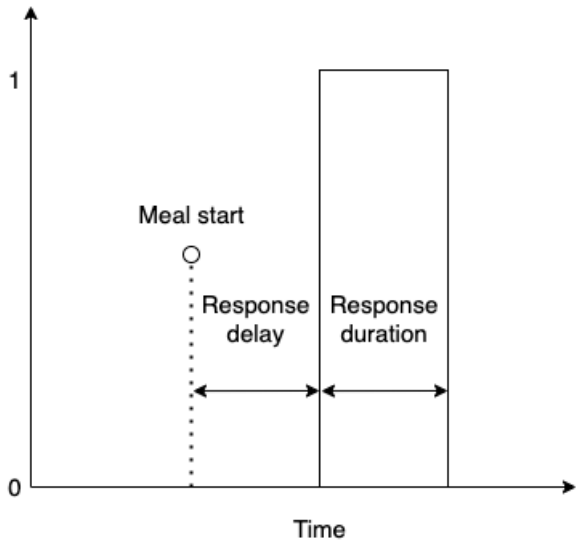


**Figure 9:** Representation of the response vector, where response delay and response duration are relative to the meal start.

There were variations in the segment length of the feature matrix, as well as variations in the response delay and response duration of the response vector, all the possible parameter values are shown in Table 2. In total there were 144 possible parameter combinations, and each combination was used to build a classifier, thus in total, there were 144 classifiers built for each validation meal.

| Parameter | Values | Unit |
|---|---|---|
| Time segment | 10, 20, 30, 60 | s |
| Response delay | 0, 4, 6, 8, 10 | min |
| Response duration | 1, 2, 5, 10, 15, 20 | min |

**Table 2:** Parameter combinations.

Before building the classifiers, the best features were selected using the MI between each feature vector (rows in the feature matrix) and the response vector. Only the three best features were selected, those with the highest cross-correlation score, as they were deemed the most relevant features. These selected features were used to train the SVM classifiers along with the response vector.

A grid search with the values $2^{[-10,-5,0,5,10]}$ for both the $\gamma$ and C was used to tune the hyperparameters of the SVM's. One SVM model was built for each grid point and parameter combination, where only the hyperparameters that resulted in the lowest MSE between the predicted output and the response were selected.

The main takeaway point in this step is that those features and parameter combinations that resulted in the highest number of TP's were selected and forwarded to the next step.

In the **second step**, the selected features and parameter combinations from the first step were used to build a new SVM model. There is no feature selection in this step. A new SVM model was then built using the selected features. Here, the hyperparameters were once again tuned via a grid search with the same values as before. The hyperparameters that resulted in the lowest MSE were selected. The point of this step is to find the tuning for each of the selected parameter combinations.

In the **third step** of the classifier training procedure, the tuned classifiers were trained once more using both training and validation sets, as there was no need for validation since the hyperparameters were already tuned in the previous step. The final classifiers were then tested using the test meals.

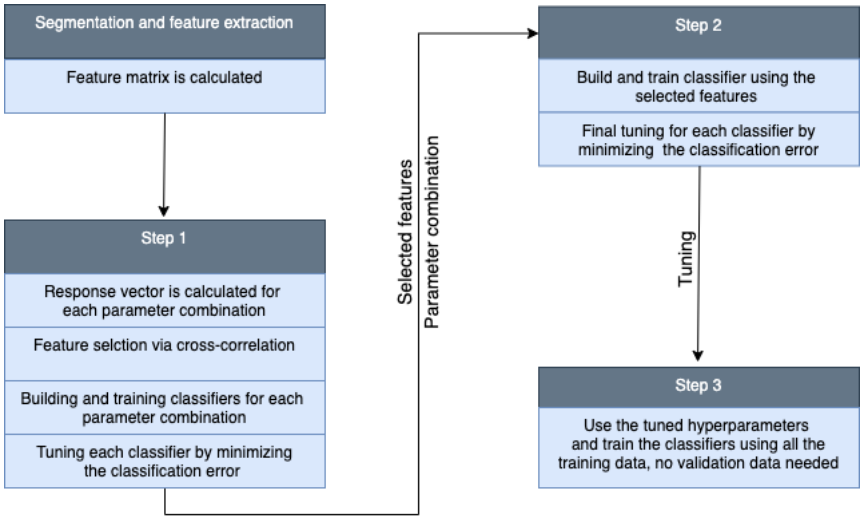


**Figure 10:** Procedure for training the classifiers.

### 4.5.3 Evaluation of the performance

The performance of the 144 classifiers built in the first step of the training procedure was evaluated by the TPR with respect to the FPR graph for both the training and validation meals. Using the graph for the average of all LOOCV runs, the parameter combinations that resulted in the four best classifiers were selected, one for each time segment. The parameter combinations for these four classifiers were then forwarded to the second and third steps of the training procedure.

The final classifiers from the third step were used on the test meals, and to assess the meal prediction a plot showing the predicted response vector against the true response vector was used. Using this plot the numbers of TP's and FP's were calculated. The way a FP was given is by the occurrence of two consecutive ones before the true meal onset, while a TP was given by the occurrence of two consecutive ones after the meal onset.

All recordings had a meal onset 15 min after the recording started, thus any detected consecutive ones after the 15'th min were counted as a TP, while any consecutive ones before that were counted as a FP. A meal is detected when the first two consecutive ones are detected and repeated consecutive ones afterward only work as a confirmation for meal detection.

# 5 Results and observations

## 5.1 Introduction

As mentioned earlier (see section 4.5.1) the data were split based on the data types, either randomly or based on the subject of the recordings. In addition, different feature combinations were tested (see section 4.4). This was used to build and test a couple of different classification systems. In this section, the results and observations from these classification systems will be presented.

## 5.2 Classification using only bowel sound recordings

The first classification system was trained using only bowel sounds, both right and left bowel sound recordings. Only the recordings without noise were used for training and testing, in total three meals were used for training and validation, and two meals were used for testing. Figure 11 shows a plot of the TPR and FPR for each parameter combination for the average of all LOOCV runs.

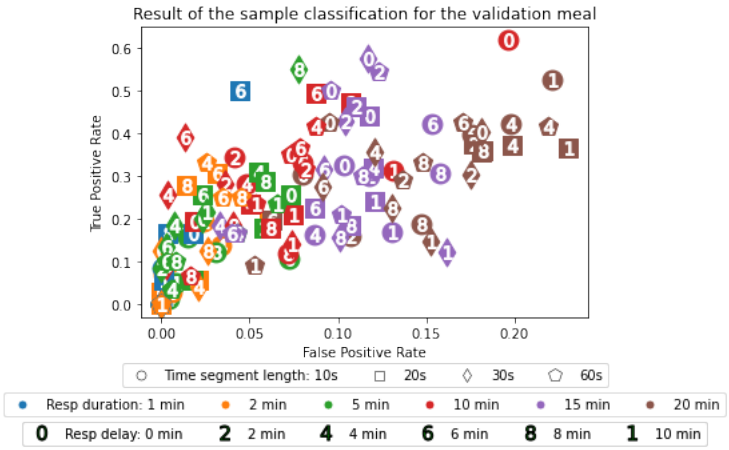


**Figure 11:** True positive rate vs false positive rate for the validation meals in step 1. Each marker represents the average of all LOOCV runs for a given parameter combination.

.

The parameter combinations that resulted in the best classifiers were picked from Figure 11, one parameter combination for each time segment. These parameter combinations were then used in steps 2 and 3 in the training process.

The selected parameter combination for each time segment is presented in Table 3.

| Delay | Duration | Segment |
|---|---|---|
| 2 | 2 | 10 |
| 2 | 15 | 20 |
| 0 | 15 | 30 |
| 2 | 15 | 60 |

**Table 3:** Best parameter combination for each time segment for classification using bowel sound recordings.

The final classifiers with these parameter combinations resulted in the following count for TP's and FP's for the test meals, Figure 12. It's clear that the classifier didn't perform well enough as there were multiple FP's that were counted. The classifier with the 10s time segment performed the worst, as there were no TP's, and only one FP was counted.

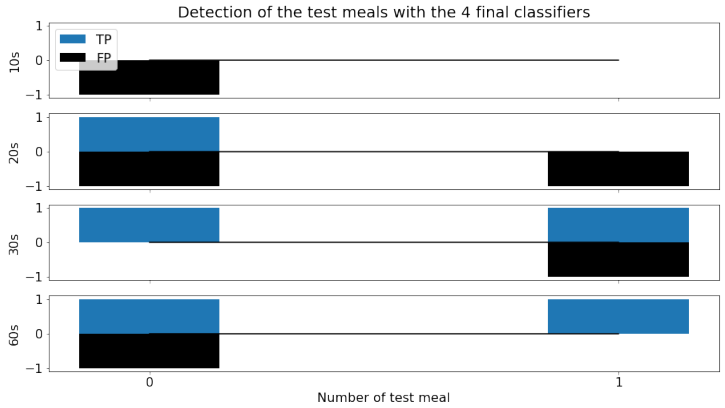


**Figure 12:** True positive and false positive meal detection for each test meal, for the four final classifiers.

The predicted response vector was plotted against the true response vector for the 10s and 20s time segment for both test meals, Figure 13 and 14. The true response vector was plotted with the same parameter combinations as the predicted response vector, however, the meal delay used was always zero. This was plotted to visualize how well the classifier predicted a meal onset. For both meals, the actual meal onset was at the 15'th min of the recording.

The classifier with 10s time segment, Figure 13, was not able to predict much, only a couple of consecutive ones were labeled at the beginning of the first test meal indicating a meal detection 15 min before the actual meal onset, and

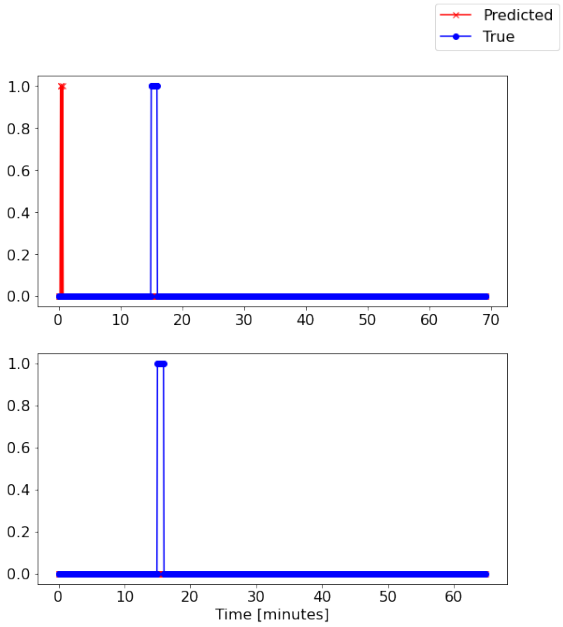


**Figure 13:** True vs predicted response vector for the test meals, for the final classifier with 10s time segment. The graph on the top and bottom are respectively the first and second test meals.

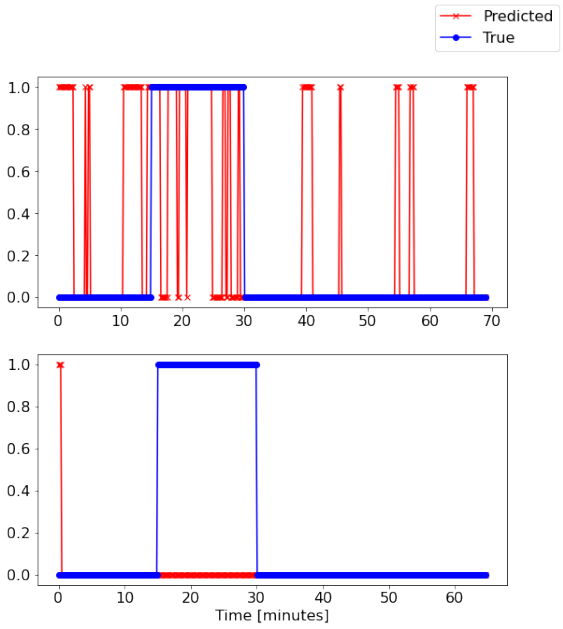


**Figure 14:** True vs predicted response vector for the test meals, for the final classifier with 20s time segment. The graph on the top and bottom are respectively the first and second test meals.

hence a FP was counted. As for the second meal no prediction was present, thus no meal was detected.

As for the classifier with the 20s time segment, Figure 14, there were multiple false meal labels for the first meal. This could be seen clearly in the graph since there were multiple labeled ones both before and after the meal onset, that is why both a TP and a FP were counted. As for the second meal, there were ones labeled at beginning of the meal recording, thus a FP was counted. For both test meals, a meal was detected 15 min before the actual meal onset.

## 5.3 Classification using bowel and swallowing sound recordings

For this classification system, swallowing sound features were combined with the features from both the right and left bowel sounds, to train, validate and test the classification system.

This classification system was built using data without noise. The same three meals from earlier were used for training and validation, and the two remaining meals were used for testing. The TPR and FPR are plotted for each parameter combination in Figure 15 for the average of all LOOCV runs. The best classifiers in this system outperformed the best classifiers in the previous classification system, with regards to TPR and FPR.

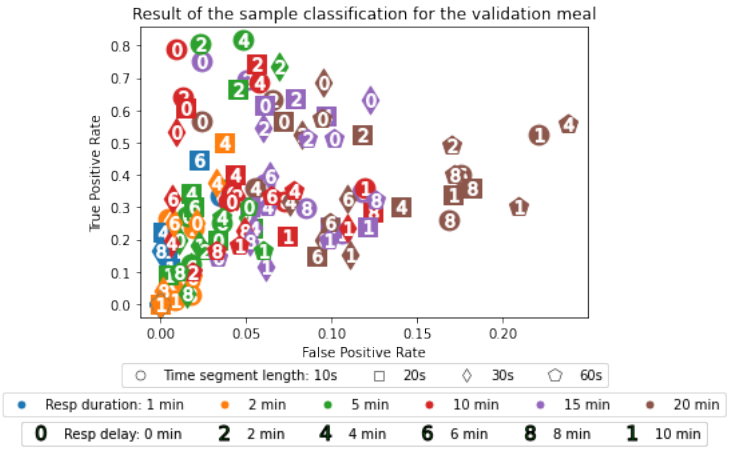


**Figure 15:** True positive rate vs false positive rate for the validation meals in step 1. Each marker represents the average of all LOOCV runs for a given parameter combination.

Once again the parameter combinations that resulted in the best classifiers were selected and used for training in steps 2 and 3. The parameter combination for this system is presented in Table 4. The four final classifiers were then used on the test meals, and the TP and FP count for each classifier is shown in Figure 16.

| Delay | Duration | Segment |
|---|---|---|
| 0 | 10 | 10 |
| 2 | 10 | 20 |
| 4 | 5 | 30 |
| 0 | 20 | 60 |

**Table 4:** Best parameter combination for each time segment for classification using bowel and swallowing sound recordings.

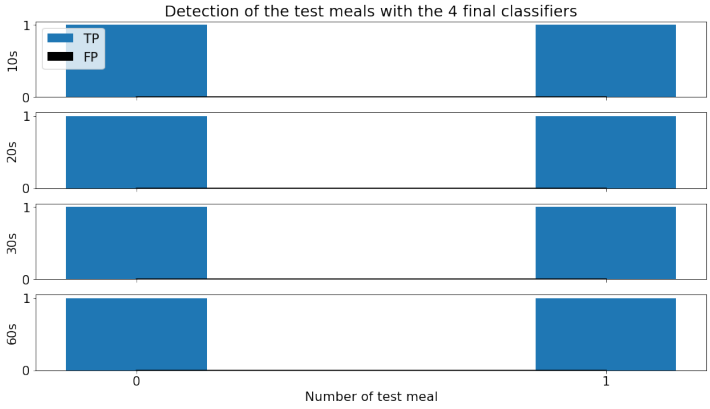


**Figure 16:** True positive and false positive meal detection for each test meal, for the four final classifiers.

The classifiers had no FPs, only TPs. To understand why the classifiers performed so well, the predicted response vector was plotted against the true response vector for the 10s and 20s time segment, Figure 17 and 18. From these plots, there was no meal labeling before the true meal onset, which explains why there were no FP's. Both classifiers were able to label the meal during the actual meal duration. There seems to be some false labeling at the end of the meals for both classifiers.

The features selected by each classifier in the LOOCV iterations are shown in Figure 19. Most features seemed to be selected from the swallowing sound features, especially the features from 0-20.

Another data splitting modality was also tested for this system. The data was split based on different subjects, the system was trained using meals from one

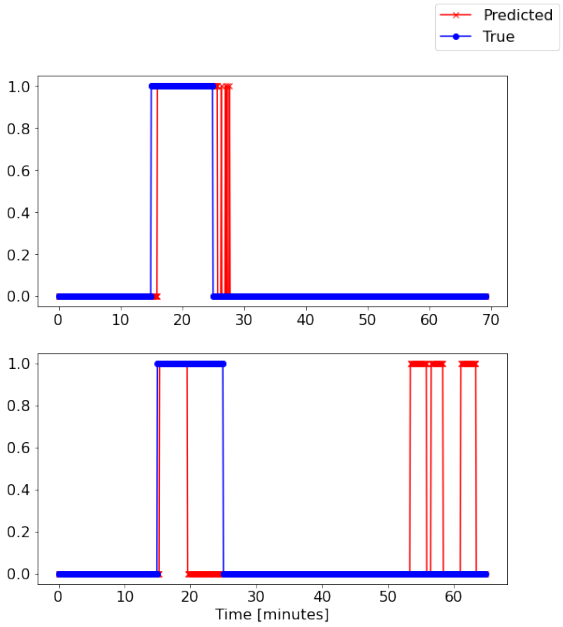


**Figure 17:** True vs predicted response vector for the test meals, for the final classifier with 10s time segment. The graph on the top and bottom are respectively the first and second test meals.

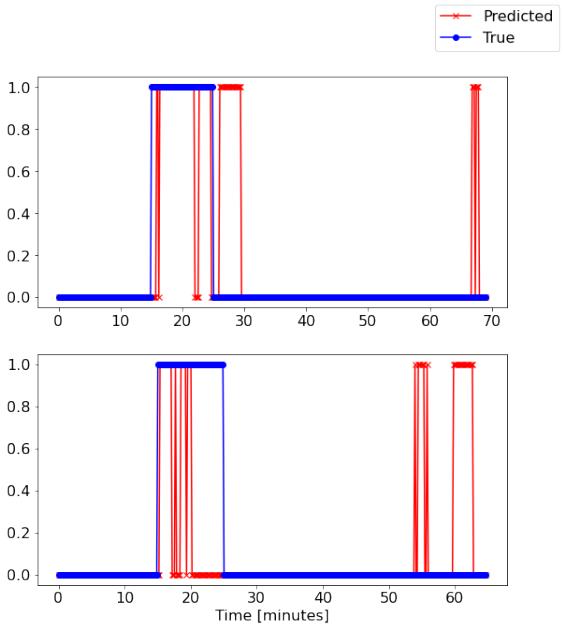


**Figure 18:** True vs predicted response vector for the test meals, for the final classifier with 20s time segment. The graph on the top and bottom are respectively the first and second test meals.

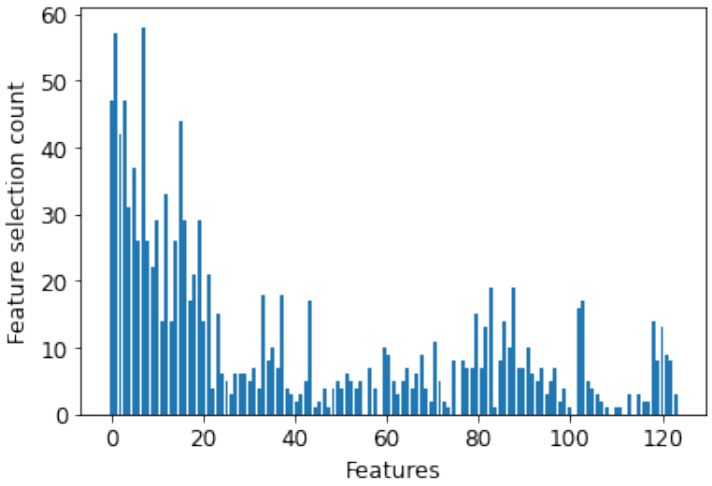


**Figure 19:** Features selected during every LOOCV iteration, where the first 41 are swallowing sound features, the next 41 are right bowel sound features, and the last 41 are left bowel sound features.

subject and tested using meals from another subject. In total, there were three training meals and two test meals. The results from this system were similar to the results shown above, that is why the results were not included in this section, but rather in Appendix A.

## 5.4 Classification using data augmented with noise

For this project, only swallowing recordings were augmented with noise, hence all classification systems built in this subsection are trained and tested using swallowing sound recordings only. Two classification systems were built, one of which was not trained using data augmented with noise, while the other was trained using data augmented with noise. Both systems were tested using data augmented with noise.

### 5.4.1 Training without data augmented with noise

The first classification system was trained and validated using data without noise, and then tested using data augmented with noise. It was trained using three training data chosen randomly from the meals without noise and tested using two meals augmented with noise. One of the test meals augmented with noise had also noise during the meal intake (the first meal). TPR and FPR for

all parameter combinations for the average of all LOOCV runs for this system are shown in Figure 20. This classifier had similar TPR and FPR values to the classifier trained with all features (bowel and swallowing sound).

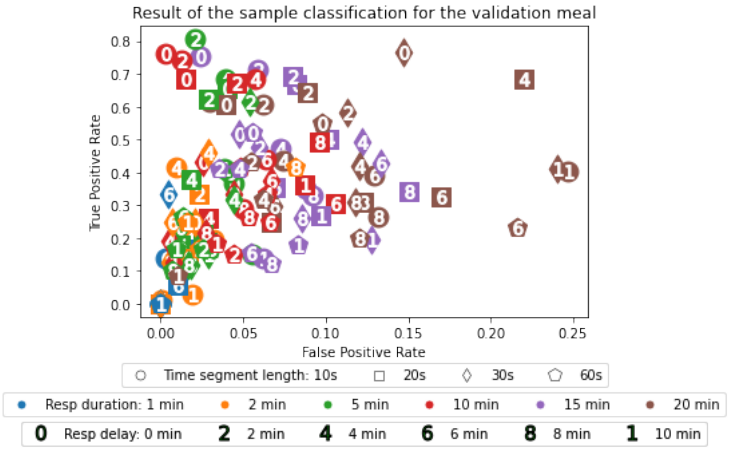


**Figure 20:** True positive rate vs false positive rate for the validation meals in step 1. Each marker represents the average of all LOOCV runs for a given parameter combination.

The counted TP and FP for each test meal are plotted in Figure 21. This plot showed that the classifiers did not label anything for the 30s and 60s time segments, while for the 10s, and 20s time segment both a FP and a TP was counted. Once again the predicted and the true response vector for 10s and 20s time segments was plotted, Figure 22 and 23. It could be seen that for both time segments, the classifier predicted the talking in the recording as a meal, both before and after the meal onset, that is why both a TP and a FP were counted.

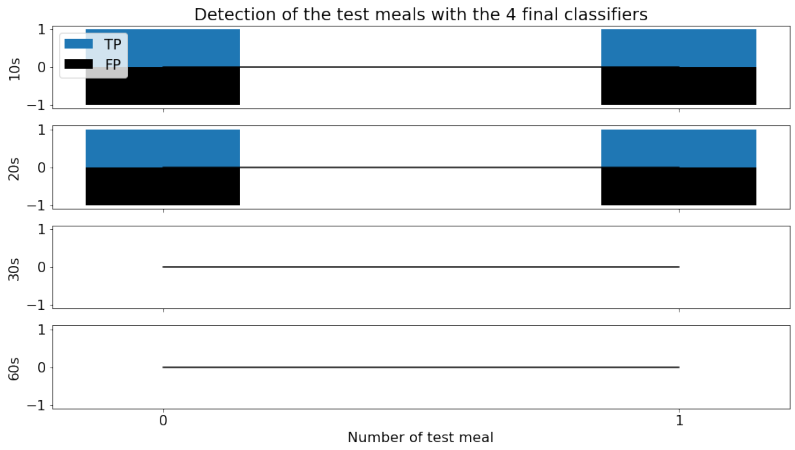


**Figure 21:** True positive and false positive meal detection for each test meal, for the four final classifiers.

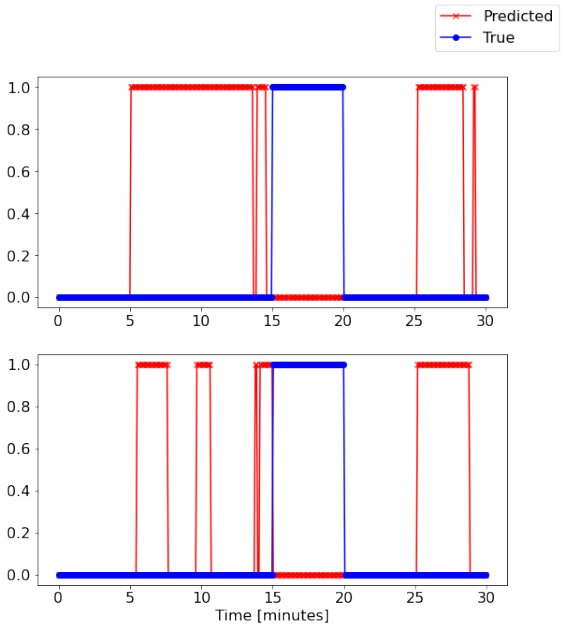


**Figure 22:** True vs predicted response vector for the test meals, for the final classifier with 10s time segment. The graph on the top and bottom are respectively the first and second test meals.

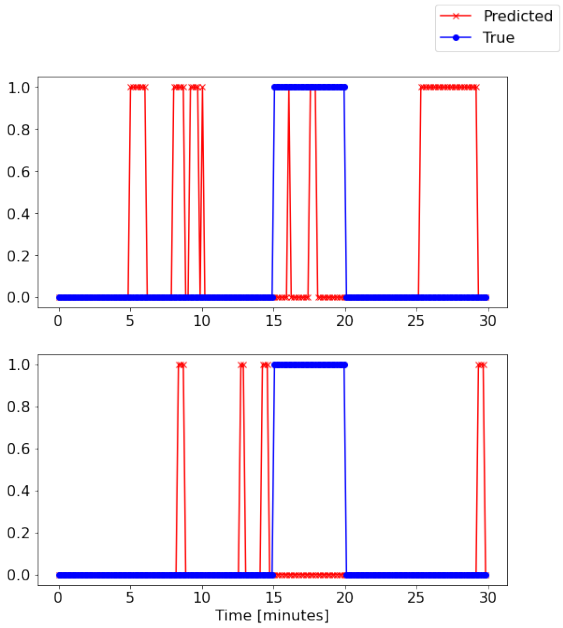


**Figure 23:** True vs predicted response vector for the test meals, for the final classifier with 20s time segment. The graph on the top and bottom are respectively the first and second test meals.

### 5.4.2 Training with and without data augmented with noise

The second classification system was built using a combination of both data augmented with and without noise. The training and validation data set consisted of six meals, three with noise, and three without noise. The test set consisted of the same test meal as in the previous classification system, two meal recordings augmented with noise.

The TPR and FPR for all parameter combinations for the average of all LOOCV runs for this classification system are plotted in Figure 24. The TPR for this system was much lower than all the previously built systems, however, the FPR was surprisingly low.

TP's and FP's were plotted in Figure 21. It seems from the plot that the four final classifiers were much better at predicting a meal onset when compared to the final classifiers in the previous system. The final classifiers were able to label almost all the meals.

To better understand why the classifiers behaved like this, once again the true response vector was plotted against the predicted response vector for the classifiers with 10s and 20s time segment, Figure 26 and 27. From these two figures,

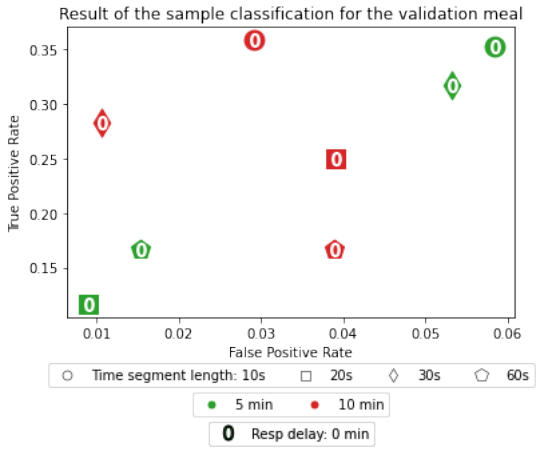


**Figure 24:** True positive rate vs false positive rate for the validation meals in step 1. Each marker represents the average of all LOOCV runs for a given parameter combination.

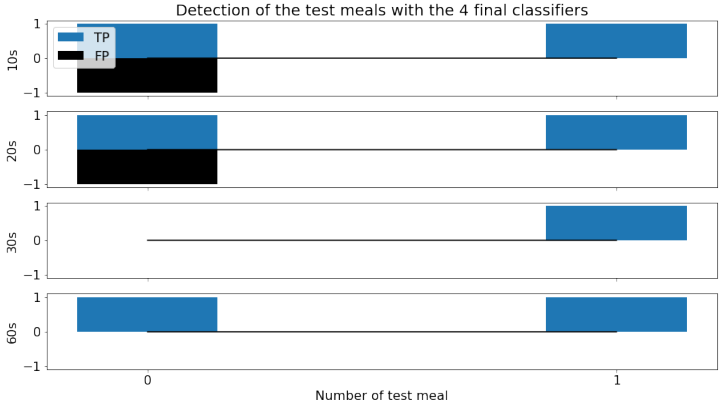


**Figure 25:** True positive and false positive meal detection for each test meal, for the four final classifiers.

it is possible to see that the classifier had a better prediction for the meal. Unlike the previous classifiers, the talking was not labeled as a meal, however, some of the silent parts were labeled as a meal. It seems that the classifiers only counted FP's for the first meal, the meal which had some talking during the meal onset.

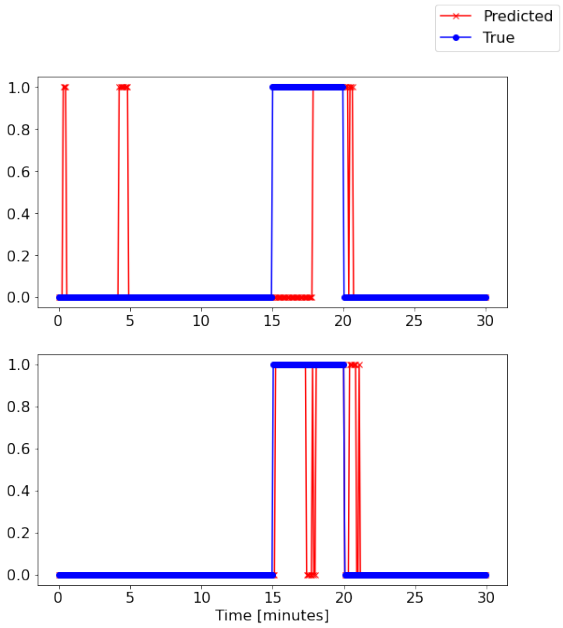


**Figure 26:** True vs predicted response vector for the test meals, for the final classifier with 10s time segment. The graph on the top and bottom are respectively the first and second test meals.

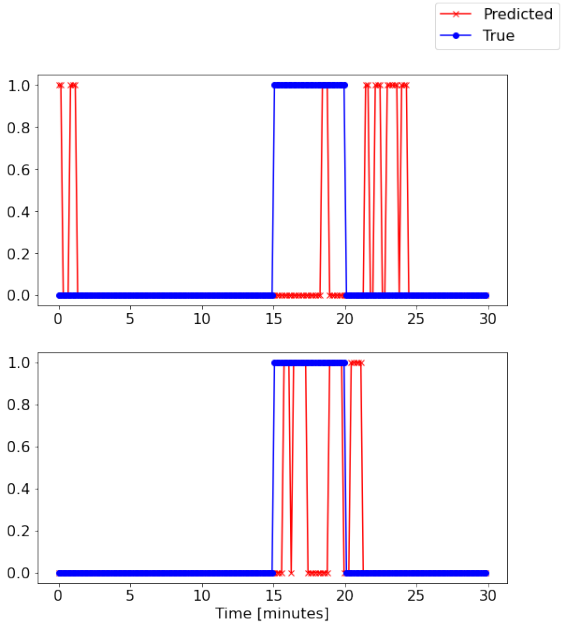


**Figure 27:** True vs predicted response vector for the test meals, for the final classifier with 20s time segment. The graph on the top and bottom are respectively the first and second test meals.

## 5.5 Additional plots and observations

Additional findings, plots, and results are presented in Appendix A, such as the 30s and 60s time segment for the classifiers predicted and true response vector, and the TPR and FPR plot for all parameter combinations for the average of the training data.

# 6 Discussion

## 6.1 Evaluation of the performance of classification using only bowel sound recordings

The classification system that was trained using only bowel sounds did not perform that well, as it struggled to detect a meal onset, even using the four final classifiers. When compared to Konstanze Kölle's system [6], which only had features from a single bowel location, the addition of an extra bowel sound(right/left bowel) feature to the feature matrix did not improve the performance of the classifiers, not one bit.

The addition of the extra recording seems to have affected how well the different parameter combinations performed. As it's clear in Figure 11 that the classifiers built were separable in duration, the higher the duration is the higher is the FPR. The classifiers with a duration of 20 min had the highest FPR. The classifiers with a duration of 10 and 15 min had low FPR and good TPR, these classifiers were also among the best classifiers. A reason for this could be that the classifier's duration aligned with the real duration of the recorded meals, which was often between 8-12 min.

Looking at Figure 12 there seems to be a couple of meals with both a TP and a FP, this follows directly from the messy true vs predicted label plots, like the one in Figure 14 for the first meal. The predicted response vector has consecutive ones all over the plot, both before and after the meal onset. This follows naturally from the way bowel sounds are, as bowel sounds may cycle from peak to peak with a period over 50-60 minutes [27]. Thus bowel sound peaks appear all over the recording, and are more apparent for the right bowel recordings, this clearly affected the feature selection and training of the SVM classifiers. Another thing that also affected the prediction was the noise in the bowel sound recordings, as any movement by the subject led to friction noise between the microphone and the skin. This noise was especially apparent at the beginning of the recordings as subjects usually adjusted their seating. This is evident in Figure 13 and 14 where meal labels are assigned at the beginning of the recording. All these effects were captured differently by the different time segments.

For this classification system, the meal detection time for the test meals varied from one classifier to another. Some had false meal detection 15 min before the actual meal onset, due to FP's, while others had no meal detection as the classifiers were not able to detect anything. Meal detections occurred seldom within the actual meal onset duration.

This classification system suffered from many FP's, thus it will not be safe to use such a system to help CGM systems with meal detection as it could lead to a false insulin dosage, which could be life-threatening for diabetes patients.

## 6.2 Evaluation of the performance of the classification using bowel and swallowing sound recordings

The classification system that was trained using both bowel and swallowing sound recordings outperformed the classification system that was trained using only bowel sound recordings. The best classifiers had TPR as high as 0.85 and FPR as low as 0.01, as can be seen in Figure 15. Just like the previous classification system, the classifiers with the duration of 20 min had the highest FPR and often performed the worst. Classifiers with the duration of 5 and 10 min were among the best classifiers, this aligns with the increase in power for the swallowing features, which lasted throughout the whole meal duration (8-12 min).

For this classification system, no FP's were detected for any of the test meals using the final four classifiers, as could be seen in Figure 16. The reason behind this could be easily understood by looking at Figure 17 and 18. There was no meal labeling by the classifiers before the actual meal onset. Most predicted meal labels are kept within the true meal duration, however, there seems to be some false meal labeling at the end of the recordings. The labeling at the end of the recordings does not affect the meal detection, it only confirms the meal detection. These predictions at the end of the recording are caused by noise due to movement at the end of the recording. It's possible to reduce such noise in the predictions by training the SVM classifiers with data that has more labeling. Meal onset, friction noise, speech, or any other disturbances in the recordings could be labeled before being fed into the SVM classifier for training, theoretically, this should improve the performance of the system.

The classifier improved performance could be explained by looking at how the addition of swallowing sounds affected the feature selection. Swallowing sound features have a large increase in power during a meal onset. Most selected features were swallowing features, as could be seen in Figure 19. The features in the frequency range of 400-1000 Hz (100 Hz power band feature) for swallowing sounds had the highest power during meal intake when compared to all the other features. For this reason, the majority of the classifiers selected features only from swallowing sounds. The power at the 100 Hz band from 1600-1700 Hz was also among the most selected features for swallowing sounds, this feature is related to the type of food that was consumed. Frequencies above 1000

Hz seemed to provide important meal information, some of this information is lost due to lowpass filtering of everything above 2000 Hz. These frequencies must be investigated as part of future work.

All recordings were performed in an almost ideal environment, without any noise, since the subjects were asked to stay quiet and minimize their movements as much as possible. All these factors could have affected the results of the classifier since the noise clearly impacted the classifier's meal labeling. Had the recording environment been more realistic, by having more natural movement by the subject and more noise, then the feature selection and the training would have certainly been affected, especially by the low-frequency noise that could overlap with the frequency range of the features. For this reason, five new recordings were obtained, all augmented with noise. The main idea was to see to what extent noise affected the system's meal labeling.

This classification system was also tested using two different data splitting modalities to test whether the selection of training and test set impacted the system's performance. The training, validation, and test data were split either randomly or based on the subject of the recording. However, both modalities provided similar results, as it seemed that the most relevant features did not vary from one subject to another. Still based on these findings, it's not possible to conclude whether the system is subject-independent or not, since all the recordings came from only two subjects. Both subjects were males and both of them were of similar age and physical condition. Therefore more data from subjects of different ages, physical conditions, and genders is needed before concluding to what extent the system is subject-independent.

This classifier had an average meal detection time of about 1-2 min and was able to correctly predict the meal duration to a certain extent. A downside to this classifier is that it was easily affected by noise, as friction noise picked up by the microphone was labeled as a meal onset by the classifier. All things considered, this system is far from perfect and could not be used for early meal detection to aid CGM systems as of now. However, as more data with more variability and better labeling is collected this might change.

## 6.3 Evaluation of the performance of the classification using data augmented with noise

### 6.3.1 Training without data augmented with noise

The new data augmented with noise was only obtained for swallowing sounds, thus there was no need to train the classification system using all the parameter combinations, as they were mainly used for bowel sounds to aid in labeling the response vector. However, all the parameter combinations were still used to ensure that they had no effect on the performance of the classifier.

Looking at Figure 24, it's clear that the best classifiers were those that actually had a meal delay of 0 and 2 min and a duration of 5, 10, and 15 min. These classifiers were only trained using the swallowing sounds from the first five recordings (no noise). During the meal duration, the power of the frequency features increases, the classifiers with meal duration and delay close to the actual meal values performed the best during the meal labeling since the training labels (response vector) were then more accurate, hence better training. Since the same recordings were used, the true meal duration was still between 8 to 12 min.

It should be noted that the classification system trained using only swallowing features performed in a similar manner to the classification system that was trained using all features. Just as mentioned earlier, most of the selected features for the best classifiers were from swallowing sounds, thus the system was already too dependent on the swallowing features.

The test set consisted of two swallowing sound recordings, both augmented with noise. One of the recordings had also some noise (reading) during the meal onset, this was to test whether the system could distinguish noise from the meal. The classifiers with the time segments of the 30s and 60s struggled with detecting a meal onset as could be seen in Figure 21. There was nothing classified in the predicted response vector for these segments, as the SVM model was completely off and therefore was not able to predict anything. As for the classifiers with the time segments of 10s and 20s, the classifier was able to classify a meal onset, and both a TP and FP were counted for both meals. Looking at Figure 22 and 23, there were almost no meal detections during the actual meal duration at all. The classifier labeled the noisy (reading) parts of the recording as a meal. This classification system had no actual reliability whatsoever, seeing that no meal was labeled correctly.

### 6.3.2 Training with and without data augmented with noise

For this system, only a few parameter combinations were used to speed up the training process. To improve the performance of the previous system, a better SVM model with proper training data was used. The training data consisted of both data with and without noise. Figure 24 at first glance may give the impression that the system is unusable, however, by looking at the TP and FP count in Figure 25 there seems to be a substantial improvement when compared to the previous classification system. There were predictions for all classifiers, even for the 30s and 60s segment.

Figure 26 and 27 shows that the classifiers were better at correctly labeling the actual meal onset, it was better at predicting the true response vector. The classifiers were also better at ignoring the noisy parts of the recordings, however, the silent parts of the recordings were falsely labeled as a meal onset. The classifier struggled with the first meal since the meal had noise during the meal onset. For the first meal, the classifier counted both a FP and a TP for the 10s and 20s time segments and struggled with predicting anything for the 30s time segment. This was probably due to the bad labeling of the data set, as the noisy parts of the recordings were not labeled properly, and thus the SVM classifiers were trained to label a meal and ignore noise while being trained with noise that is labeled "falsely" as a meal.

Even though the classifier was limited by the quality of the data labeling, it performed quite well, as the average meal detection time was about 2-3 min for the meals that were only given a TP. This shows promising results, as with a better training set, the system's performance would improve a lot.

# 7 Conclusion

This project investigated the feasibility of early meal detection by the use of both swallowing and bowel sounds. The support vector machine classifiers built for this project showed promising results even with the limited amount of data that were available. The best classifiers were able to achieve a recall as high as 0.85, with an average meal detection time of 1-2 min. These results proved that it is possible to improve Konstanze Kölle's system by introducing swallowing sound features to the system.

The classifiers performed well without noise, but once the noise was introduced the performance of the classifiers was degraded drastically, as it struggled to distinguish between noise and meal onset in the recordings. The system was trained using a limited set of data, and due to the lack of proper labeling, the noise was often predicted as a meal.

The classification system shows great potential, however, as it stands, the system can not be used to aid CGM systems since the system is easily affected by noise, hence more investigation is required to have a better assessment of the system's performance. To conclude, in a feature work, as the system is improved and more data is collected, the feasibility of early meal detection based on bowel and swallowing sounds will be assessed.

# 8 Suggestions for future work

## 8.1 Testing different filtering methods for swallowing sound recordings

As the system stands right now, both bowel and swallowing sound features are pre-processed in a similar manner. This is not ideal as there is some meal information in the frequency range of 2 kHz and above which are missed upon swallowing sounds due to filtering. This frequency range provides direct information about the meal, such as the hardness or softness of the food, this information might be helpful in meal detection. This needs to be further investigated, and ideally different pre-processing should be used for swallowing sound recordings.

## 8.2 Testing different feature calculation and selection methods

The feature calculation should also be changed to take into account the difference in the frequency content of bowel and swallowing sounds. Swallowing sounds features should include higher frequency content since the higher frequency content provide meal information,

Another thing that must be investigated is the features that are extracted for this project, as the power spectrum features were introduced and used mainly for bowel sound recording. There might be other features for swallowing sounds, that reduce the impact of noise in the features. Other features should be considered such as nonlinear features, and non-frequency-based features because features affect the training of the classifiers.

Another thing that also must be considered is how features are selected, as the system is right now, only the three best features are selected, most of the time only swallowing features are selected. Different feature selection methods should be considered to see how it impacts the classification system. Because as it stands right now there is almost no need for combining bowel and swallowing recording. This should be investigated.

## 8.3 Collecting more data

Most data for this experiment were obtained in a controlled environment and does not represent a realistic recording environment. This was a major drawback for the system, as when new data (data augmented with noise) were added the classifier performance was degraded drastically. Recordings should include more natural movement, noise such as talking during the meal, and also more natural noise from the environment.

Another thing that could affect the classifier was the lack of labeling. There was a lot of noise due to friction between the sensors and the skin, and since it was not labeled, the system was not able to distinguish between noise and meal during predictions. All training data must be properly labeled.

The amount of variability in the data that was available for this experiment was also lacking. There were in total 10 recordings that were used in this project, all these recordings were from two subjects, hence there was little to no variability in the data. New data must be recorded, the data should include people of different age groups and physical conditions. The new data should also include an equal amount of data from both genders.

## 8.4 Testing different parameter combinations

The parameter combinations that were used in this project were proposed for the bowel sound recordings by Konstanze Kölle [6], however, they are irrelevant for swallowing sound recordings. Swallowing sounds features seemed to only be affected by the segment length, as the longer the segment length is the more swallowing information is captured, and vice versa. Different time segments lengths should be tested, to see what works the best for both swallowing and bowel sound features.

## 8.5 Testing different classifiers

Different classifiers should be tested to see whether the performance of the system could be improved. SVM classifiers worked well for meal detection using bowel and swallowing sound recordings, but there might be other alternatives that perform even better. The use of other classifiers, such as hidden Markov models and Gaussian mixture models should be investigated.

## 8.6 What should the final system look like

Another thing that should be looked into is what the final system could look like. If the system is gonna measure both swallowing and bowel sounds, in what way should the system be designed, with regard to both comfort and robustness. Should the sensor be acoustic, or would a less visible sensor be more appropriate for the patient? Also, another thing to look at would be the sensor placement, as it might not be that comfortable to have the swallowing sensor just above the collar bone for a long period. Other approaches such as placing it at the backside of the neck or at the chest should be investigated.

# 9 Bibliography

[1] M. A. B. Altaf and J. Yoo, "A 1.83 j/classification, 8-channel, patient-specific epileptic seizure classification soc using a non-linear support vector machine," *IEEE transactions on biomedical circuits and systems*, vol. 10, 02 2015.

[2] U. D. of Health & Human Services, "What is diabetes?," 2021.

[3] M. clinic staff, "Hyperglycemia in diabetes," 2020.

[4] unkown, "Gestational diabetes," 2019.

[5] Steven J. Russell, MD, PhD, Harvard Medical School, "Continuous glucose monitoring," 2017.

[6] K. Kölle, "Feasibility of early meal detection based on abdominal sound," Master's thesis, Norwegian University of Science and Technology, 2019.

[7] O. Makeyev, P. Lopez-Meyer, S. Schuckers, W. Besio, and E. Sazonov, "Automatic food intake detection based on swallowing sounds," *Biomedical signal processing and control*, vol. 7, pp. 649–656, 11 2012.

[8] H. Khalifi, "Automatic detection and recognition of swallowing sounds," *Reaserch gate*, 2021.

[9] T. Olubanjo and M. Ghovanloo, "Real-time swallowing detection based on tracheal acoustics," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4384–4388, 2014.

[10] Y. Bi, M. Lv, C. Song, W. Xu, N. Guan, and W. yi, "Autodietary: A wearable acoustic sensor system for food intake recognition in daily life," *IEEE Sensors Journal*, vol. 16, pp. 1–1, 01 2015.

[11] Trief, P. M., Cibula, D., Rodriguez, E., Akel, B., & Weinstock, R., "Incorrect insulin administration: A problem that warrants attention," *Clinical diabetes : a publication of the American Diabetes Association*, vol. 34, pp. 25–33, 01 2016.

[12] R. Triggs, "What you think you know about bit-depth is probably wrong," 2021.

[13] D. Greene, *Decimation and Downsampling*. Rice, 2021.

[14] U. Jaitley, "Why data normalization is necessary for machine learning models," 10 2018.

[15] R. Ranta, V. Louis-Dorr, C. Heinrich, D. Wolf, and F. Guillemin, "Digestive activity evaluation by multichannel abdominal sounds analysis," *IEEE transactions on bio-medical engineering*, vol. 57, pp. 1507–19, 02 2010.

[16] H. Yoshino, Y. Abe, T. Yoshino, and K. Ohsato, "Clinical application of spectral analysis of bowel sounds in intestinal obstruction," *Diseases of the Colon & Rectum*, vol. 33, pp. 753–757, 1990.

[17] M. Taniwaki and K. Kohyama, "Fast fourier transform analysis of sounds made while swallowing various foods," *The Journal of the Acoustical Society of America*, vol. 132, no. 4, pp. 2478–2482, 2012.

[18] S. D. Yongxin Luo, "Power spectral density," *Science Direct*, 2007.

[19] E. Burns, "machine learning," 2021.

[20] I. cloud Educatuin, "Supervised learning," 2020.

[21] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 2007.

[22] Ross, Brian C., "Mutual Information between Discrete and Continuous Data Sets," *PLOS ONE*, vol. 9, pp. 1–5, 02 2014.

[23] T. Huijskens, "Mutual information-based feature selection," 10 2017.

[24] K. Kölle, "Protocol for the pilot study: Analysis of bowel sounds related to meal onset," 2021.

[25] Octopart, *SPM0687LR5H-1*, 2018.

[26] Roland, *Roland UA-1010 Octa-Capture*, 2020.

[27] A. Godfrey, "Listening to bowel sounds: An outdated practice?," 2017.

# A Additional results and observations

In this section, additional results and observations are presented for the classification systems that were built in this project.

## A.1 Classification using only bowel sound recordings

First, the TPR and FPR for all parameter combinations for the average of all training meals are plotted, Figure 28. The TPR and FPR were similar to the average over the LOOCV, Figure 11.

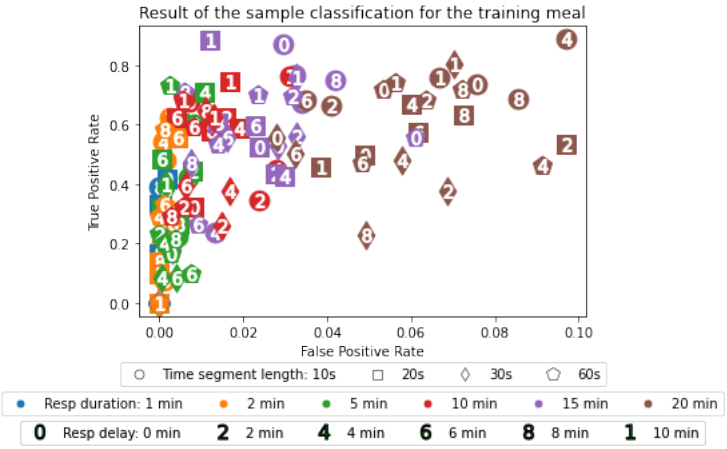


**Figure 28:** True positive rate vs false positive rate for the training meals in step 1. Each marker represents the average of all training meals for a given parameter combination.

.

The predicted and the true response vector is plotted for the 30s, and 60s time segments, Figure 29 and 30, as only the 10s and 20s segments were shown in the main part of the report. For the 30s time segment, Figure 29, only a TP was counted for the first meal, and the meal was detected almost 42 min after the meal onset. As for the second meal, both a TP and FP were counted and the meal was detected 14 min before meal onset. While for the 60s time segment, Figure 30, both a TP and a FP were counted for the first meal, and the meal detection time was around 14 minutes before the actual meal onset. As for the second meal, only a TP was detected, no FP was issued as there were no consecutive ones before the meal onset, meal detection time was 34 minutes after the meal onset. The meal detection time for the classifiers was either long after, or long before the actual meal onset.

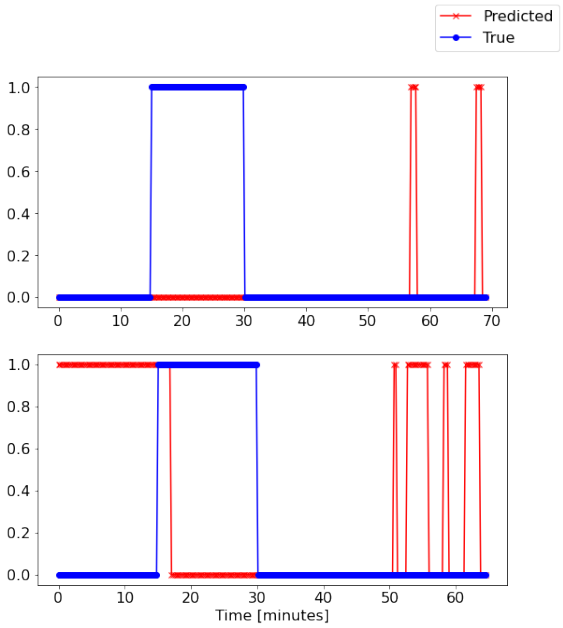


**Figure 29:** True vs predicted response vector for the test meals, for the final classifier with 30s time segment. The graph on the top and bottom are respectively the first and second test meals.

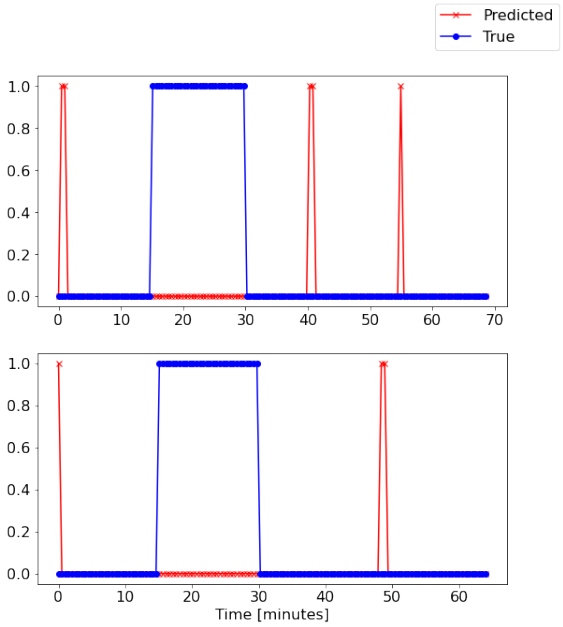


**Figure 30:** True vs predicted response vector for the test meals, for the final classifier with 60s time segment. The graph on the top and bottom are respectively the first and second test meals.

## A.2 Classification using bowel and swallowing sound recordings

Two data splitting modalities were tested for this classification system, in the first one the data was split randomly, while in the second one the data were split based on the subject of the recording.

### A.2.1 Splitting training and test data randomly

This system is the same as the one described in the main part of the report. For this system first the TPR and FPR for all parameter combinations for the average of all training meals are plotted, Figure 31. Similar TPR and FPR were observed for the validation meals average, Figure 15.

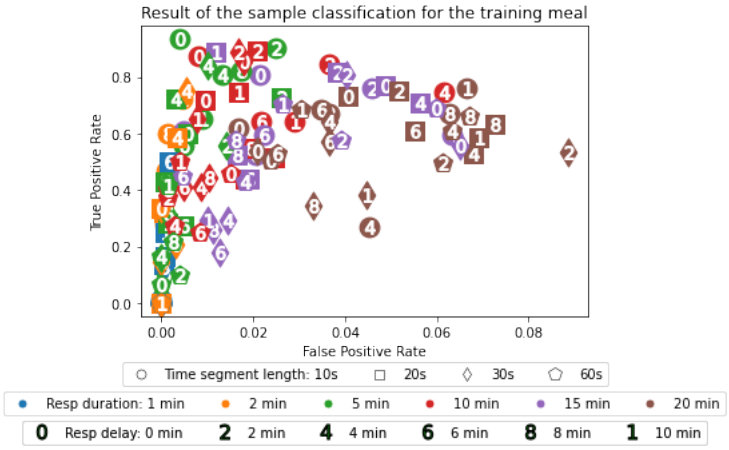


**Figure 31:** True positive rate vs false positive rate for the training meals in step 1. Each marker represents the average of all training meals for a given parameter combination.

Since the predicted and true response vector for the 10s and 20s segment was already plotted, the 30s and 60s time segments were plotted here, Figure 32 and 33. For both classifiers, only TP was counted, with an average meal detection time of 4 min. Unlike the classifiers with the short time segments, these classifiers labeled more of the noise in the recordings as a meal. It seemed that the longer the time segment the more events are captured, both meal and unwanted events, such as noise are captured in the calculated features. This in turn affects the training of the SVM classifiers.

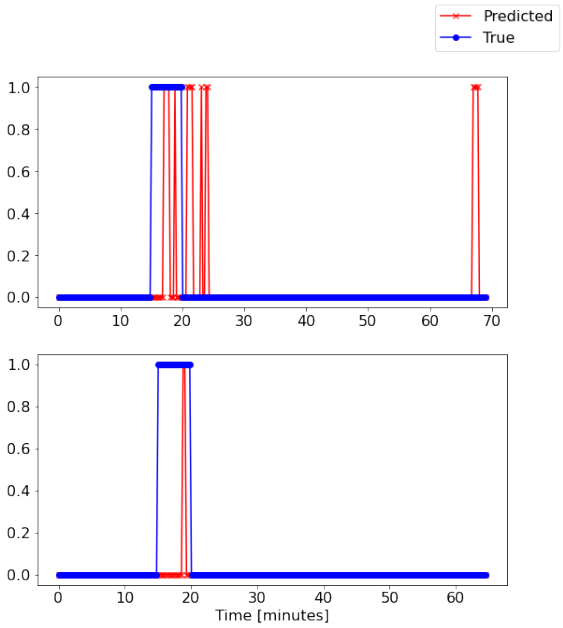


**Figure 32:** True vs predicted response vector for the test meals, for the final classifier with 30s time segment. The graph on the top and bottom are respectively the first and second test meals.

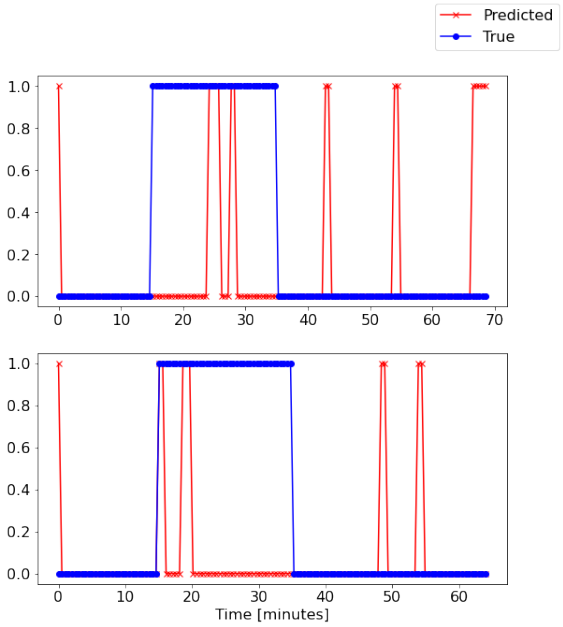


**Figure 33:** True vs predicted response vector for the test meals, for the final classifier with 60s time segment. The graph on the top and bottom are respectively the first and second test meals.

### A.2.2 Splitting training and test data based on the subjects

This classification system was briefly mentioned in the main part of the report, but none of the results were included as they were similar to the classification system above. This system was trained using three meals from a single subject and then tested using two meals from another subject, to test whether the system is subject-dependent. The classifier performance did improve with the training data as can be seen in the TPR and FPR for all parameter combinations for the average of all training meals plot, Figure 34. The TPR and FPR were a little higher for the training meals, compared to the preceding system, however, similar results were observed when the TPR and FPR were plotted for the average of all LOOCV runs, Figure 35.

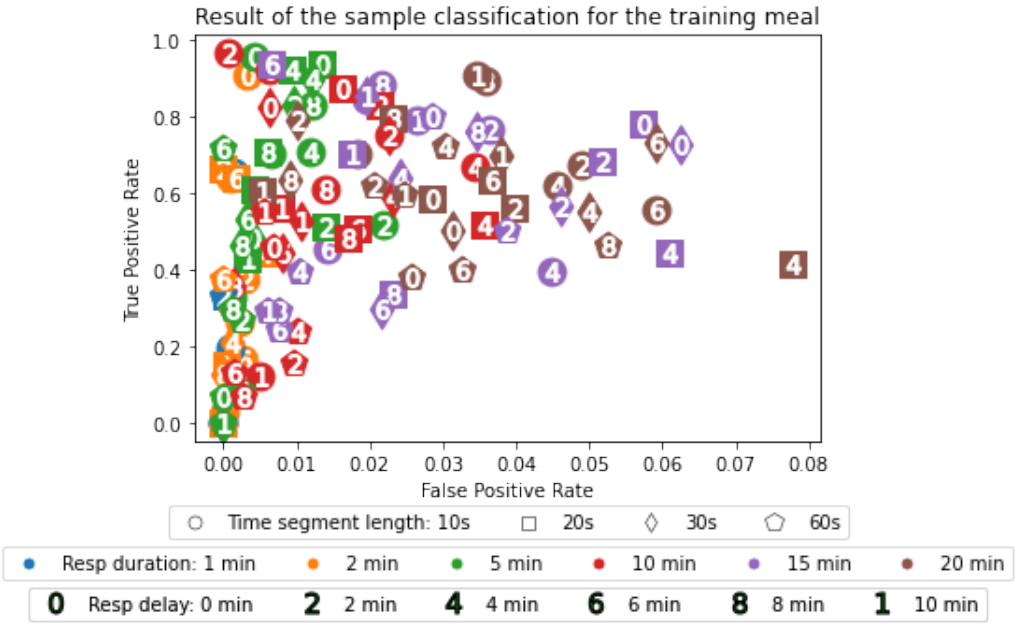


**Figure 34:** True positive rate vs false positive rate for the training meals in step 1. Each marker represents the average of all training meals for a given parameter combination.

Other than the difference in TPR and FPR plot for the training data, the classifiers performed exactly alike. To ensure that the performance is not different, The TP and FP count for the test meal was plotted, Figure 36. For all test meals, only TP's were counted. Based on these findings, the system was assumed to be subject-independent. Nevertheless, more data is required to make a genuine assessment of the system's subject dependency.

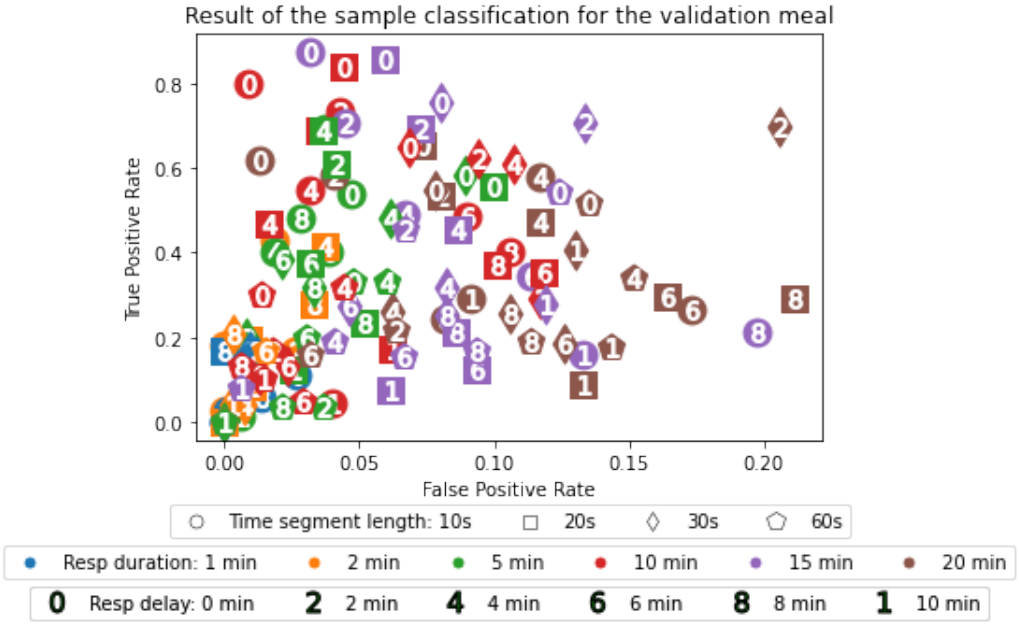


**Figure 35:** True positive rate vs false positive rate for the validation meals in step 1. Each marker represents the average of all LOOCV runs for a given parameter combination.

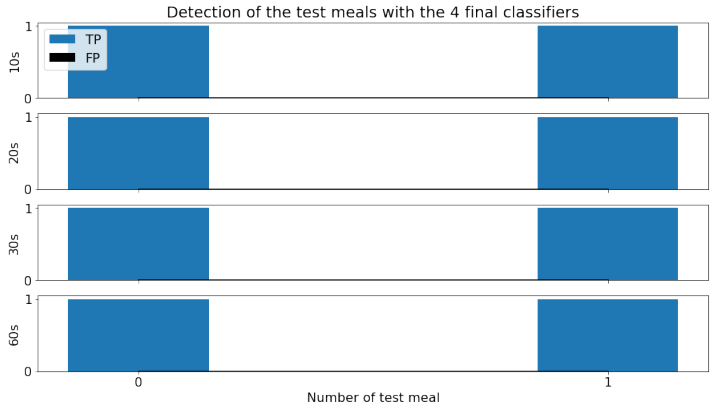


**Figure 36:** True positive and false positive meal detection for each test meal, for the four final classifiers.

## A.3 Classification using only swallowing features

The final classifiers in the previous classification system used only swallowing features. The system proved to be dependent on the swallowing features, that's why a classification system using only swallowing sounds was tested. It was built to see if there was at all a need for bowel sound recordings for meal detection. The TPR and FPR for all parameter combinations for the average of both training meals and all LOOCV runs are plotted in Figure 37 and 38.

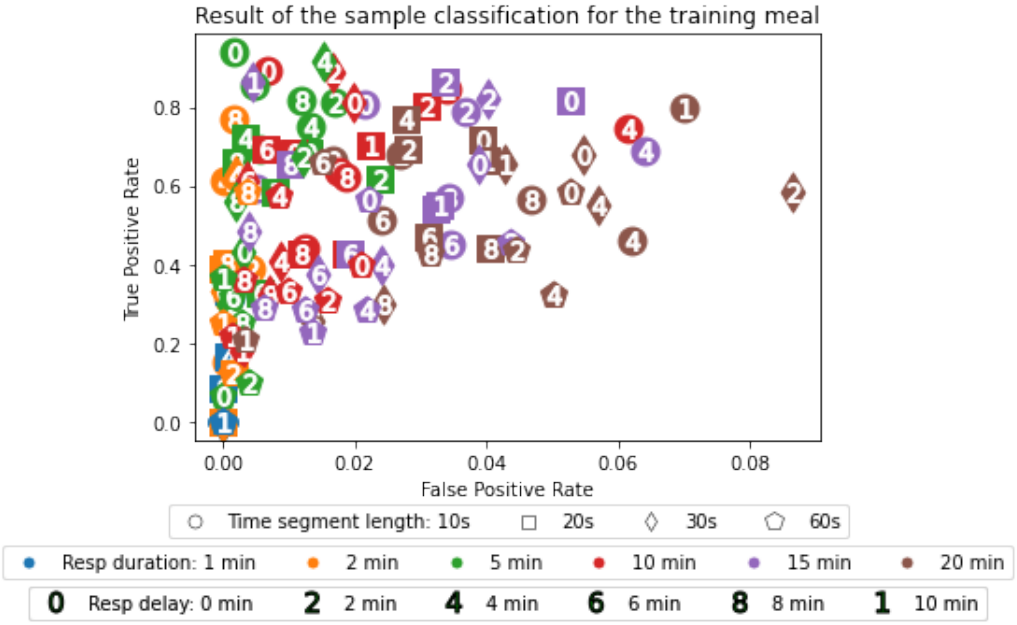


**Figure 37:** True positive rate vs false positive rate for the training meals in step 1. Each marker represents the average of all training meals for a given parameter combination.

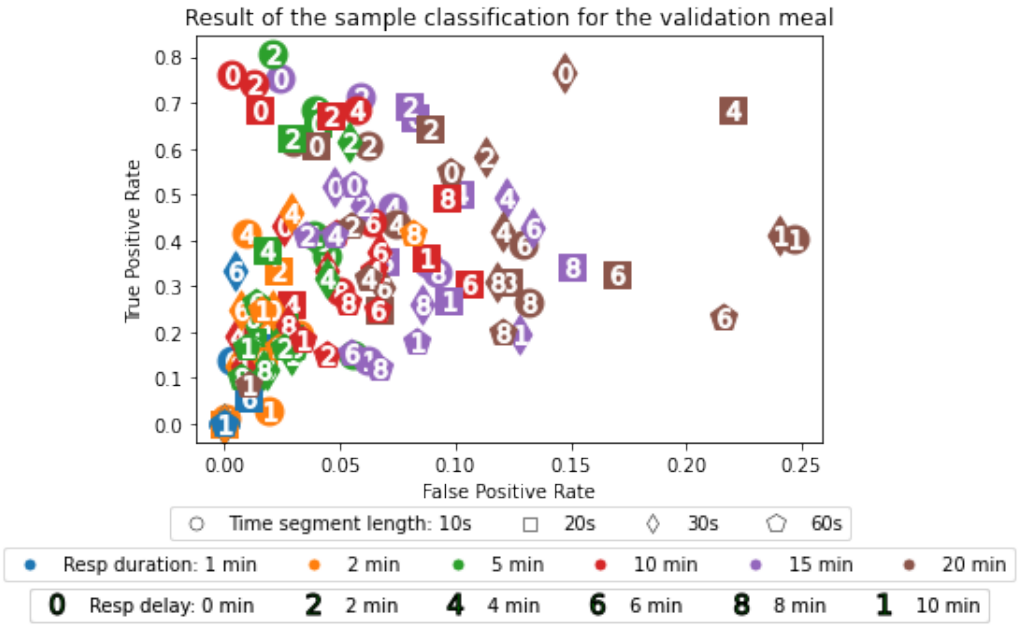


**Figure 38:** True positive rate vs false positive rate for the validation meals in step 1. Each marker represents the average of all LOOCV runs for a given parameter combination.

The four final classifiers TP and FP plot is shown in Figure 39. This classifier had similar results as the previous classifier with all features. These results confirmed that there is no need for bowel sounds, due to the way the different features are selected and incorporated into the system. The features from different recordings should be incorporated into the system in a better way, this was included in the future work section.

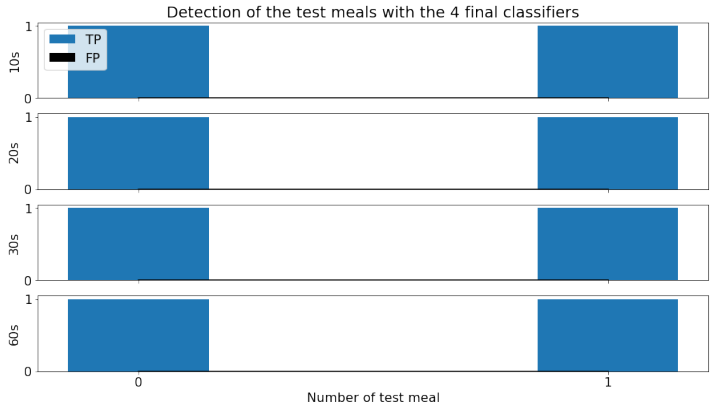


**Figure 39:** True positive and false positive meal detection for each test meal, for the four final classifiers.

## A.4 Classification using data augmented with noise

Two classification systems were built for testing data augmented with noise. The plots that were not included in the main part of the report are included in this part.

### A.4.1 Training without data augmented with noise

The TPR and FPR for all parameter combinations for the average of all training meals in this system is the same as the one shown in the previous section, Figure 37. As described in the main part of the report the classifiers for the 30s and 60s time segments were not able to predict anything, since the SVM classifiers were not provided with good training data. For that reason, there was no meal labeling in the predicted response vector, as could be seen in Figure 40 and 41.

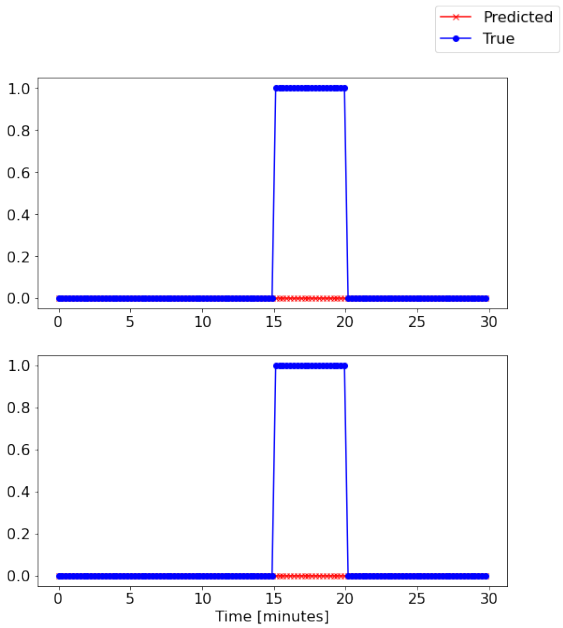


**Figure 40:** True vs predicted response vector for the test meals, for the final classifier with 30s time segment. The graph on the top and bottom are respectively the first and second test meals.

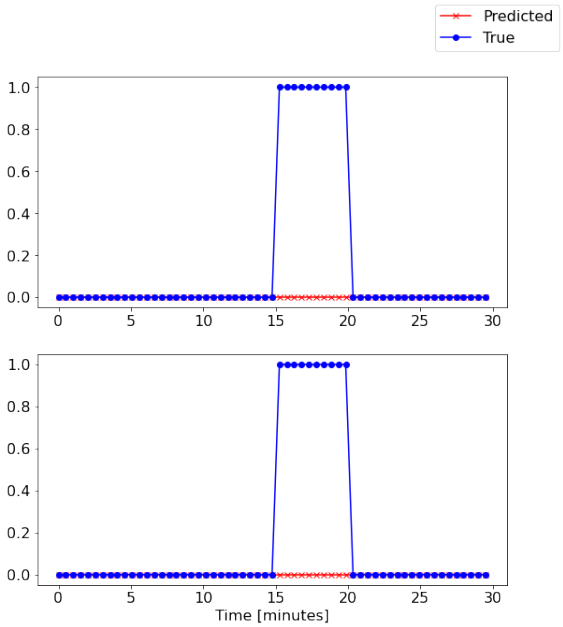


**Figure 41:** True vs predicted response vector for the test meals, for the final classifier with 60s time segment. The graph on the top and bottom are respectively the first and second test meals.

### A.4.2 Training with and without data augmented with noise

When the training and validation set included data augmented with noise, the performance of the classifier with regard to TPR and FPR for the average over the training meals was poor, Figure 42, just as it was for the average over the LOOCV runs, Figure 24.

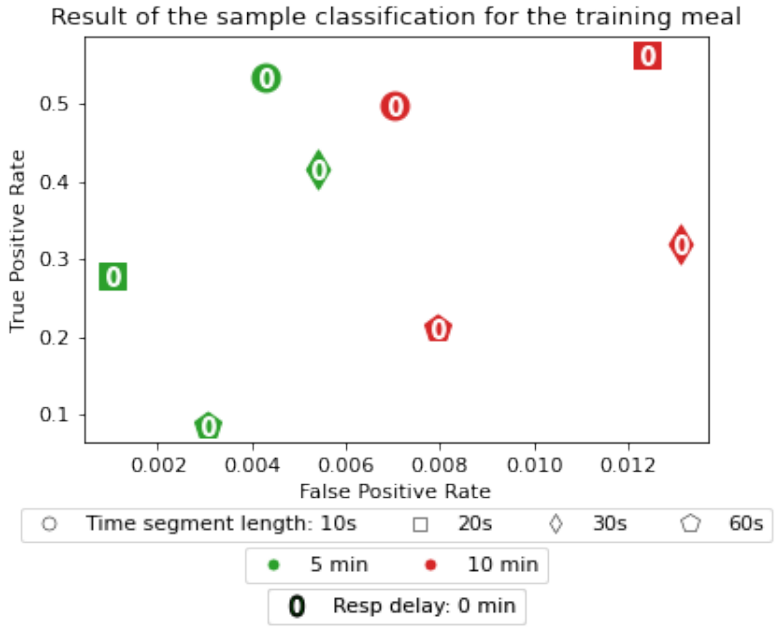


**Figure 42:** True positive rate vs false positive rate for the training meals in step 1. Each marker represents the average of all training meals for a given parameter combination.

However, unlike the final classifiers for the 30s and 60s time segments in the preceding system (see Appendix A.4.1), the classifiers in this system were able to predict a meal onset. This could be seen in the predicted and true response vector plot, Figure 43 and 44, there was a meal prediction for both classifiers. Similar to the 10s and 20s time segments, the classifiers struggled with the first test meal, since the meal had also reading (noise) during the meal. Thus the 30s time segment classifier was not able to detect a meal for the first meal. The average meal detection time for the three meals that were detected is 2 min. Unlike the classifiers with the 10s and 20s time segments, these classifiers did not label the talking as a meal, which is why the testing of different segment lengths was included in future works.

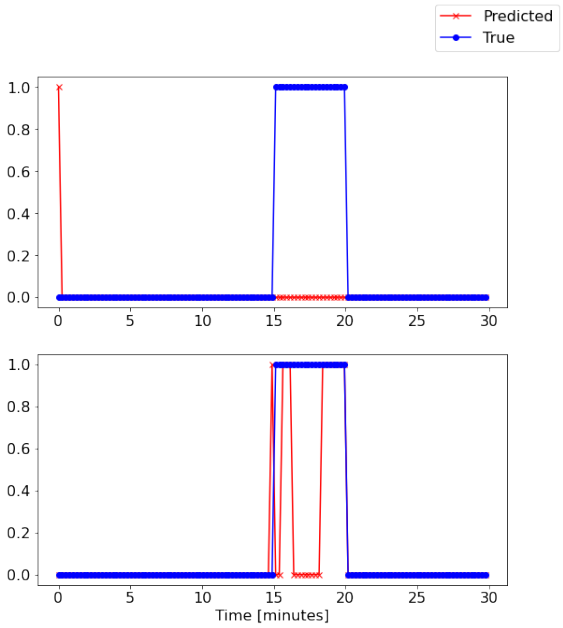


**Figure 43:** True vs predicted response vector for the test meals, for the final classifier with 30s time segment. The graph on the top and bottom are respectively the first and second test meals.

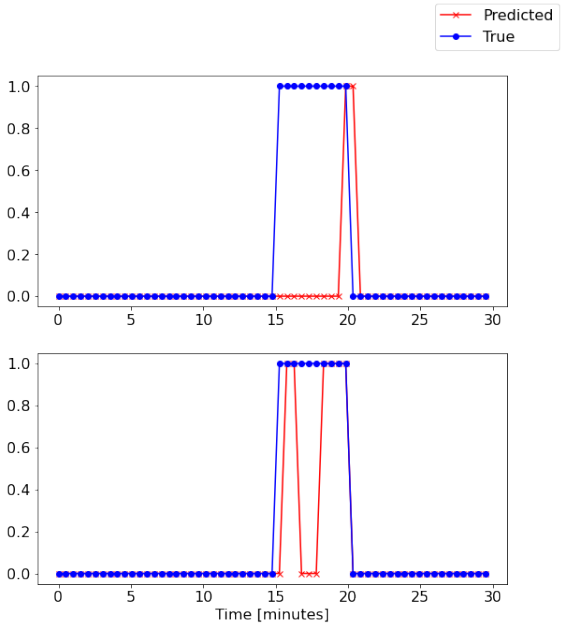


**Figure 44:** True vs predicted response vector for the test meals, for the final classifier with 60s time segment. The graph on the top and bottom are respectively the first and second test meals.

# B Zip file

This thesis comes with a zip file that contains all the code used for this project, in addition to all the features that were used in this project. The code is included in the file "Code", while the features are included in the file "Features" as shown in Figure 45. Konstanze Kölle's paper [6] is also included in the zip file as "Konstanze thesis.pdf".

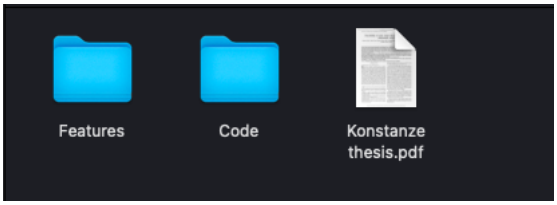


**Figure 45:** Included zip file content