Master's thesis

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

Elise Søiland

Early Meal Detection Based on Abdominal Sound

Master's thesis in Cybernetics and Robotics
Supervisor: Anders Lyngvi Fougner
June 2019

NTNU
Norwegian University of Science and Technology

ARTIFICIAL PANCREAS TRONDHEIM

Elise Søiland

Early Meal Detection Based on Abdominal Sound

Master's thesis in Cybernetics and Robotics
Supervisor: Anders Lyngvi Fougner
June 2019

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Engineering Cybernetics

NTNU
Kunnskap for en bedre verden

NTNU
Norges teknisk-naturvitenskapelige universitet

Fakultet for informasjonsteknologi, matematikk og elektroteknikk
Institutt for teknisk kybernetikk

MASTEROPPGAVE

Kandidatens navn: Elise Søiland

Fag: TTK4900 Teknisk Kybernetikk, Masteroppgave

Oppgavens tittel (norsk): **Tidlig måltidsdeteksjon basert på buklyder.**

Oppgavens tittel (engelsk): **Early meal detection based on abdominal sound.**

Oppgavens tekst:

This thesis is affiliated with Artificial Pancreas Trondheim (APT).
In this work, the student will use advanced signal processing and pattern recognition (or similar methods) for early meal detection based on recorded abdominal sounds. This will be useful in a system for closed-loop glucose control in diabetes. Specifically, the student will perform the following tasks:

1. Find and read literature regarding sound processing, while focusing on signal features and classification methods especially useful for sound-based classification.
2. Get familiar with APT's existing data set with abdominal sounds from meals in healthy people, and the signal processing methods applied in:
   a. Two recent paper manuscripts from Konstanze Kölle (submitted in Aug 2018 and Sep 2018).
   b. The MSc thesis by Kaja Kvello (June 2018).
3. Consider if the data set is suitable for further work or whether you need to collect more or different data.
   a. You may put weight on how the data set is "tagged" (with respect to having a reference for when bowel sounds actually occur) – and whether the data set needs to be tagged at all.
   b. If more data is needed/wanted, apply the regional ethical committee (REK) before starting the data collection.
4. Evaluate the suitability of the signal processing methods and classification algorithms found in 1 and 2 above, and implement the methods found to be most relevant.
5. Evaluate the methods on the data set.

Oppstart: 7. januar 2019

Innleveringsfrist: 3. juni 2019

Utført ved Institutt for teknisk kybernetikk

Veileder: Anders Lyngvi Fougner
Biveileder: Sunilkumar Telagam Setti

Trondheim, 10. desember 2018

Preface

This thesis is submitted as a part of the degree Master of Science in Engineering Cybernetics at the Norwegian University of Science and Technology (NTNU).

The problem description was given by the Artificial Pancreas Trondheim (APT) research group, which revealed the potential benefit of a sound based meal detection system as a part of an automatic insulin pump. This thesis is based on previous work by APT members (PhD candidate Konstanze Kölle and MSc student Kaja Kvello). The APT group also provided a data set containing 10 recorded meals of 10 different subjects and 17 recorded meals of one subject as .wav-files. Throughout the semester, I attended regular meetings which provided a medical and overall understanding of their project. My final results were presented for the group and they shared ideas on how to include a meal detection system in an artificial pancreas.

The Department of Electronic Systems provided the following equipment for recordings of a new data set:

- Sennheiser MKE2 P-C condenser microphone attached in the center of a chest-piece of a classical stethoscope
- Roland Octa-Capture sound card
- Access to an echo-free room

My supervisor, Anders Lyngvi Fougner, provided relevant background literature in the initial stages of the work. He has shared his expertise on diabetes, signal processing techniques and classification methods at weekly supervision meetings. Through the process, he has contributed with thoughts and discussion on my various ideas, which made me reflect and evaluate the methods considered.

Sunilkumar Telagam Setti, my co-supervisor, shared ideas of relevant signal processing and feature extraction techniques, especially those commonly used in sound analysis(e.g mel-frequency cepstral coefficients, wavelet transform, and entropy). He also contributed with thoughts on report structure and formulations during the completion of my thesis.

The final choice of feature extraction and classification method were done by me, as well as the implementation in MATLAB.

I want to thank Anders Lyngvi Fougner and Sunilkumar Telagam Setti for excellent supervision and for always answering my list of questions on all supervision meetings. Also, I hope that they can publish the results of this project as a conference paper. Lastly, I want to thank the Department of Electronic Systems for providing equipment and the echo-free room.

Abstract

This thesis describes the development and implementation of a sound based meal detection algorithm. The long term aim is to include such an algorithm in an automatic insulin pump, also called an artificial pancreas. Insulin pumps are currently used by many diabetic patients and require manual dosing at mealtime. Including a meal detection algorithm will be an important step towards an automatic insulin pump. Previously, continuous blood glucose monitoring has been used to detect meals, but as it takes time before the blood sugar rises after a meal, the detection can be too late. In this thesis, sound recordings from a stethoscope attached to the abdominal wall are used to achieve faster meal detection.

An existing data set with recorded meals from one person, and four new recordings from a second subject, were used for testing the system. The algorithm developed, consists of a combination of preprocessing audio signals, feature extraction, signal processing techniques, and pattern recognition to classify audio segments as either meal or no meal. Two different sets of features were tested as input to a neural network for classification. A threshold was set for the output of the net, and if four consecutive segments exceeded the threshold value, a meal was detected.

Both sets of features gave promising results for meal detection, but only one set provided good results for both subjects. Generally, meals were detected after only 1-4 minutes with a few false positive results. The detection was significantly faster than meal detection based on blood glucose measurements. Suggestions for further improvements and important steps towards the integration of a meal detection system in an artificial pancreas are also being discussed.

Sammendrag

Denne oppgaven beskriver utvikling og implementasjon av en algoritme for måltidsdeteksjon basert på lyd fra magen. Målet er at en slik algoritme skal være en del av en automatisk insulinpumpe, også kalt en kunstig bukspyttkjertel. Insulinpumper brukes i dag av mange diabetespasienter og krever manuell dosering ved måltid. Å inkludere en algoritme for deteksjon av måltid vil være et viktig steg i arbeidet mot en kunstig bukspyttkjertel. Tidligere har kontinuerlig blodsukkermåling blitt brukt for å detektere måltid, men ettersom det tar tid før blodsukkeret stiger etter et måltid, vil deteksjonen ofte komme for sent. I denne oppgaven er lydopptak fra et stetoskop på magen brukt for å oppnå raskere måltidsdeteksjon.

Et eksisterende datasett med opptak fra én person ble brukt, samt fire nye opptak ble gjort på en ny forsøksperson. Algoritmen ble til slutt testet på 4 måltider fra hver av de to forsøkspersonene. Algoritmen utviklet i denne oppgaven består av en kombinasjon av preprosessering av lydsignal, signalbehandlingsteknikker for egenskapsuttrekking og mønstergjenkjenning for å klassifisere segmenter med lyd som enten måltid eller ingen måltid. To ulike sett med egenskaper av lydsignalene ble testet som inngang til nevralt nettverk for klassifisering. En grense ble satt for utgangen av nettverket, og dersom fire segmenter på rad oversteg denne verdien, ble et måltid markert.

Begge settene med egenskaper gav lovende resultat for måltidsdeteksjon, men bare det ene settet virket like bra på lydopptak fra begge forsøkspersonene. Generelt blir måltidene detektert etter bare 1-4 minutter med noen få falske positive resultat. Deteksjonen var generelt betydelig raskere enn måltidsdeteksjon fra blodsukkermåling. Forslag til forbedring av deteksjonsalgoritmen, samt viktige steg som gjenstår før en slik algoritme kan inkluderes i en kunstig bukspyttkjertel, er også diskutert.

Contents

List of Figures

List of Tables

Nomenclature

ANN	Artificial Neural Network
AP	Artificial Pancreas
APT	Artificial Pancreas Trondheim
BG	Blood Glucose
CGM	Continuous Glucose Monitoring
CWT	Continuous Wavelet Transform
DCT	Discrete Cosine Transform
DWT	Discrete Wavelet Transform
HMM	Hidden Markov Model
IIR	Infinite Impulse Response
IMF-FD	Intrinsic Mode Function-fractal Dimension
LTI	Linear time-invariant
MEMD	Multivariate Empirical Mode Decomposition
MFCCs	Mel-frequency cepstral coefficients
NN	Neural Network
PSD	Power Spectral Density
STFT	Short Time Fourier Transform
SVM	Support Vector Machine

1 Introduction

About 245 000 people in Norway have a diabetes diagnosis. Nearly 28 000 of them are diagnosed with type 1 diabetes[1]. These patients need external insulin administration to control their blood sugar level, which requires multiple injections of insulin daily or use of an insulin infusion pump.

The research group Artificial Pancreas Trondheim (APT) aims to develop an automatic glucose control system, called an artificial pancreas. The goal is to mimic the natural glucose control of the body without depending on daily information and administration from the user.

1.1 Glucose regulation

In the human body, two hormones are controlling the blood glucose (BG) level; insulin and glucagon. Both are produced and secreted by the pancreas. After a meal, the BG level will increase due to carbohydrates in the food. Insulin is then released into the bloodstream to promote absorption of glucose from the blood into skeletal muscle and other tissues in the body. The liver stores excess glucose as glycogen, which can be released if the blood sugar drops. The rest of the glucose is converted to fatty acids which are stored as body fat called adipose tissues[2].

If the BG level drops, glucagon is secreted from the pancreas. Glucagon acts opposite of insulin to increase the BG level. This hormone mobilizes glucose, fatty acids and amino acids stored in tissues to the blood circulation. Primary, glucose is released from the liver by converting glycogen back to glucose[2].

1.2 Diabetes

Diabetes, formally called *Diabetes Mellitus*, is caused due to lack of production or absorption of insulin[2].

Diabetes is mainly divided into two types; type1 and type2. Type 2 is the most common type. It is caused by too little production of insulin, often combined with insulin resistance. This disease is most often treated by changing the patient's lifestyle and diet. Reducing the carbohydrate intake and increase the patient's activity level, is often sufficient for reducing the insulin resistance and the level of glucose in the blood[2][3]. Some patients will, however, experience decreased insulin production and may need supplemental insulin[4].

Type 1 diabetes, often called insulin-dependent diabetes, is caused by a dysfunctional pancreas. The insulin-producing beta cells are destroyed, resulting in no insulin production, and a need for external insulin administration[2][3]. Insulin has to be injected as stomach enzymes will break down the insulin if it is taken orally. Usually, insulin is injected in the subcutaneous tissue right under the skin.

1.3 Current insulin administration

Insulin treatment aims to mimic the body's natural secretion of insulin. Insulin is available in rapid-, short-, intermediate-, and long-acting types, or as a mix. Which type one wants to use is dependent on individuals' preferred insulin injection technique and lifestyle[4].

The conventional insulin administration involves injection with syringes or pens with needle and cartridge. The injections are done in the subcutaneous tissue, typically in the upper arm, the thigh or the abdomen. The abdomen is often preferred as it has the fastest rate of absorption. For most type 1 diabetes patients, three or more insulin injections daily is required to meet glycemic goals. The dosage should be based on BG measurements as the BG level is hard to predict due to day-to-day variations. The insulin absorption rate, insulin sensitivity, exercise, stress, food absorption, and hormonal changes are some of the factors contributing to the variation in BG level[4].

In the later years, insulin pumps with continuous subcutaneous insulin infusion, have become more popular. An insulin pump is a wearable electromechanical pump, including processing module, batteries and insulin reservoir. A disposable tube with a cannula connected ensures the delivery of insulin under the skin at the abdomen. Insulin is infused constantly with a preselected rate. Besides, boluses activated by the user has to be infused at mealtimes. Insulin pump therapy has been shown to provide better glycemic control, especially for reducing hypoglycemia[5].

Which treatment method to use is an individual choice. It depends on the individual's symptoms, lifestyle, and preferences. The treatment has to be individualized to each individual to ensure the best glycemic control and self-monitoring.

1.4 Importance of tight BG control

The lack of insulin production or low insulin sensitivity puts persons with diabetes in the risk of elevated BG level, called hyperglycemia. Symptoms of hyperglycemia may not become noticeable until reaching very high values (>11.1mmol/l or >200mg/dl) or having high BG level over a longer period. Health problems caused by hyperglycemia are often related to the heart, blood vessels, eyes, kidneys, and nerves[2]. The combination of reduced blood flow and nerve damage in the feet increases the risk for foot ulcers as the nerve damage reduces the ability to feel pain. This can lead to infection or a need for limb amputation. Long-Term accumulated damage to the small blood vessels in the retina can cause blindness[3]. Actually, 2.6% of global blindness in 2010 was attributed to diabetes[6].

Complications with elevated BG level can not only occur in diabetes patients, but also critically ill patients. Both hyperglycemia and insulin resistance are common among this patient group. Elevated BG level effects wound healing and increase the risk of infections. Studies have shown intensive insulin therapy to be effective. In a study done by Berghe et al. in 2001, the BG level of critically ill patients in the surgical care unit was kept within the range of 4.44 and 6.1 mmol/l (80 - 110mg/dl). The morbidity and mortality were significantly reduced compared to the patients who received conventional insulin therapy[7].

Short term complications related to BG regulation are mainly caused by a too low BG level, called hypoglycemia. Type 1 diabetes patients are dependent on external insulin administration, putting this patient group in the risk of injecting a too high dose of insulin. Because insulin lowers the blood sugar, a high dose can lower the BG level under 70mg/dl (3.88mmol/l), which can have serious consequences. Hypoglycemia may result in trouble talking, clumsiness, loss of consciousness, or even death[2]. Tight glucose control is crucial to avoid low BG levels.

1.5 Abdominal sound

Sounds from the abdomen, or gastrointestinal sound, are normally caused by transport of food, liquids, and gas in the intestines during digestion. Sounds from the abdomen can often be heard by ourselves as gurgling or rumbling. Auscultation of the abdomen is frequently used by health professionals to detect abnormal activity in the gastrointestinal tract. Recorded signals of abdominal sound consist of bursts of energy which constitute a non-stationary, transient signal.

The term bowel sound (BS) is used to describe loose successions or clusters of sudden bursts in the sound signal[8]. Bowel activity has been described as the occurrence rate of bowel sounds. A peak in bowel activity has been seen immediately after a meal intake[9]. The increased occurrence of bowel sounds has been used as a feature for meal detection[10].

The frequency of bowel sounds have been found to lie between 50Hz and 1500Hz[11][12]. Dalle et al. reported that no bowel sound exceeded 3000Hz. The energy of signals is often used as a measurement of activity and has been used to localize the frequency band of bowel sounds. Ranta et al. reported that only 2% of the energy is located beyond 500Hz, meaning that the main part of the energy was found from 100Hz to 500Hz[13].

1.6 Meal detection

The main motivation for the development of meal detection systems is in general to prevent eating-related diseases such as diabetes type 2, obesity, and cardiovascular disease. Continuous food-intake monitoring can provide more reliable user information, resulting in more effective health intervention services[14]. However, meal detection has also been shown to be an important sep towards the development of an automatic insulin pump[15].

Researchers have investigated a wide spectre of wearble sensors to obtain accurate noninvassive meal detection. Wrist bands have been found to count "bites" with ca. 80% accuracy by tracking wrist motion[16]. Audio-based detection has also been investigated by using a smartwatch device's build-in microphone. By applying signal-processing techniques chewing and swallowing were identified with high accuracy[17]. Microphones has also been attached to the subject's neck to record acoustic signals during eating[18]. Piezoelectric sensors have also been attached on the jaw, below ear and on the neck[19]. Piezoelectric sensors and microphones have also been attached to the abdomen to identify bowel sounds and meal related sound[9][15][20].

2 Theory

In general, meal detection is a signal classification problem. The major steps involved in such classification problems are preprocessing, feature representation and classification. The objective of the feature representation is to obtain discriminative information from the signals, while classifiers such as support vector machine, nearest neighbor classifier and neural networks utilize this information for identifying the category of the signals. The performance of a signal classification approach depends crucially on the combination of the aforementioned steps. More importantly, it depends crucially on feature representation.

2.1 Preprocessing

Decimation Decimation is one way of reducing the sampling frequency. Aliasing is prevented by lowpass filtering before the result is downsampled. A lower sampling frequency results in fewer samples which require less processing and memory resources[21].

Interpolation Interpolation is the opposite of decimation. The original sampling rate is being increased by inserting zero-valued samples between original samples. The primary motivation to interpolate is to increase the sampling rate at the output of one system so that another system, operating at a higher sampling rate, can input the signal[21].

Linear time-invariant filters When dealing with recorded audio signals, the signal consists of many frequencies. Often, some of the frequencies are noise signals which are not of interest. Different filters allow us to filter out certain signals on the basis of frequency. These filters are useful tools for noise reduction and feature extraction as the frequencies of interest can be extracted.

[21](p. 330) describes a linear time-invariant filter as a system which "modifies the input signal spectrum $X(\omega)$ according to its frequency response $H(\omega)$ to yield an output signal with spectrum $Y(\omega) = H(\omega)Y(\omega)$". $H(\omega)$ acts like a weight function which weights different frequency components in the signal. Filters generally named after their frequency-domain characteristics. The following sections will describe some commonly used LTI filters.

The probably most known LTI-filters are lowpass and highpass filters. An ideal version of such filter will only allow the signal with frequencies within the filter passband to pass. Outside the passband, the filter will ideally have a zero gain. The *cutoff frequency* defines the passband. For low pass filters, the signal components with frequencies under the cutoff will pass. Highpass filters have the opposite function and pass signals of higher frequency[21].

2.2 Feature representation

The raw data can rarely be fed directly into a classification algorithm. Most often, a lot of signal processing is done to identify the measurable quantities that make the classes distinct from each other. These quantities are often described as the *features*. Some selected techniques for signal feature representation in the time, frequency, and time-frequency domain will be described in the following subchapters.

An extensive set of features will directly cause a large set of classifier parameters. Keeping the number of features as small as possible will help with designing classifiers with good generalizations capabilities[22]. Only the features which correlate with the meal should be selected. There exist different algorithms, e.g. principal component analysis (PCA), which helps to select the features which correlate the most with the classification problem. For less complex cases, the features can be plotted, and the ones separating the different classes should be selected.

2.2.1 Time domain features

Representing a signal in with time domain features is often beneficial as the signal changes over time. These features are often easy to implement as the features can be extracted directly from the signal. A major disadvantage comes from the non-stationary property of sound signals, changing the statistical properties over time, but time domain features assume the data as stationary signals. In general, time domain features are calculated form signal amplitude values, meaning that interference acquired through the recording can cause another challenge by using these kinds of features[23].

Entropy Entropy describes the average rate of information produced in a signal. A low probability event will cause a high entropy, while an event with high probability will result in low entropy. The entropy is computed from the normalized histogram using Equation 1. Commonly, there are several entropy types such as Shannon, log energy, sure, threshold, etc.[24][25].

$$H = -\sum_{i=0}^{N-1} p_i log(p_i) \tag{1}$$

Zero-crossing Zero-crossing is the number of times the amplitude value of the signal crosses the zero amplitude level. This is a representation of frequency information of the signal in time domain[23].

2.2.2 Frequency domain features

In signal processing, it can be advantageous to represent the signals in the frequency domain. As the rain separates the light into colors, making the rainbow, the light is separated based on different frequency component in the light. A signal, for example, a sound signal from a guitar chord, can be represented in the frequency domain to tell which frequency components, or tones, the signal consists of.

Fourier transform The Fourier transform is a well known mathematical tool, used for representing signals in the *frequency domain*. The signals are decomposed in terms of sinusoidal (or complex exponential) components, representing the original signal's frequencies as magnitude components. The Fourier transform of a continuous signal can be seen in Equation 2.

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

The Fourier transform of a discrete signal can be calculated using Equation 3, where W_N is defined as in Equation 4[21].

$$x(k) = \sum_{n=0}^{N-1} x(n)W_N^{kn} \tag{3}$$

$$W_N = e^{-j2\pi/N} \tag{4}$$

2.2.3 Time-frequency domain features

The obvious advantageous with these features is that they represent the signal in both the time and frequency domain. A time-frequency representation enables to set the time when a change in frequency response occurs.

Short time Fourier transform The short time Fourier transform(STFT) extracts data segments at regular intervals using a time-limited window. A discrete Fourier transform is carried out for each frame of the signal. This method provides the local time-frequency behavior of the signal, which makes it possible to tell how the frequency characteristics are changing with time[26].

The wavelet transform In later years, Wavelet transform has become a more popular alternative to the STFT for processing no stationary signals[27][28]. Similar to the STFT, wavelet transform also maps a time function in a two-dimensional function of τ and β (instead of ω and τ). τ is called the translation parameter and is the translation of the wavelet function along the time axis. β is called the scaling parameter and scales the function by compressing or stretching it[28][29]. As STFT uses same window size, the wavelet uses a flexible size for dissimilar frequency bands. Wider windows are used for lower frequency, and narrow windows are used for high frequency. The varying window size will yield higher time resolution for high-frequency signals[28].

A wavelet is an oscillation wave-like function which starts at zero, increases, and goes to zero again. There are various versions of wavelets, but they are all scaled and shifted versions of the "mother wavelet" function [28]. The mother wavelet, Equation 5, is centered at $t = 0$ with an average value of zero. The Continuous Wavelet Transform (CWT) for a signal $s(t)$, where $\psi \in R^2$, can be defined as in Equation 6.

$$\psi(t) = \frac{1}{\sqrt{\beta}}\psi\Big(\frac{t-\tau}{\beta}\Big) \tag{5}$$

$$CWT(\beta,\tau) = \frac{1}{\sqrt{\beta}}\int s(t)\psi\Big(\frac{t-\tau}{\beta}\Big)dt \tag{6}$$

β, the scaling parameter, gives the size of the wavelet, while τ gives the location. The discrete version, Discrete Wavelet Transform (DWT), can be defined as in Equation 7. The discrete wavelet, ψ_k, can be a sampled version of the continuous wavelet [29].

$$DWT(m,n) = 2^{-\frac{m}{2}} \sum_k s(k)\psi(2^{-m}k - n) \quad (7)$$

The DWT is a filtering process with two filters; highpass (wavelet) and low pass (scaling). The signal can be decomposed into a high pass component and a low pass component in the first level of decomposition. The second level of decomposition is done on the low pass components from the first stage. The frequency of the signal is downsampled with a factor of 2 for each decomposition level[29][30].

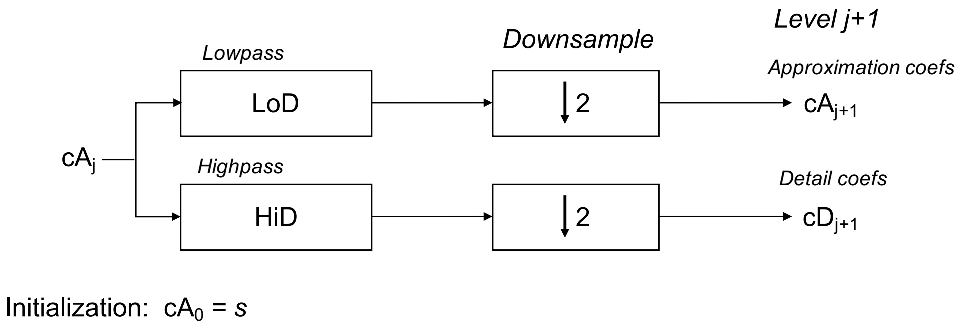


Figure 1: One-Dimensional DWT of signal s.

Mel-frequency cepstral coefficients Mel-frequency cepstral coefficients (MFCCs) are *cepstral* coefficients scaled on the *mel scale*. The mel-scale was constructed to reflect the relationship between the perceived magnitude of pitch and the physical measure of it. Equal steps in mels should be equal in subjective size throughout audible frequency range[31]. The cepstrum is defined as a power spectrum of the logarithm of the power spectrum of a signal[32]. Meaning that MFCCs can be described as a spectrum of a logarithmic spectrum mapped to the mel-scale.

MFCCs are frequently used as features in speech recognition due to the scaling based on the perception of the human ear[33]. It has also been frequently used on biomedical signals and is one of the most common features used for computer-based lung sound analysis[34].

For the calculations, the signal has to be framed into a decided window length with a chosen overlap. The segment length and overlap have been shown to impact the MFCCs directly, and different lengths might be tested during feature selection phase[33]. The coefficients can then be derived by a set of signal processing tools which is applied to each segment of the data, giving one set of coefficients for each window.

The first step is to take the discrete Fourier transform and discard the phase information as perceptual studies have shown that the amplitude is much more important. Next, the logarithm is taken of the amplitude spectrum because the perception of loudness has been found to be approximately logarithmic. Then, the signal is wrapped to the mel-scale before the last processing method, discrete cosine transform(DCT), is applied. The mel-filter bank has overlapping segments, causing a high correlation between each other. The DCT is therefore applied to decorrelate the signal[35]. The entire MFCC calculation process is illustrated in Figure 2.

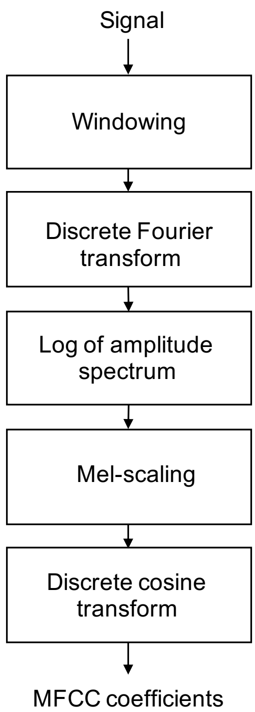


Figure 2: MFCC flowchart

2.3 Classification methods

As more industrial processes are being automatized, the field of machine learning has become important. Pattern recognition methods are integrated into complex "intelligent" systems, designed for decision making. The machine learning process is often challenging, as often a large amount of complex data is being used to find meaningful patterns. The whole process is more complicated than applying an algorithm on a set of data. It is an iterative process of getting the data needed, pre-process the data, and extract well-structured data which can be fed into a classification algorithm. This chapter will focus on the process of making an algorithm classify a set of data.

2.3.1 Training, validation and test set

The first step is to identify, choose and get the data needed for the classification algorithm. In order to train and test the classifier, a test set might be required. A test set consists of labeled data which can be used to train a classifier and to check whether it classifies correctly or not. The test set is often divided into training, validation and test data.

The training set is used to train the classification algorithm. While the validation data is used as a guide for stopping the training process in time. If the classifier is trained too much, it might have become overtrained and cannot be generalized to new data. The training process can instead be stopped when the validation set gets proper results. The test set is excluded from the training and validation process. It is used to control the accuracy of the classifier by comparing the classification with the previously marked class. If the classification is accurate enough, it can be implemented and used in real applications.

2.3.2 Pattern recognition methods

The main goal with pattern recognition is to classify objects into a number of classes or categories. The objects can be images, sound records, or any other signal depending on the application. The classes or categories can be chosen based on what you want to find or identify. For this theses, the object is the sound recording and the classes are whether the subject is eating or not.

The first step in a classification problem is to identify measurable quantities that make the classes distinct from each other. The measurements used for the classification are known as the features. Features can be plotted to visualize the separation between the classes. If the classes are entirely separated, a straight line will be able to separate the groups. The separating line, called the decision line, constitutes the classifier. The role of the classifier is to allocate new feature samples to one of the classes defined. The training of the classifier can be divided into supervised, unsupervised, and semi-supervised methods, based on the a priori information available.

Supervised learning

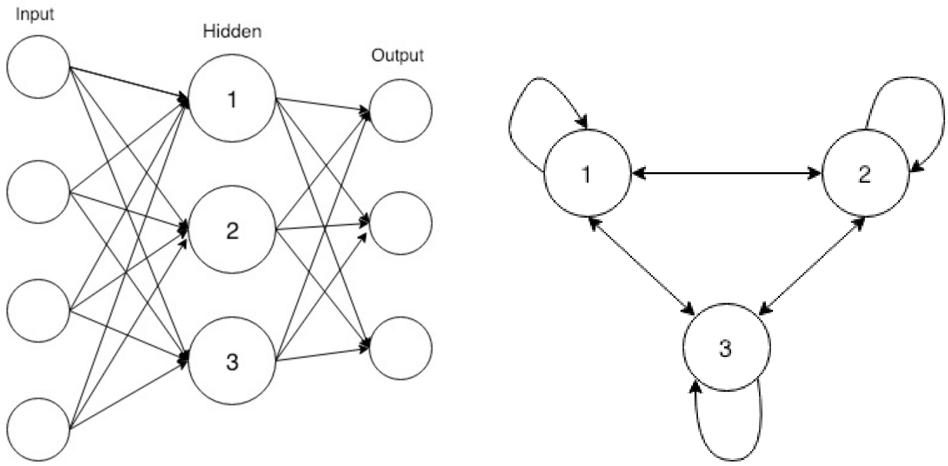
In supervised learning, signals labeled with known classes are used to train the classifier. These algorithms require a set of example data with labeled events or classes. In addition, the training set has to be of sufficient size and quality. These methods are used if there exists a data set with measurements and known classes or if it is easy to collect such data [22].

Artificial neural network The discipline of artificial neural network (ANN) is inspired by the nerve cells in the human brain and cortex. The nerve cells, or the neurons, are quite similar to a normal body cell except their to types of appendages: multiple dendrites and an axon. The dendrites branch out from the cell to get input from other neurons. The axon, on the other hand, branches out as an output channel to other cells. By shearing signals through the dendrites and the axon, the neurons build a network [36](ch 1).

An ANN has most often perceptron architecture. The network consists of nodes and weighted connections(arrows) between the nodes. The nodes represent the neurons, while the arrows represent the connection between the dendrites of one neuron and the axon of another[37]. The types of networks can grossly be divided into feedforward neural networks, recurrent neural networks, and their hybrids[36]. An example of two typical architectures is pictured in Figure 3. If a network, as in Figure 3a, consists of multiple hidden layers, it is called a *deep neural network*[22].

The transmitting of signals between neurons in the brain may have either an excitatory or inhibitory effect on neighboring neurons, depending on if the synapse increases the signal or decreases it. The connection between the neurons (the synaptic connections) shows plasticity due to long term changes in the strength of the connections as a result of the pattern of stimulation. This plasticity is thought to form the basis for learning and memory in the brain[36].

The learning process of an ANN is based on the learning and memory process in the human brain. The weights are changing according to the data in the training set. A common algorithm used for calculating the weights of a multilayered feedforward network is *Backpropagation*. A cost function is defined, often as the square of the error between the calculated output and the actual output of the training data. The algorithm seeks to minimize this error by changing the weights between the nodes. A gradient-descent-based algorithm adjusts the weights because the minimum of such complex cost function cannot be found explicitly.



3a: Layered feedforward network 3b: Recurrent network

Figure 3: Architecture of artificial neural networks.

Support vector machine Support vector machine (SVM) is a discriminant model for supervised learning, which means that the model uses the training data to find lines or hyperplanes which separates the classes. Discriminant models do not find any information about the distribution of the data within each class. A separating hyperplane is not unique as Figure 4 illustrates. SVM uses optimization techniques to find the optimal hyperplane[38][22].

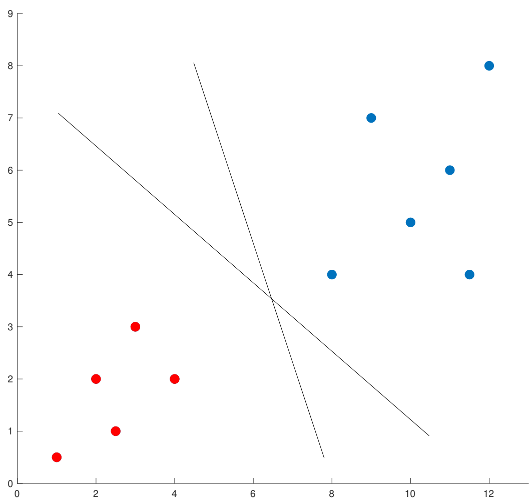


Figure 4: Two classes separated by to different hyperplanes.

The best separation between two classes, is the hyperplane with the greatest distance to each class. The distance from the separating hyperplane to the closest sample is called the *margin*. Figure 5 illustrates the margin of the hyperplane, which separates the blue and the red group. The best hyperplane is found by applying an optimization algorithm.

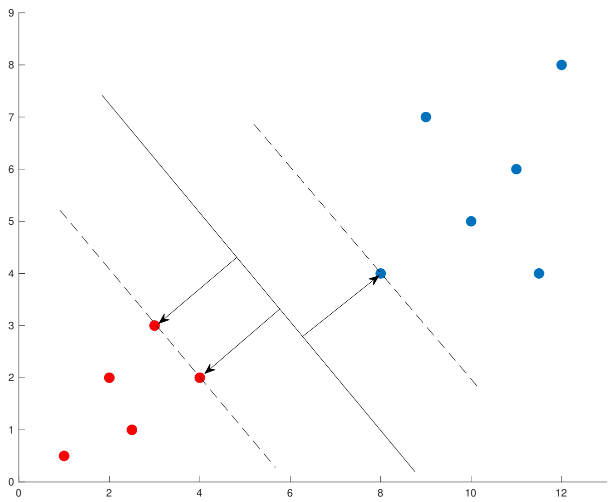


Figure 5: A hyperplane separating the blue and the red class. The margins are illustrated with dotted lines.

Hidden Markov model Hidden Markov models (HMM) are generative models, which seek to model the joint distribution in order to approximate the data generating process itself. The classification can be done by choosing the class which the object most likely belongs to[38].

unsupervised learning

It is not always possible to get a set of data which can be used to train a classifier. It might be impossible, very difficult or expensive to do measurements or tag the data with the correct classes for a training set. Even the number of classes or categories might be unknown. The goal in such cases is to find the underlying similarities in the data set and cluster (group) the data.

K-means clustering K-means is a well-known algorithm for clustering data points. The goal is to divide the data into clusters, or classes, such that objects in the same class have a high degree of similarity and dissimilarity with the objects in the other groups. The algorithm requires a number of clusters, k, as input. The algorithm starts with choosing a random location for the centers of all k clusters, called the *centroids*. Then, each data point is assigned to its closest centroid. The centroid values are then recomputed to fit in the center of the data points assigned. Samples are again assigned to the currently closest centroid before they are recomputed again. The algorithm continues with reassigning the samples and recomputing centroids until no change is obtained[39].

semi-supervised methods

The third group of pattern recognition algorithms is semi-supervised methods. These algorithms have the same goal as supervised methods, but a complete training set is not available. The training set available for the designer has a small number of labeled samples. This is often the case when labeling has to be done manually as this can be a difficult and time-consuming task[36](ch 2, p18). The training set contains some a priori knowledge which can be used to define constraints for a clustering algorithm[22]. Utilizing the little apriori information available can increase the efficiency and the accuracy of the classifier.

3 Aim of the study

This thesis aims to detect meals using sound recordings from the abdominal wall. Such meal detection system can be used as input to an artificial pancreas, or an automatic insulin pump, to aware the pump that a higher dose of insulin is needed due to a meal. The goal for the detection algorithm is to detect the meals as fast as possible, as accurate as possible.

3.1 Motivation

For diabetes patients, tight glucose control is crucial for avoiding hypoglycemia and hyperglycemia. The current semi-automatic insulin pumps require the patients to announce meals and exercise [40]. Such AP-systems rely on accurate information given by the user. The main goal for an artificial pancreas is to make the insulin dosing fully automated. Detecting the meals will be an important step towards this goal.

Meal detection based on data from a subcutaneous CGM has earlier been investigated. The latency of this measurement has been shown to influence meal detection. Studies have reported a delay of 20 to 40 minutes from the meal onset and until it is detected by an algorithm [15] [41] [42]. In order to avoid too high BGL after a meal, it is essential to detect the meal as soon as possible. The sound made by the intestines and stomach is investigated as a method for faster meal detection.

4 Previous Work by APT

In recent time, the APT research group has been working with the development of a sound based detection algorithms. A set of data have been recorded and tested for both bowel sound and meal detection. This chapter will give a brief description of previous studies done related to abdominal sound as well as a description of the available data set.

4.1 Existing data set of abdominal sound recordings

The provided data set was recorded by the APT research group in 2018. The data set includes:

- One recorded meal of 10 subjects enrolled for a pilot trial.
- 17 meals recorded on one subject.

Giving a total of 27 recordings. The following description of the setup is based on the explanation of the pilot trial described in [20].

All subjects were self-declared as healthy with respect to gastrointestinal functions. The volunteers were asked to keep their regular lifestyle prior to the recording and on the day of recording. After breakfast, the subjects were fasting until lunch. The recording sessions were scheduled to fit with the typical lunchtime for each subject. All subjects brought their own lunch to the recording session. A silent meeting room was used for all the recordings.

The recordings were done using a condenser microphone, of the type Sennheiser MKE2 P-C. (Sennheiser Electronic GmbH & Co. KG, Wedemark, Germany.), fixed in the center of the chest piece of a classical stethoscope. The stethoscope was attached to the upper right quadrant of the abdominal wall, using medical tape. During the recordings, the subjects were seated in a reclined position with elevated legs. They were asked to move as little as possible. After 30 minutes, they started to eat their lunch and finished their meal within 15 minutes. The recording continued for 45 minutes, resulting in a total of 90 minutes recording. Unexpected events, such as nose-clearing, coughing, changing the posture, or exogenous disturbances from the surroundings, were logged by the subjects. Furthermore, the subjects were not allowed to use electronic devices as it might interfere with the recording device.

4.2 Detection of bowel sound

MSc thesis by Kaja Kvello

The aim of Kaja Kvello's master theses [43] was to implement a bowel detection algorithm. Bowel sound occurrence frequency might be used for meal detection, as it has been observed increased bowel activity after a meal [13]. Based on literature and previous research, she found that the following traits can characterize bowel sounds:

- Frequency range of 100 - 500 Hz.
- Varying amplitude, but generally higher than surrounding noise.
- Sound durations varying between ca 10 ms and 1 s.

One recorded meal, from the data set described in section 4.1, was used to train and test a classifier. The bowel sounds in the recording were marked by auditory inspection, which she reports as challenging. Without medical expertise, she reports difficulties with separating bowel sounds from noise. Some bowel sounds might even be so small that they are difficult to notice.

First, a 50Hz highpass filter was applied for noise reduction. Secondly, the signal was segmented in frames of 30ms, giving 120 samples in each frame. Next, a set with features was extracted based on the mean amplitude, mean power spectral density (PSD) for different frequency bands, and the ratio of PSD around different frequencies. A feature selection algorithm, provided by Scikit-learn in Python, was used to select the best features. The selected features were used to train an SVM model for the classification.

The implementation gave promising results for the one recording used. The mean accuracy was 94% successfully detected bowel sounds. The main problem described was the false positive results giving a precision of 78%. 28% of the detected bowel sounds were marked in the test set as no bowel sound. However, this can be due to some incorrect markings of the test set.

Kölle, Data driven filtering of bowel sounds using multivariate empirical mode decomposition

Kölle et al. [20] Investigates a filtering method, intrinsic mode function-fractal dimension (IMF-FD), to differentiate bowel sounds from noise and artifacts. This method behaves like a series of bandpass filters as it utilizes the property of the multivariate empirical mode decomposition (MEMD), which decomposes the signal into its different frequency components. The resulting intrinsic mode functions (IMFs) are modulated in amplitude and frequency where transient sonic events occur. Based on the complexity of these functions, the information-carrying IMFs are selected. This filtering method enhances the nonlinear components of the original sound signal and segments them from the rest. Artifacts which occurred in the same frequency range as bowel sounds were eliminated by heuristic rules. The method was tested on the data collected, and close to 100% of the labeled bowel sounds were detected.

4.3 Meal detection

Kölle, Feasibility of early meal detection based on abdominal sound

This paper [15] investigates whether the patterns of bowel sounds allow fast meal detection. Data from the pilot trial described in section 4.1 were analyzed. Spectrogram analysis suggested that the power distribution over the frequency range changed shortly after the meal start. The power in different frequency bands was used as features. Different SVM models were trained based on a leave-one-out cross-validation procedure using ten recorded meals from one subject. The best suited parameter set was used to train the final SVM model. The final model was tested on the recorded meals from the pilot trial with ten recorded meals from different subjects. A few recordings had to be excluded due to background noise or the absence of audible bowel sounds. Five of the seven remaining meals were successfully detected on average 10 minutes (std: 4.4min) after the actual meal onset.

5 Method

Most of the previous work done, concerning sound from the digestion system, is related to the clusters of sudden bursts, called the bowel sounds. The methods used by Kvello[43] and Kölle[20] are both related to the detection of bowel sounds. In contrast, this project aims to detect a meal by evaluating the entire recording from the abdomen. The second paper from Kölle,[15], describes features and classification methods used for sound-based meal detection. This paper was used as a guideline for some of the choices made in this project.

Making a meal detection algorithm is an iterative process. It contains steps as collecting data, apply signal processing methods for feature extraction, define a training set, and choose and tune a classification algorithm. This chapter will describe the approach used to obtain a meal detection algorithm based on the sound measurements.

5.1 Evaluation of available data set

The studies done by Kölle reports a high variability in bowel activity among subjects and between different recording sessions[15][20]. She obtained a significantly lower accuracy when testing the classifier on different subjects, compared to the validation which used measurements from the same person as the training set[15]. Due to the variation among individuals, this thesis will use data from the same subject for training and testing, making a person specific detection algorithm. Before a detection system possibly is being generalized to detect meals on different subjects, it has to work for one subject. In order to have some recordings for training and some others for testing, more than one recording from each subject is needed. The pilot trial, recorded on ten different subjects, were therefore excluded from this study. The remaining 17 recordings, from the same subject, were evaluated for future work.

To ensure similar training data, the same recording procedure has to be followed. Consequently, the stethoscope has to be attached at the upper right quadrant of the abdomen, the subject has to be seated during the entire recording and be fasting from breakfast until lunch. Some recordings deviated from this procedure as, for example, the subject was lying, or the stethoscope was attached to the left side of the abdomen. All recordings which deviated from the procedure were excluded from the data det.

In order to have enough data for both the "meal" and "no meal" class, the recordings must include a proper amount of data both prior to the meal and from the digestion phase. Recordings with less than 15 minutes of either class had to be excluded.

Recordings used for training require high quality. If the training data is of bad quality, typically containing much noise, it is hard to separate noise from the event which will be detected. The provided data set were recorded in a regular meeting room. Even though the room was quiet, noise from outside was heard. For example, people walking or talking outside the room and doors opening and closing in the same corridor. Another source of noise is the attachment of the stethoscope, causing mechanical noise. Any movement or change of position was heard in the recording. This was due to movements of the skin under the stethoscope and the tape which slightly loosens. The stethoscope was also covered with clothes, which touched the sensor and caused some artifacts.

The remaining recordings, which were of a proper length and followed the same setup, were further evaluated to ensure a good recording quality. The recordings were plotted for visual inspection and played through headphones for auditory inspection. Some recordings contained less variation which could be caused by poor contact between the skin and the microphone or a too low gain on the recorder. A few contained longer periods of continuous noise from, for example, music or ventilation systems. Recordings of bad quality were not included in the data set used in this study.

The following list summarises the numbers of recordings which had to be excluded, and the reason why:

- 7 recordings were excluded due to not following the same recording setup.
- 3 recordings were excluded due to weak signal caused by loose connection of the stethoscope or a too low gain on the recorder.
- 2 recording were excluded due to longer periods of noise.
- 2 recordings were excluded due to a too short recording.

After the exclusion of the improper data, there are four recordings from one subject left. In order to test whether the method can be generalized to other subjects, data from more than one person is needed. It was therefore decided to record data on one more person. When being aware of the limitations with the existing data set, an effort was made to achieve data containing less noise.

5.1.1 Regional ethical committee approval

Before starting to collect sound data from a new subject, Regional ethical committee(REK) has to approve the experiment. The previous REK approval did cover one more subject, and it was therefore not necessary to apply for a new approval. However, The research group want to do more recordings in the future and did send a new application. REK is not evaluating this application before June.

5.2 Acquisition of new data

The same procedure was followed, as described in section 4.1, for the collection of new data. The recording session included fasting, eating, and digestion. The meal was chosen to be the same as the lunch reported for the first subject; bread with cheese. Approximately three slices of bread were eaten during each recording.

The same microphone and stethoscope as the previous recordings were used. The microphone was connected to a *Roland Octa-capture* sound card. The data were then saved on a laptop using the audio software *Audacity*[44]. No settings which can influence the recording were changed compared to the recordings done on the first subject. The sample rate was 48000Hz, with 24 bits per sample. The gain was constantly kept on 46.5.

An echo-free room, available at the department of electronic systems NTNU, was used in order to reduce the external noise. The walls, the roof, and the floor are covered by wedged shaped isolation which absorbs sound and electromagnetic waves. The room is very quiet, making it ideal for sound recordings. The sensor was not covered with clothes which eliminated the artifacts caused by the clothes touching the sensor. However, the mechanical noise caused by the attachment of the stethoscope was not eliminated as the same attachment with medical tape was used.

Four new recordings were collected. Slightly shorter duration for the new recordings was chosen to ensure the comfort of the subject. Table 1 summarises the total amount of data used in this thesis.

	Amount	**Recording session**
Subject 1	4 recordings, one meal pr. recording	30min fasting, $\leq$15min eating, $\geq$45min digestion
Subject 2	4 recordings (noise-free), one meal pr recording	20min fasting, $\leq$10min eating, 30 min digestion

Table 1: Recordings used in this project.

5.3 Software

The implementation was done using MATLAB R2018b with the following toolboxes installed: Audio, Signal Processing, Wavelet and Deep Learning.

5.4 Preprocessing

A sample rate of 48000Hz was used for the recordings. Since bowel sounds have been found to lie between 50 and 1500Hz [11] [12], a sampling frequency of 4000Hz will accelerate the processing without loss of information. The MATLAB function *decimate(a,r)* was used to reduce the sampling rate of signal x, by a factor of r corresponding to 48000/4000. *decimate* lowpass filters the input to guard against aliasing and downsamples the result. The default lowpass Chebyshev Type I infinite impulse response (IIR) filter of order 8 was used.

The recordings have to be segmented into time segments in order to classify a segment of sound as "meal"-sound or "no meal"-sound. Kölle et al. tested segment lengths of 10sec, 20sec, 30sec and 60sec. Successive segments were overlapping by half of their length. For the feature used in the study presented, a window length of 20sec with 10sec overlap was found to give the best results[15]. The same segment length was chosen for this project.

5.5 Feature extraction

The downsampled signal itself could not directly be used as input to a classifier. First, the elements in the signal which correlates with the meal intake had to be found. There are many signal processing tools which can generate different features of a sound signal. Obviously, not all the processing and feature extraction techniques could be tested during this project. Literature regarding bowel sound detection, lung sound classification, and speech recognition was read to gain information about common features for sound classification. The most common features used for similar problems were used as a first guide for selecting features.

Additionally, spectrograms were plotted to see how the signal changed with respect to time and frequency. The power spectral density (PDS) were plotted using the MATLAB function *spectrogram(signal, window, overlap, f, fs)*. A change in power density according to a meal intake, will indicate the frequency band of the gastrointestinal sound. The features should be extracted from these frequency bands.

Before the recordings of subject 2 were done, it was decided to find features based on the increase in power which Kölle reports in the paper [15]. As reported, the spectrogram shows that the gastrointestinal sound activity increases a few minutes after the meal. This can be seen in all the four recordings from subject 1 used in this project. A clear example is pictured in figure 6.

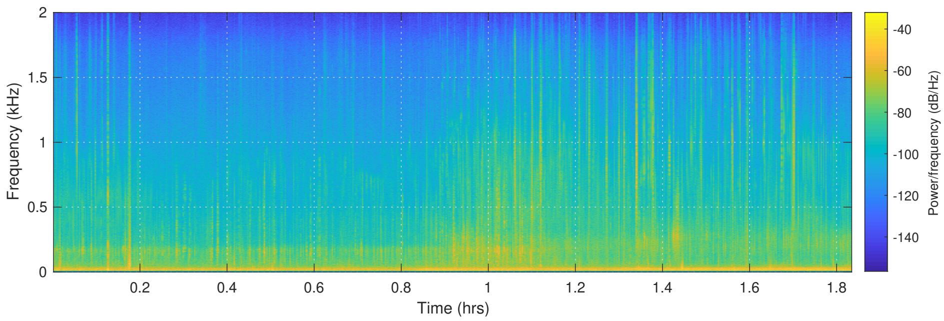


Figure 6: Power spectral density of one recording from subject 1. The meal started at 51 minutes (0.85hrs).

MFCC is a method used by the majority of studies regarding audio signals, especially in speech recognition. This method is also used on biomedical data and is listed as one of the most common features for lung sound signals (together with autoregressive model, energy, entropy, spectral features, and wavelet.) [34]. Because of the popularity of this method, it was decided to test if MFCCs can be utilized in sound based meal detection.

The wavelet decomposition is a common way to decompose the signal in terms of high and low-frequency components. As the meal information has been shown to lie in the lower frequency range, the detailed information from higher frequencies is irrelevant for meal detection. Approximation coefficients were therefore chosen and were used for entropy computations.

Spectrograms with logarithmic axes were later plotted to show the activity for the lowest frequencies of the signals. An example is pictured in Figure 7, and it is possible to see an increase in activity immediately after meal onset at 20minutes. This frequency range is outside the specifications of the microphone. Hoewer, attachments to the microphone might increase the sensitivity and changes the specifications. Since a change is observed due to a meal, it is likely that the microphone is able to measure lower frequencies than the ones described in the product specification[45]. Based on these findings, it was decided to extract the power of narrow frequency bands in the frequency range 1-55Hz and investigate this feature set separately.

A summary of the two chosen feature sets are pictured in Figure 8. Both sets of features were tested on both subjects. The following sections describe in detail how each set of features were extracted.

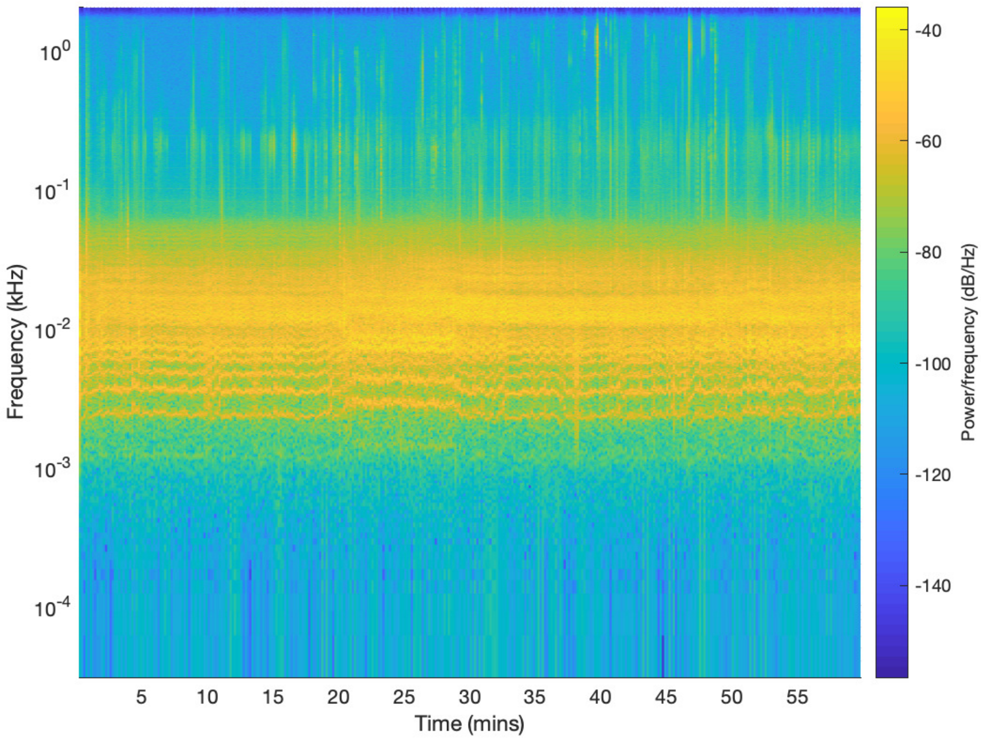


Figure 7: Spectrogram with logarithmic frequency scale. Meal started after 20minutes.

1.	**2.**
- MFCC - Wavelet decomposition + Entropy	- Power of the lowest frequencies

Figure 8: The two sets of features tested.

5.5.1 First set of features

MFCCs The mel-frequency sepstral coefficients were extracted using the *mfcc(audioIn, fs)* build in function in MATLAB [46]. The function returns the log energy, MFCCs, *delta* and *deltaDelta* values. Delta values describes the change in MFCCs, while the *deltaDelta* describes the change in *delta* values. Only the MFCCs were investigated in this thesis. The line of code, showed in Listing 1, was used to generate the MFCCs. *signal* is the recording, downsampled to 4000Hz. The window length was set to 80000 samples with 40000samples overlap, which corresponds to 20seconds frames with 10 seconds overlap. *Coeffs* is a matrix with coefficients corresponding to each frame of data in each row. The first column is the log energy, while the rest is MFCCs. The MFCC function can return as many MFCCs as there are valid passbands. A passband is valid if its edges fall below $2/fs$, where fs is the sampling frequency. The default number of 13 coefficients were used, corresponding to 13 passbands which are defined in Table 2.

```
coeffs = mfcc(signal, 4000, 'WindowLength', 80000, '
    OverlapLength', 40000);
```

Listing 1: MATLAB code for calculations of MFCC coefficients.

Filters	Passband Edges (Hz)
Filter 1	[133 267]
Filter 2	[200 333]
Filter 3	[267 400]
Filter 4	[333 467]
Filter 5	[400 533]
Filter 6	[467 600]
Filter 7	[533 667]
Filter 8	[600 733]
Filter 9	[667 800]
Filter 10	[733 867]
Filter 11	[800 933]
Filter 12	[867 999]
Filter 13	[933 1071]

Table 2: Filter bank frequency edges.

Wavelet decomposition The signals were first divided into overlapping frames of 20seconds with 10 seconds overlap using the MATLAB function *buffer(x,n,p)*. The function returns a matrix with data from one frame in each row. The first and the last row of data were removed as the first frame started with *p* zeros, and the last one was an incomplete segment. For each frame of data, a 3-level wavelet decomposition was computed. The first order Daubechies wavelet, also known as the Haar wavelet pictured in Figure 9, was used. Daubechies wavelet is one of 16 wavelet families available in the Wavelet Toolbox in MATLAB. This particular wavelet was selected as a first choice as it was the first option.

The entropy was calculated of the coarse scale approximation coefficients using the MATLAB function *wentropy(X,T)*. The function returns the entropy of type T of a vector or matrix *X*. The *Shannon* entropy was chosen among the six available entropy types. Also, different entropy types could have been tested,

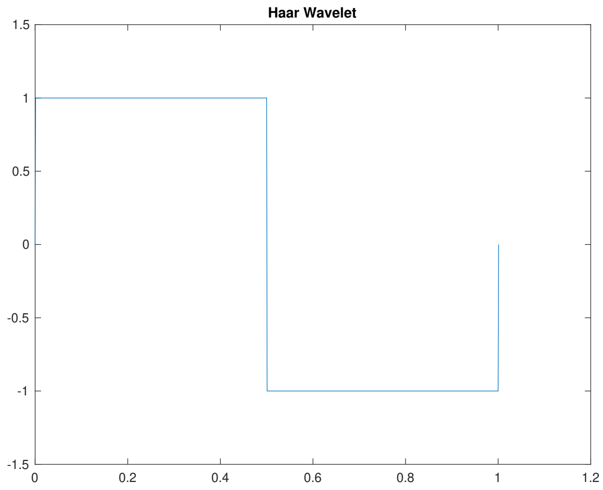


Figure 9: Daubechies's wavelet

but the first choice gave good results. Listing 2 shows the code used to segment the signal, decompose it into a three-layer wavelet decomposition, and calculate the entropy.

```
seg = buffer(signal,80000,40000);
segmentedSignal = seg(:,2:end-1)';
ent = [];
for i = 1: size(segmentedSignal,1)
     [c1,l] = wavedec(segmentedSignal(i,:),3,'db1');
        ent=[ent wentropy(c1(1:10000),'shannon')];
end
```

Listing 2: MATLAB code for wavelet decomposition and calculation of entropy.

Feature selection Kvello reports in her thesis that 100-500Hz is a common frequency range of focus when working with bowel sound [43]. Based on her findings, the first five MFCCs were selected, corresponding to 133-533Hz (filter 1-5 in Table 2). Later, a classification algorithm was tested using both all MFCCs and the five selected coefficients. The best result was obtained when using the selected MFCCs together with the entropy.

5.5.2 Second set of features

An increase in power was seen immediately after the meal onset in the logarithmic spectrograms. Training data should ideally have been extracted immediately after the meal onset, but due to the difficulties of avoiding noise in the data set from subject 1, the training data extracted 10 minutes after the meal onset had to be used. For subject 2, the training data were extracted directly after the meal onset.

The signal was segmented in the same way as for the first set of features, frames of 20 seconds with 10 seconds overlap. The average band power was calculated for each frame of data using the function *bandpower(x, fs, freqRange)*. The frequency range was defined as a two-element vector as "freqRange" in the code listed in listing 3. The power corresponding to all the frequency bands defined in *freqRange* were saved in matrix *y*.

```

freqRange = [1 5; 10 15; 20 25; 30 35; 40 45; 50 55;
             5 10; 15 20; 25 30; 35 40; 45 50;
             1 10; 15 25; 30 35; 40 55];

for i= 1: size(reqRange,1)
y =[y; bandpower(segmentedSignal, 4000, freqRange(i,:))];
end
```

Listing 3: MATLAB code for power calculations in the lowest frequency range.

Feature selection Fifteen combinations of overlapping frequency bands in the range of 1Hz to 55Hz were extracted for each recording from both subjects. A feature selection was done based on the features extracted for subject 2, as they contained less noise. Figure 10 shows all the features, in feature set 2, for one of the recordings. An increase in power appears immediately after the meal started (at the 119th segment) for most of the lowest frequency bands. Similarly, a change in power for the lowest frequency bands was observed for all recordings done on subject 2. Only the frequency bands which provided visible meal response were selected. Table 3 summarises the selected frequency bands.

Frequency bands [Hz]
5-10
10-15
15-20
1-10
15-25

Table 3: Selected frequency bands for power computations.

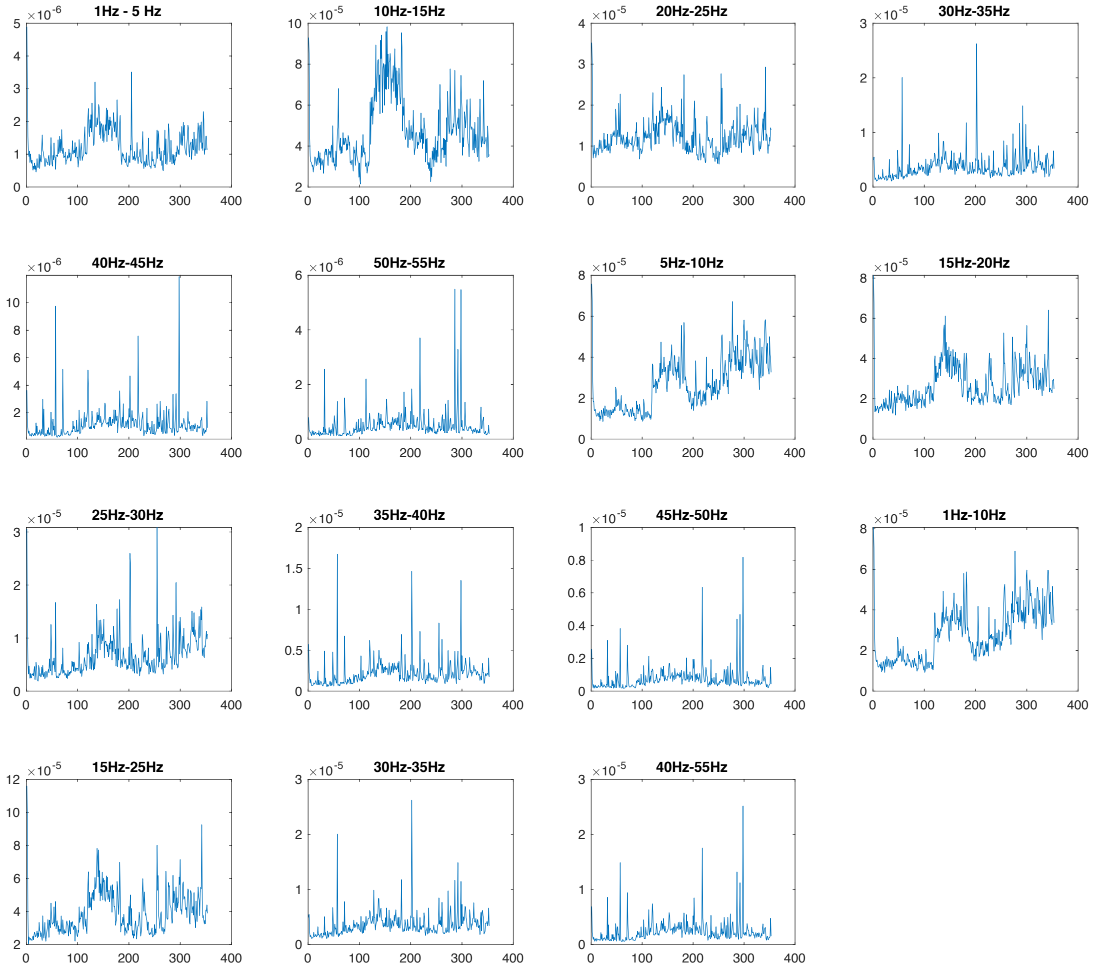


Figure 10: The power in different frequency bands in one recording (subject 2).

5.6 Defining a training set

For supervised learning, it is important that the training set is marked correctly such that the classifier can be trained to do the correct classification. Labeling the recordings with "meal" and "no meal" was a challenging task as it is not known when the response of a meal is reflected in a sound signal nor the duration of the response. The only a priori information is the time the subject started eating. It is natural to expect a time delay from the meal starts until a response is found in the recording. Kölle defines the two terms, *response delay* and *response duration* which describes the delay and duration of the response relative to the meal start. Figure 11 visualises these terms.

Kölle et al. tested different values for response duration and delay, and con-

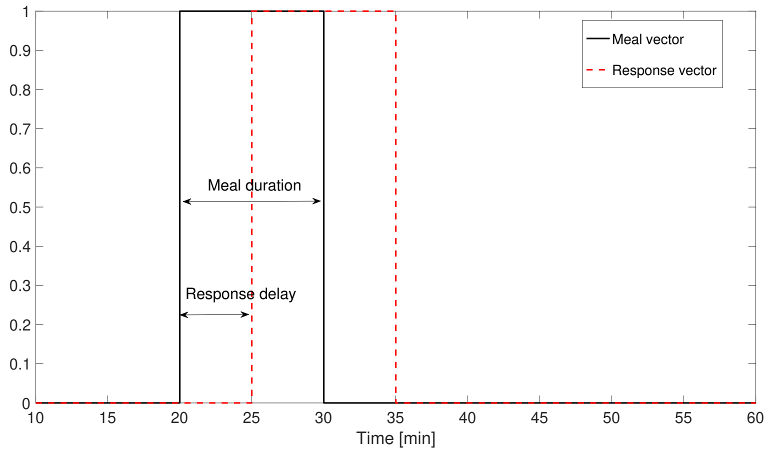
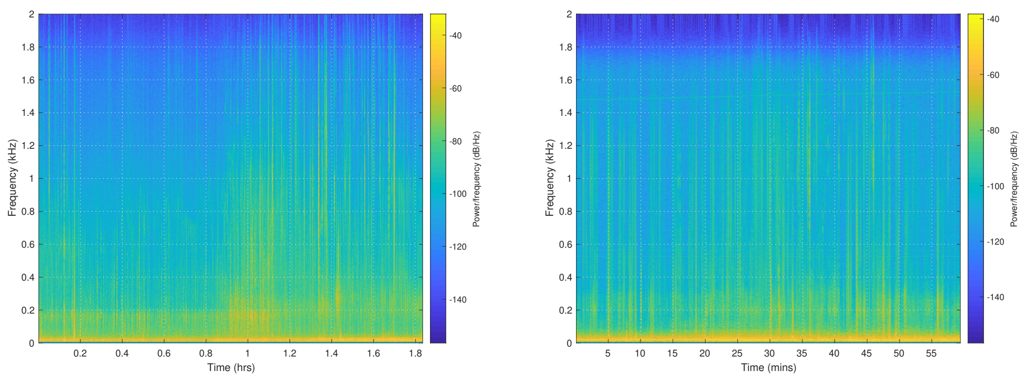


Figure 11: Response vector relative to meal intake.

cluded that 10 minutes response delay and 20 minutes duration gave the best result for the available data and the chosen features [15]. The spectrograms plotted verified that a response was seen approximately 10 minutes after the meal onset. However, the response and the delay varied among the recordings and the subjects. For subject 2, it was impossible to see the same response in a linear scaled spectrogram. A comparison of a spectrogram from each subject is pictured in Figure 12.



12a: Subject 1, meal onset at 0.85hrs

12b: Subject 2, meal onset at 20min

Figure 12: Spectrograms from subject 1 and subject 2.

Even though the response delay was visible in the recordings of subject 1 only, "meal"-data were extracted 10minutes after the actual meal onset and with a duration of 15minutes for both subjects. "No-meal"-data were extracted 15minutes prior to the meal onset. Figure 13 visualises the labelling. Since the first set of recordings contained external noise, this data had to be carefully inspected to avoid noise in the training set. It was a difficult and time-consuming task to separate artifacts from bowel sounds, but clear artifacts were avoided from the training data.

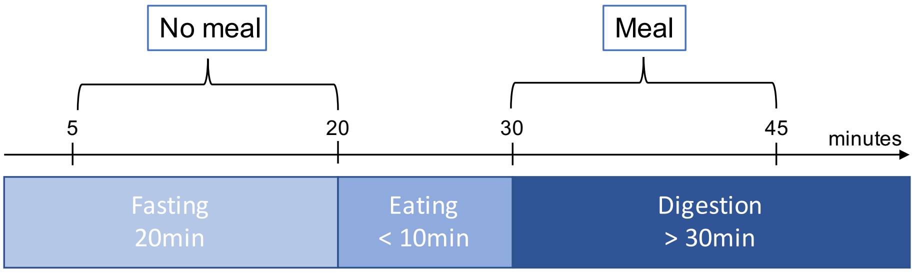


Figure 13: labelling recordings from subject 1.

The new recordings done contained less noise, which made it possible to extract noise-free training data without listening to the recordings. The labeling process was significantly simplified, and "meal"-data were extracted both with 10 minutes response delay and directly after the meal as pictured in Figure 14.

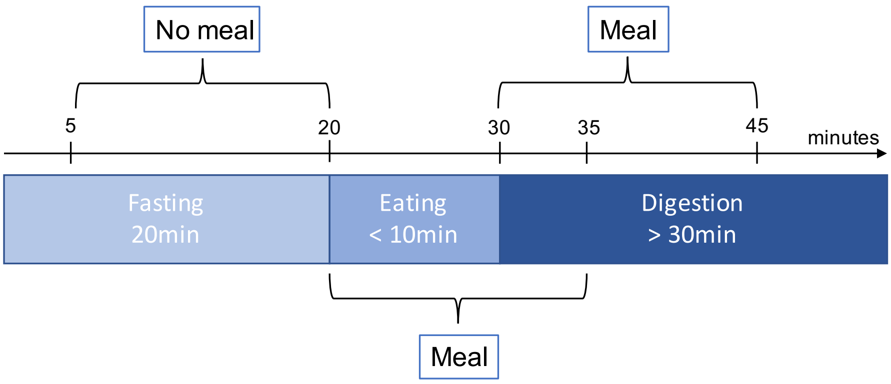


Figure 14: labelling recordings from subject 2.

5.7 Choice of pattern recognition method

Both supervised and unsupervised methods were considered. Kvello reports in her thesis [43] that k-means clustering gave poor results for bowel sound detection. Due to the available training data, with known meal onset, supervised learning was a natural choice.

Different methods, as described in section 2.3.2, were evaluated. Initially, different classification methods were tested using the application *classification learner* in MATLAB. This application makes it possible to train different models at the same time, and select the one giving the best result. Different supervised methods gave a minor difference in classification accuracy. A neural network was tested separately, as it was not included in the *classification learner* application. One method had to be chosen for further tuning, and a neural network was chosen as it gave slightly better results initially than the ones tested with the *classification learner* application.

5.7.1 Neural Network

After extracting features and label the training data, a neural network could be built. The *Deep Learning Toolbox* from MATLAB was used to build, train and test the network. A feed-forward network was chosen and the function *patternnet(hiddenSize,TrainFcn)* was used to build the network. The two input parameters, *hiddenSize* and *TrainFcn*, were later tuned to find the size of the network giving the best meal classification.

Initialization The network was initialized by the default initialization function *initlay*. Each layer is initialized according to its own initializing function *net.layers*{*i*}*.initFcn*, where *i* indicates the layer. The initialization method for each layer can be defined using *net.layersi.initFcn*

In this project, *'initnw'* was used to initialize the weights and biases in each layer. This algorithm was chosen as it chooses parameters according to the Nguyen-Window initialization algorithm, which aims to distribute the active region of each neuron evenly in the input space such that fewer neurons are wasted. The even distribution makes the training work faster as each area of the input space has neurons.

The Nguyen-Window initialization algorithm also includes a degree of randomness, meaning that the initialized weights and biases are slightly different for each initializing. Since it is impossible to know which initializing weights and biases giving the best result on beforehand, the randomness makes it possible to test different initializing parameters. However, this gives a different network for each training, meaning that different results can be obtained even though the same data is used for training, validation, and testing.

Training After being initialized, the network can be trained. Backpropagation was a natural choice, as the most efficient algorithms are based on backpropagation. MATLAB provides a lot of different training functions, based on backpropagation, in the Deep Learning Toolbox. MATLAB recommends *trainlm* as a first choice as it provides the fastest training, even though it requires more memory than other algorithms. This algorithm updates the weights and biases according to Levenberg-Marquardt optimization. Since memory was not a concern in this project, *trainlm* was the chosen training function.

The network was tuned by changing the number of hidden layers and the number of neurons in each layer. As a starting point, a network with one layer and three neurons was tested. The number of neurons was increased as long as the performance improved. Two layers were tested, and the size was expanded until the performance stopped to improve. A network, pictured in Figure 15, with two hidden layers with five and three neurons, respectively, was found to give the best result.

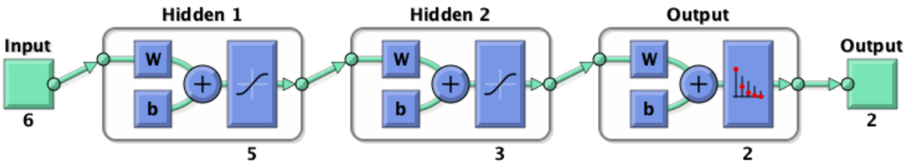


Figure 15: The final network used.

Validation The validation aims to stop the training process when the network is properly trained and before it gets overtrained. *trainlm* uses a validation set to check when the training process should be stopped. When a chosen number of validation epochs in a row fail to improve or remain the same, the training process is stopped. In this project, the default number of 6 epochs were used. In general, 10-14 epochs were needed before the validation stopped to improve.

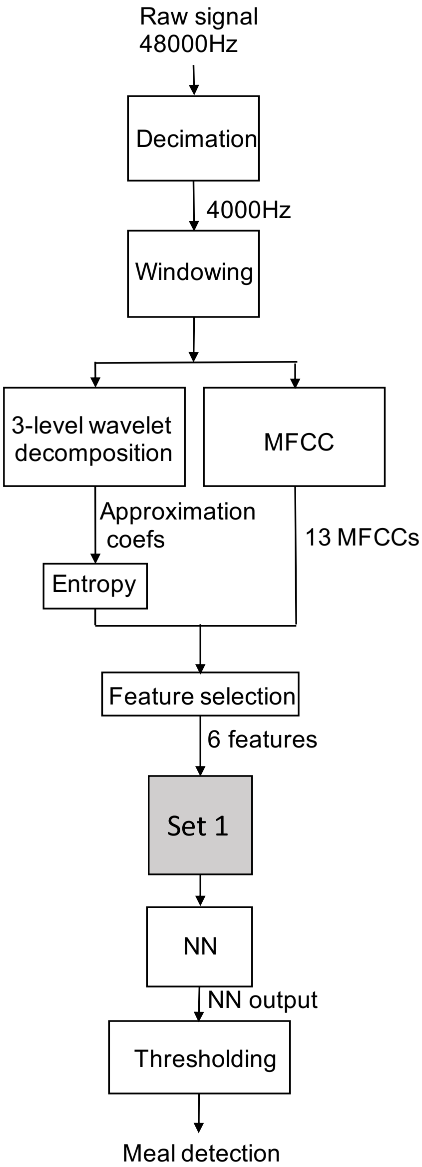
Testing The testing is done to check that the network is generalizing well, and does not influence the training. For each subject, one recording was left out while the remaining three recordings were used for training and validation. This was done for all four recordings for both subjects. Because of the randomness in the initializing of the network, the result can be different each time a network is trained. Therefore, a network was trained and tested three times for each test-recording.

5.7.2 Thresholding

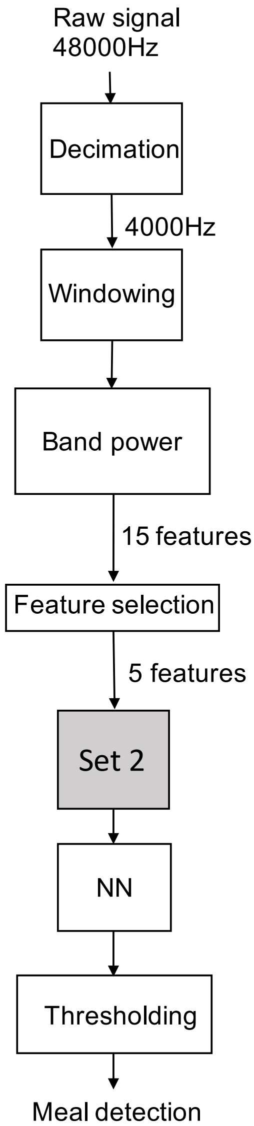
The output of the neural network were vectors with estimated values between 0 and 1 for each class. A value above 7.0 for the meal class, was defined as a *meal classification*. In order to obtain a more robust detection, with less false positive results, it was defined that at least 4 segments in a row had to be classified as a meal before the detection algorithm will detect the meal. As a result, the fastest meal detection possible was after 4 segments of data, corresponding to 40 seconds. The MATLAB code used for the classification is listed in Listing 4. *y_est* is the estimation provided by the neural network and *MealDetect* is the vector saving each segment which was detected as a meal.

```
MealDetect = [];
for i=4:size(y_est,2)
    if (y_est(1,i-3)>0.7) && (y_est(1,i-2)>0.7) && (y_est(1,i-1)>0.7) && (y_est(1,i) > 0.7)
        MealDetect = [MealDetect i];
    end
end
```

Listing 4: MATLAB code for a meal detection algorithm.



16a: Illustration of method 1



16b: Illustration of method 2

Figure 16: Flowchart summarising the two methods tested.

6 Results

This chapter will describe the improvements in the new recordings, and the results of the two different feature set tested for early meal detection.

The results are listed in tables for each set of features separately. The tables include which recordings (as a number of 1 to 4) the training and validation data were extracted from, as well as the results for the test meal. As described in section 5.7.1, the randomness in the initialization phase of the network results in different results for each training. Consequently, the results for three different initialization parameters are included in the tables to show the impact of the initialization. Meal detections within 10 minutes are marked with a green color. Late detections, later than 20 minutes after the meal onset, are marked with yellow. The plots of the output of the neural network are plotted with the segment number on the x-axis.

6.1 New recordings

The new recordings, done in the echo-free room, gave significantly less noise. No noise from outside was heard. The number of artifacts caused by clothes touching the sensor was also eliminated by not covering the sensor with clothes. However, the mechanical noise made by the attachment of the stethoscope was not improved as the same attachment was used. Even though the subject tried to sit as still as possible, a few movements were made. The few movements were logged so that it could easily be avoided from the training set.

It was significantly easier to extract training data from the new sound measurements, as it was not necessary to listen through the recordings to avoid external noise. Much time was saved compared to the time spent on tagging the original dataset. As a result, it was easy to test response vectors with different response delays. "Meal"-data for the training set were therefore extracted both directly after, and 10 minutes after the meal onset for subject 2.

6.2 Output of the neural network

The output from the meal class of the neural network is in this thesis defined as a meal *estimation* as it does not provide a crisp classification. The meal estimation varied for each recording tested, as well as for each training due to the randomness in the initial parameters. A typical estimation which provided a successful detection is pictured in Figure 17. An estimated value of 1 can be seen immediately after the meal onset. Some separate segments are estimated

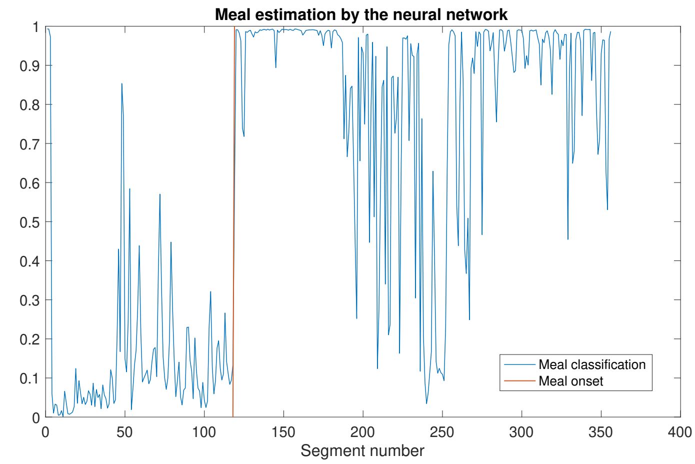


Figure 17: Typical meal estimation by the neural network.

to a value close to 1 before the meal, which can be seen as spikes in the plot. The detection algorithm prevented these spikes from being false detections as four segments in a row had to be estimated to a value above 0.7.

6.3 Meal detection with first set of features (Entropy + MFCCs)

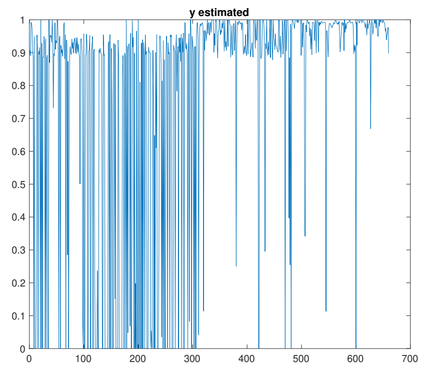
Subject 1

The following results were obtained using the first set of selected features. Table 4 shows the time of the first meal detected by the meal detection algorithm for subject 1. In general, the algorithm detects the meals only a few minutes after the meal onset. In this case, "meal" data in the training set were extracted 10 minutes after the actual meal onset. Nevertheless, the algorithm detects some meals already af5DEE5Der one minute, which indicates that the response of a meal appears earlier than 10 minutes after starting to eat.

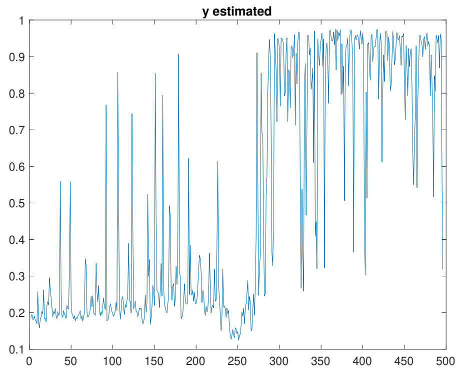
				Time of first meal detection [min]			Actual meal onset [min]
Training rec.		Validation rec.	Test rec.	1	2	3	
4	1	3	2	60	1	53	51
3	2	1	4	38	49	49	40
3	2	4	1	31	31	31	30
1	2	4	3	1	71	4	50

Table 4: Subject 1: Meal detection results using feature set 1 and 10min response delay.

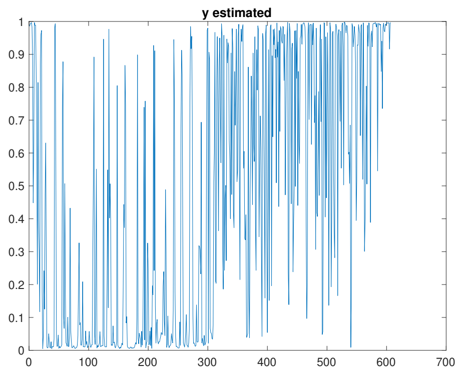
Table 4 shows four false detections and one late detection. The meal estimation, done by the neural network, which provided false meal detection is plotted in Figure 18 for further inspection. The meal estimates increases after the meal onset in all four false detected meals. This indicates that another detection algorithm, or filtering the NN output before thresholding, could have avoided some of the false detections. Possible reasons for false detections and suggestions of improvements are further discussed in section 7.2 and chapter 9, respectively.



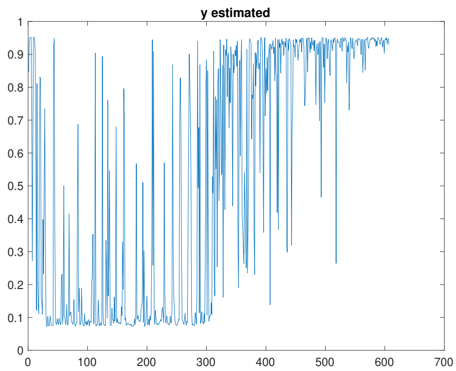
18a: Meal detection at 1min,
actual meal onset at segment 306



18b: Meal detection at 38min,
actual meal onset at segment 240



18c: Meal detection at 1min,
actual meal onset at segment 300



18d: Meal detection at 4min,
actual meal onset at segment 300

Figure 18: NN output from false meal detections in Table 4.

Subject 2

The same set of features (entropy and MFCCs) were extracted and tested on the new data set. First, training data were extracted in the same way as for subject 1, 10 minutes after the meal onset. The results can be seen in Table 5. Similarly, as for subject 1, the algorithm provides fast meal detection, although there are some false detections.

				Time of first meal detection [min]			Actual meal onset [min]
Training rec.		validation rec.	Test rec.	1	2	3	
1	2	3	4	22	21	21	20
2	3	4	1	1	1	1	19
3	4	1	2	15	NA	21	20
1	4	2	3	16	22	24	20

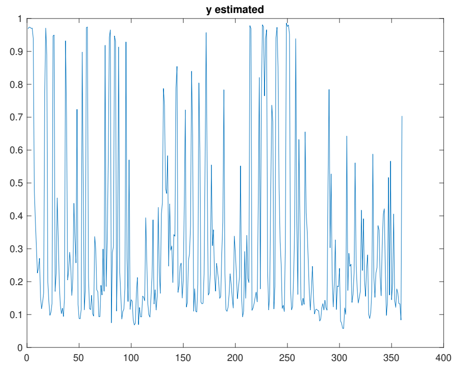
Table 5: Subject 2: Meal detection results using feature set 1 and 10min response delay.

Since the recordings of subject 2 were noise-free, it was easy to extract training data without careful inspection of the recordings. Based on the findings so far, which indicates that a meal response is visible directly after the meal onset, a response vector without response delay was tested. "Meal"-data were extracted directly after the start of the meal. The results were significantly improved, which can be seen in Table 6. Only three false detections occurred, and one late detection.

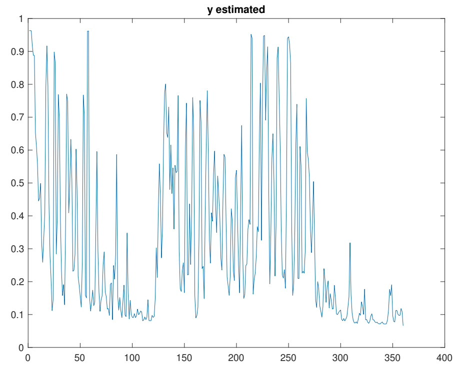
				Time of first meal detection [min]			Actual meal onset [min]
Training rec.		validation rec.	Test rec.	1	2	3	
1	2	3	4	22	21	28	20
2	3	4	1	1	1	42	19
3	4	1	2	21	15	20	20
1	4	2	3	21	21	21	20

Table 6: Subject 2: Meal detection results using feature set 1 and no response delay.

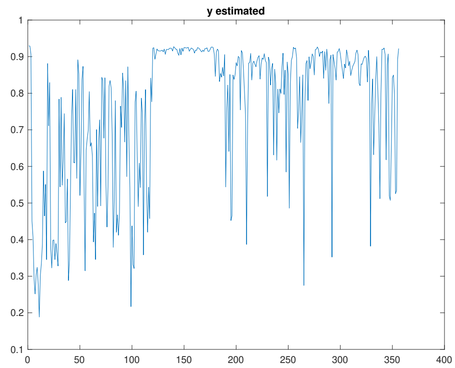
Figure 19 shows the output from the neural network, which resulted in false detections. The meal detection after 15 minutes, Figure 19c, could probably have been avoided by either filter the output of the NN before the thresholding or choose different detection criteria (threshold and number of segments). The meal in recording 1 is often detected incorrectly. Contrary to the false detection in recording 2, the output of the neural network does not provide a visible meal response, which can be seen in Figure 19a and 19b.



19a: Meal detection at 1min,
Actual meal onset at segment 120



19b: Meal detection at 1min,
Actual meal onset at segment 120



19c: Meal detection at 15min,
Actual meal onset at segment 120

Figure 19: NN output from false meal detections in Table 6.

6.4 Meal detection with second set of features (power)

Subject 2

The second set of features tested, power of a selection of frequency bands, gave the fastest meal detection for subject 2. Only three false positive results occurred, which can be seen in Table 7. The output from the neural network, which provided false detection of meals, are pictured in Figure 20. The output from the neural network of the first false detection, Figure 20a, shows a completely wrong classification by the neural network. It might be due to poor training or bad initializing for that particular training. Once again, the meal in recording 1 fails to be detected. The output of the neural network, 21b and 20c, has several spikes of false classifications before the meal onset.

				Time of first meal detection [min]			Actual meal onset [min]
Training rec.		**validation rec.**	**Test rec.**	**1**	**2**	**3**	
1	2	3	4	22	1	22	20
2	3	4	1	1	1	28	19
3	4	1	2	21	22	21	20
1	4	2	3	23	21	21	20

Table 7: Subject 2: Meal detection results using feature set 2 and no response delay.

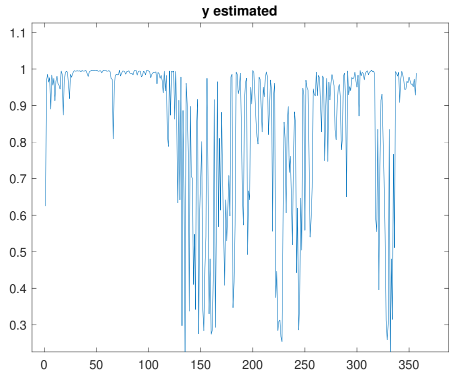
Subject 1

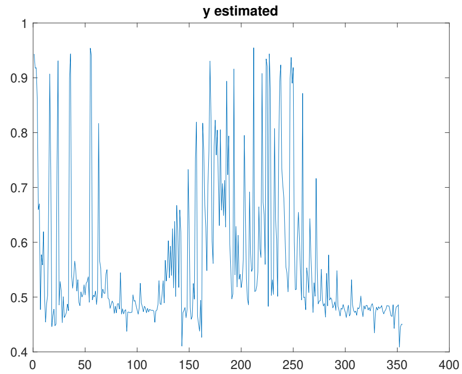
The same good results were not found for subject 1 when using the second set of features and a response delay of 10 minutes. In general, no meals were detected using the same algorithm as for subject 2. Figure 21 shows two examples of output from the neural network. Response of the meal can be seen in Figure 21a, while Figure 21b shows no meal classifications. In addition, the meal estimate does never exceed 0.7, which is the threshold value for meal detection.

Summary

The meals which were successfully detected were in average detected within 4 minutes. The results give a strong indication that meal related sounds appear shortly after meal intake and that it is possible to separate the meal-related sound from regular gastrointestinal sound activity. A meal detection within 4 minutes is a major improvement compared to meal detection using CGM.

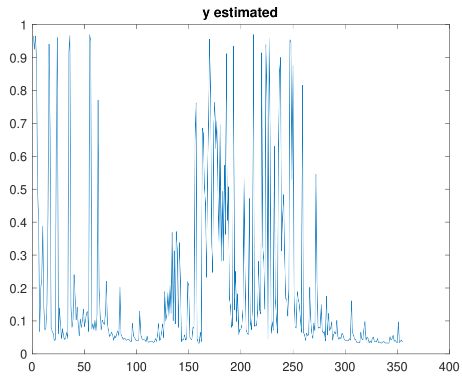
The first set of features (entropy and MFCCs) provided the best result for both subjects combined. The neural network provided estimations with visible meal response in close to 100% of the tested recordings. However, some false detection occurred. Chapter 9 will discuss possible improvements of both the classification and the detection algorithm.





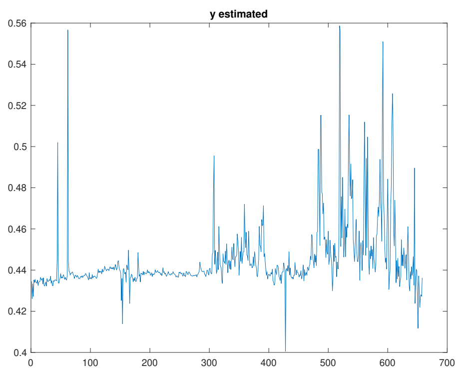
20a: Meal detection at 1min,
Actual meal onset at segment 120

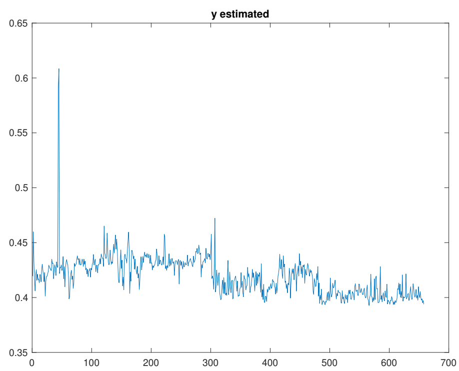
20b: Meal detection at 1min,
Actual meal onset at segment 120



20c: Meal detection at 1min,
Actual meal onset at segment 120

Figure 20: NN output from false meal detections in Table 7.





21a: Meal onset at 306 segments

21b: Meal onset at 306 segments

Figure 21: Subject1: NN output when using feature set 2 and 10min response delay.

7 Discussion

The present study demonstrates that abdominal sound recordings can potentially be utilized for early meal detection. This chapter will discuss the results in more detail as well as important aspects of the integration of a meal detection system in an artificial pancreas. There were also some limitations related to the present study, which will be further discussed. Suggestions for improvements and further steps towards the integration of a meal detection system in an artificial pancreas are topics discussed in the next chapter, Future work.

7.1 Limitations associated with the data set

The recordings used can, of course, not be related to a real-life scenario. No one sits as still as possible in a completely silent room while eating their lunch. In this early stage of the development of a sound based meal detection algorithm, the classification problem has to be as simplified as possible. One has to make sure that it is possible to record meal related sound and classify meals based on the recordings before noise and movements are included. Noisy signals will increase the complexity of the classification problem, making it difficult to extract the meal-related sounds of interest. However, making a classification separate all sources of noise from gastrointestinal sound might require a lot of training data, which were not available for this study.

Another simplification made, is the meals. Different food may create different sound during digestion. In this project, the meals were standardized to only contain bread with cheese. Soup, crisp bread or a big portion of dinner, might cause sound signals with completely different frequency response. The features used in this project were selected based on the available data set, and might not give a desirable result for other types of food or drinks. To create a system which detects different kinds of food, a variety of meals have to be recorded and included in the training process.

Even though the data set does not represent real-life scenarios, it indicates whether it is possible to identify meal related sound or not. Simplifications are needed to prove a concept before adding complexity. An additional simplification made in this study was to use data from the same subject for training and testing, as person specific detection is less complex. The variations among individuals are then canceled, giving a less complicated classification problem. However, this requires multiple recordings done on one each subject. In this study, only four recorded meals were used from each subject. The recordings have to be divided into training, validation and test recordings, giving only two recordings for the training and one for validation.

7.2 False detections

Even though the variation among different subjects is canceled, there will be a daily variation in bowel activity for each person. It is well known that the digestive system is influenced by both physical and mental health, as well as food or drinks consumed the past days. The varying amount of gas passing through the intestines is one example of factors which causes variations in the sound heard from the abdomen. In order to make a robust meal detection systems, the algorithm has to separate the meal-related sound from the general gastrointestinal sound. To overcome that challenge, the training data has to include data from all scenarios. Only two recordings for training, recorded within a few days, may not provide sufficient variation to detect all test meals. The lack of variation in the training set might be a reason why some meals fail to be detected more often than others.

As the results show, the meal in recording nr. 1 of subject 2, is detected incorrectly with both sets of features. Sometimes, only one false detection occurred before a correct detection, but a few times, multiple false detections occurred. The lack of variation in the training set might be the reason that this particular meal often fails to be detected. In recording 1 of subject 2, many clear rumbling sounds, or "hungry sounds", can be heard prior to the meal. Usually, the occurrence of bowel sounds increased after the meal, which made this recording deviate from the others. Because frequent "hungry sounds" only occurred for one recording, the network had no training data with similar sounds. The classifier might separate "hungry sounds" from meal related sound if it is included in the training set.

Another reason for false detections might be the mechanical attachment. If the stethoscope is loosely attached, the gastrointestinal sound will not be properly recorded. Small movements or changes in position might have resulted in periods with more or less tight contact between the microphone and the skin. A loose connection was difficult to prevent and notice during the recording session as the stethoscope head was covered with tape. However, all recordings used in this project were plotted to make sure that the sound signals were of proper amplitude and quality.

As seen in Figure 18, 19 and 20, the output of the neural network has, in close to all cases, a visible meal response even though there were a false detection. This indicates that the output of the neural network could have provided meal detection with higher accuracy if the threshold algorithm was changed. By filtering the output of the network, change the threshold, or change the number of segments which had to exceed the threshold, the number of false detection could have been improved. However, increasing the number of segments, which have to exceed the threshold, will cause later detections. Increasing the thresh-

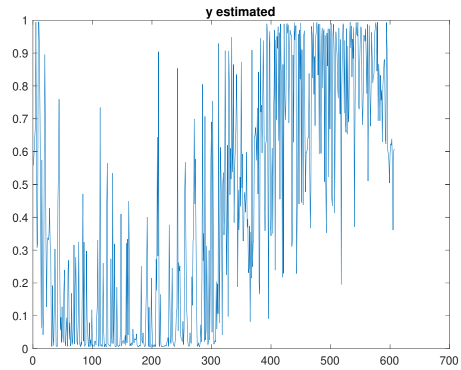


Figure 22: Output from the "meal"-class of the neural network for a late detection.

old value might cause that some meals are not detected. Possible improvements for the detection algorithm are further discussed in chapter 9.

To sum up, false detection might be caused by the lack of variation in the training data, making the classifier not generalize for daily variations. Variation in the recording could also be caused by a partly loose connection between the microphone and the skin. However, since the output of the network most often provide a meal output, which makes it possible to see when the meal happens, another detection algorithm might solve the problem with false detections.

7.3 Late detection

For feature set 1, two meals were detected more than 20 minutes after the meal onset (21 and 23 min). The late detected meals were the same as the ones which often were detected incorrectly. The network failed to give a clear estimate of the meal class, which resulted in late detections. Possible reasons for the poor meal estimation might be the same as for the false detections; lack of variation in the training set, noise, loose attachment of the stethoscope or incomplete training process.

7.4 Neural network

When using a neural network, there are a lot of hyperparameters to tune. For example, initial weights, number of neurons, number of layers, and the duration of the training. The initialization weights and biases were initialized by an initialization function containing a degree of randomness. Ideally, the initialization parameters which gave good results should have been investigated in order to find the optimal set of initialization parameters. Due to time constraints, there was no time to find this optimal set for this thesis. However, training and testing the classification system with three different initialization parameters gave a good indication of the performance.

The validation function used, stopped the training process early, typically after computing 12 epochs. Since the output of the network had a visible meal response in close to 100% of the test results, overtraining was prevented. However, due to the low number of recordings from each subject, only one recording could be used for validation. This might have caused that the training stopped too early if the validation set was very similar to the training set. Figure 20a shows one example of poor classification which might be caused by stopping the training process too early. Since the algorithm succeeded to detect the meals for two other sets of initializing parameters, the false detection in Figure 20a might be caused by a combination of inadequate initialization parameters and too short training process. Insufficient training might also be the reason of some false detections.

7.5 First set of features

MFCCs and the entropy calculated from a wavelet decomposition provided successful meal detection for both subjects. This indicated that the first set of features can be generalized to multiple individuals. However, more data, from a multitude of people, are needed before one can conclude if the features can be generalized to the population or not.

For subject 2, higher accuracy was obtained when extracting the training data for the meal class directly after the meal onset. For subject 1, a response delay of 10 minutes was used, and nothing else was tested due to the difficulties of avoiding noise for the "meal"-class in the training set. However, the early detection of several meals for both subjects indicates that the meal response starts shortly after starting to eat.

As previously mentioned, the features are extracted based on a dataset which is collected following a standardized procedure in a laboratory condition. Meaning that the same set of features might not provide a good result when the algorithm is applied to data containing noise or other meals.

7.6 Second set of features

As presented in the results in section 6.4, the second set of features did not provide good results of the meals eaten by subject 1. Only very few meals were detected, and most often, the meals were not detected at all. An increase in power in the lowest frequency bands, directly after the meal onset, were found in both subjects. However, the selection of frequencies was based on the features extracted from subject 2. In addition, the size of the network was tuned based on data from subject 2. In that way, the recordings from subject 2 were included in the training process. The selected frequencies might have been individual and the same set of features might be generalized to subject 1. Without medical expertise, it is hard to tell what causes this activity in the lowest frequency range. It might be the chewing frequency, which is personal and dependent on the food eaten.

Another reason might be the response delay used when extracting "meal"-data from subject 1. Meal-related training data were only extracted 10 minutes after the meal onset. The increase in power was observed immediately after the meal onset, and started to decrease after ca. 10minutes. Training data from the "meal"-class were then extracted after the characteristic response, which will not contribute to sufficient training of the network.

Another reason that the second set of features failed to detect meals from subject 1, is due to the noise and artifacts. The noise typically caused high spikes when plotting the features from subject 1. A comparison of the power in one frequency band(1-5Hz) is pictured in Figure 23. The four upper plots are the power extracted from subject 1, while the bottom ones correspond to subject 2. The recordings from subject 1 have a higher amount of spikes caused by noise. The orange line indicates the meal onset. A meal response is obvious in three of the four recordings done on subject 2. For subject 1, the meal response is less visible. This example of the feature set indicates that the poor meal detection obtained for subject 1 might be caused by noise, inappropriate features, or even a combination.

7.7 Integration of a meal detection system in an artificial pancreas

If a bolus of insulin is set due to detection of a meal, a high accuracy for the meal detection algorithm is required. Otherwise, the BG can drop below the critical level and cause hypoglycemia. Sound measurements alone might not provide the accuracy required as noise will always influence the signal. When integrating a sound based meal detection system in an AP, there will be other measurements available, such as CGM measurements and previous insulin boluses. A combination of all measurements can be used to obtain a safe AP, which avoids both hypoglycemias and elevated BG level after meals.

A suggestion from my supervisor is to use a fast and "unreliable" meal an-

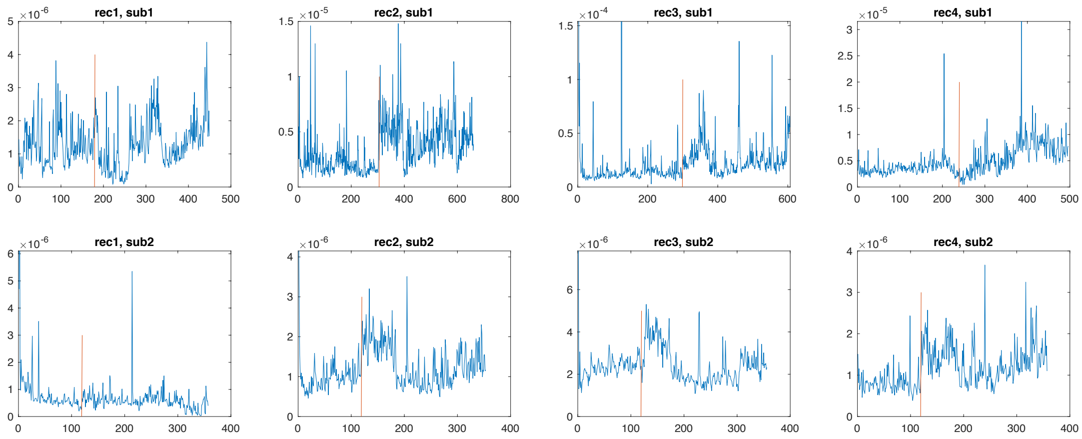


Figure 23: Comparison of power in 1-5Hz for both subjects. The yellow line indicates the start of the meal.

nouncement for a first insulin bolus while waiting for a more trustful detection by the CGM. After the first bolus, the basal infusion can be stopped in order to prevent hypoglycemia if the meal detection was incorrect. The total amount of insulin in the initial bolus will replace the basal insulin for a short period. No additional insulin will then be given until CGM measurements confirm a meal. If the meal classification were correct, the initial bolus will prevent the BG level rising too high, and an additional bolus can be infused when the BG starts to rise.

Another option is to introduce a filtered version of the meal estimation as a probability function in the BG control algorithm. The dosages might be increased if it is likely that the patient is eating. Previous boluses might be used to see the time of previous meals. If the previous meal was eaten a long time ago, and the sound based estimate indicates a meal, it is probably a correct meal estimate, and a higher dose of insulin might be infused.

Other measurements might also be added to the AP to obtain even better BG control. An example is pulse measurements or activity measurements from smartwatches. The activity level might be used to adjust the insulin dosage as glucose absorption changes according to the activity level. However, the complexity of the implementation will increase for each sensor added. Also, the user experience has to be in focus. Patients might not accept to wear multiple sensors attached to different places on their body.

7.8 Alternative applications

A sound-based meal detection system can also be advantageous in the treatment of nutrition-related diseases, e.g. obesity, eating disorders, and diabetes type 2. A wide range of approaches has been investigated for automatic food intake monitoring, as traditional note-based approaches rely heavily on user memory and recall [14]. Remembering is difficult for many, especially patients with memory disorders. A wearable device which continuously logs meal onsets, such as the sound-based system described, might be useful for this patient group.

8 conclusion

This thesis describes a method for person specific meal detection based on sound recordings from the abdominal wall. The long term aim is to include a non-invasive meal detection system in an insulin pump in order to make a closed-loop system for blood glucose control. An automatic insulin pump can improve the blood glucose administration for diabetes type 1 patients.

The two sets of features tested gave promising results for meal detection. The feature set using MFCCs and wavelet transformed based entropy provided good results for both of the two test subjects. Generally, meals were detected within a few minutes, which indicates that meal related sound appears shortly after the meal onset, or even immediately when starting to eat. The sound based meal detection obtained is significantly faster than meal detection based on CGM.

A few false positive results occurred. Suggestions for improvements and future work are described in chapter 9.

9 Suggestions for future work

The results of this project show a meal detection algorithm which aims to detect a meal within a few minutes in most cases. However, there is a lot of work remaining before such a system can be integrated into an artificial pancreas. There are also a huge number of signal processing techniques, feature extraction methods, and feature selection methods which can be tested for improvements. Suggestions for possible improvements of the method implemented and the integration of a meal detection system in an AP will be discussed in the following chapter.

9.1 Future recordings

More noise-free data is needed to test features and classification methods on different subjects. If person-specific data is used to train each classifier, multiple recorded meals are needed from more people. Otherwise, the features found to work for a few subjects, might not be generalized to fit for the users. In the future, recordings should be done on subjects of different body mass indexes to see how the meal-related sound signal changes due to different bodies. Future recordings should also include diabetes type 1 patients as they are the target users and might suffer from other gastrointestinal disorders as a consequence of the primary disease.

Future recordings should also contain a variety of gastrointestinal sound such that the training data contain more variation. It will then be possible to check whether the algorithm manages to separate, for example, hungry sounds from meal related sounds. Later, the recordings should be moved out from a laboratory condition to collect data in a natural environment which contains noise.

Various types of food and drinks should also be included in the future. It should be investigated whether sound measurements from the abdomen can be used to classify meals when eating different meals and if it is possible to differentiate different types of food (e.g. separate solid food from drinks) based on the sound measurements.

9.2 Preprocessing

Different segment length should be further investigated, to find the optimal segment length and overlap for meal detection.

9.3 Features

A combination of the two feature sets presented might be tested in the future. "Meal"-data should then be extracted directly after the meal onset for all features. If the two sets are combined, the network size and parameters have to be tuned to fit the current inputs. The two sets were of different scaling and might have to be normalized. A greater set of features might provide a more robust and versatile detection algorithm.

Features which separate hungry sounds from meal related sound should be investigated. It might be useful to extract hungry sounds and meal-related sound from the training set for closer inspection. Features which separates the two sounds, if possible, might be proper features for a meal detection algorithm.

Other preprocessing and filtering techniques should also be investigated in the future, especially when introducing noise to the system. Features which is less affected by noise might provide meal detection in real-life scenarios.

9.4 Future pattern recognition methods

The thresholding algorithm used in this thesis was not ideal as some false detections occurred even though a visible meal response was seen in the output of the network. Including the tuning of the threshold value and the number of successive segments in the training process might provide a better detection algorithm. The output of the neural network might also be filtered to avoid spikes causing false meal detections. Another alternative is to use a different network which will provide a crisp classification. A deeper neural network, e.g. a convolutional neural network (CNN), was suggested from my supervisor. This network utilizes information from neighboring inputs (e.g. previous segments of data) which might be advantageous when looking for a change in pattern.

Different training and initializing functions for a NN might be tested in the future. If a larger training set or a deeper NN is being used, a more efficient training and validation algorithm might be advantageous to save time on training. However, a slow training process will not cause any delay in meal detection. Other classification models which require less computational and memory resource might also be investigated in the future.

9.5 Implementation in an artificial pancreas

A question to ask is whether one classification model can be trained to detect meals on different subjects, as in the pilot trial described in [15], or if a classifier should be person-specific as in this thesis. Because of the high variability in the recordings done on different subjects, a tailored classifier might provide better accuracy. As long as the same feature extraction techniques can be used, a personalized classifier can be obtained by introducing a "training period" for the system. The microphone can collect data, while the user reports the onset of each meal for a period of time until there are enough training data. However, introducing a "training period" will require an extra effort by the user and demand a portable device with enough memory and processing resources. An evaluation of whether it is necessary and feasible to individualize a meal detection algorithm should be further investigated.

Making one classifier fit all subjects, will not require any effort by the user, and will maybe be experienced as an easier solution. It will be easier to integrate the system into an insulin pump as the same algorithm can be used. However, it might be challenging to make a fast and accurate detection system which is generalized for all individuals. Much training data will be needed in order to train such complex classifier.

When a meal detection system is included in an AP, other information from the pump might be used. A combination of other available measurements for a more accurate meal detection should be investigated. Sensors which are easy to add, such as an additional microphone for noise cancellation, smartwatch for activity measurements, etc. should be further investigated. However, one has to keep in mind that a complex sensor system might not be comfortable and accepted by the users. The use of microphones might cause privacy concerns. Other alternatives, e.g. piezoelectric sensors or accelerometers, should be investigated to find the best suited sensor which is most comfortable for the user.

The design and placement of the microphone should be investigated to find an attachment, location, and design, which makes the device comfortable to wear. A location close to the CGM sensor might be more accepted. If the discomfort is minimized, many patients will likely accept such a non-invasive add-on to their insulin pump.

10 Bibliography

[1] "Diabetes i norge," (Accessed on 23.4.2019), 2014. [Online]. Available: https://www.fhi.no/nettpub/hin/ikke-smittsomme/diabetes/

[2] F. Chee and T. Fernando, *Closed-loop control of blood glucose*. Springer, 2007, vol. 368.

[3] "Who, diabetes," (Accessed on 23.4.2019), october 2018. [Online]. Available: https://www.who.int/en/news-room/fact-sheets/detail/diabetes

[4] A. D. Association *et al.*, "Insulin administration," *Diabetes care*, vol. 26, no. suppl 1, pp. s121–s124, 2003.

[5] J. Pickup and H. Keen, "Continuous subcutaneous insulin infusion at 25 years: evidence base for the expanding use of insulin pump therapy in type 1 diabetes," *Diabetes care*, vol. 25, no. 3, pp. 593–598, 2002.

[6] R. Bourne, G. Stevens, R. White, J. Smith, S. Flaxman, H. Price, J. Jonas, J. Keeffe, J. Leasher, K. Naidoo *et al.*, "Causes of vision loss worldwide, 1990–2010: a systematic analysis. lancet glob health. 2013; 1 (6): e339–49," *This meta-analysis describes the prevalence and causes for visual impairment and blindness worldwide in various world regions in*, 2010.

[7] G. Van Den Berghe, P. Wouters, F. Weekers, C. Verwaest, F. Bruyninckx, M. Schetz, D. Vlasselaers, P. Ferdinande, P. Lauwers, and R. Bouillon, "Intensive insulin therapy in critically ill patients," *New England journal of medicine*, vol. 345, no. 19, pp. 1359–1367, 2001.

[8] C. A. Dimoulas, G. Papanikolaou, and V. Petridis, "Pattern classification and audiovisual content management techniques using hybrid expert systems: A video-assisted bioacoustics application in abdominal sounds pattern analysis," *Expert Systems with Applications*, vol. 38, no. 10, pp. 13 082–13 093, 2011.

[9] C. Vasseur, G. Devroede, D. Dalle, N. Van Houtte, E. Bastin, and R. Thibault, "Postprandial bowel sounds," *IEEE Transactions on Biomedical Engineering*, no. 5, pp. 443–448, 1975.

[10] N. Henry, N. Paul, and N. McFarlane, "Using bowel sounds to create a forensically-aware insulin pump system," in *Presented as part of the 2013 {USENIX} Workshop on Health Information Technologies*, 2013.

[11] 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, no. 9, pp. 753–757, 1990.

[12] T. Tomomasa, A. Morikawa, R. H. Sandler, H. A. Mansy, H. Koneko, T. Masahiko, P. E. Hyman, and Z. Itoh, "Gastrointestinal sounds and migrating motor complex in fasted humans," *The American journal of gastroenterology*, vol. 94, no. 2, p. 374, 1999.

[13] R. Ranta, V. Louis-Dorr, C. Heinrich, D. Wolf, and F. Guillemin, "Digestive activity evaluation by multichannel abdominal sounds analysis," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 6, pp. 1507–1519, 2010.

[14] T. Vu, F. Lin, N. Alshurafa, and W. Xu, "Wearable food intake monitoring technologies: A comprehensive review," *Computers*, vol. 6, no. 1, p. 4, 2017.

[15] K. Kölle, A. L. Fougner, R. Ellingsen, S. M. Carlsen, and Ø. Stavdahl, "Feasibility of early meal detection based on abdominal sound," *submission ID: JTEHM-00041-2019*, 2019.

[16] Y. Dong, A. Hoover, J. Scisco, and E. Muth, "A new method for measuring meal intake in humans via automated wrist motion tracking," *Applied psychophysiology and biofeedback*, vol. 37, no. 3, pp. 205–215, 2012.

[17] H. Kalantarian and M. Sarrafzadeh, "Audio-based detection and evaluation of eating behavior using the smartwatch platform," *Computers in biology and medicine*, vol. 65, pp. 1–9, 2015.

[18] 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, no. 3, pp. 806–816, 2016.

[19] J. M. Fontana and E. Sazonov, "Detection and characterization of food intake by wearable sensors," in *Wearable Sensors*. Elsevier, 2014, pp. 591–616.

[20] K. Kölle, M. F. Aftab, L. E. Andersson, A. L. Fougner, and Ø. Stavdahl, "Data driven filtering of bowel sounds using multivariate empirical mode decomposition," *Biomedical engineering online*, vol. 18, no. 1, p. 28, 2019.

[21] J. G. Proakis and Manolakis, *Digital signal processing: principles, algorithms, and applications.*, 3, Ed. Pearson Prentice Hall, 1996.

[22] S. Theodoridis and K. Koutroumbas, *Pattern Recognition*, 4, Ed. Elsevier, 2009.

[23] C. Altın and O. Er, "Comparison of different time and frequency domain feature extraction methods on elbow gesture's emg," *European journal of interdisciplinary studies*, vol. 2, no. 3, pp. 35–44, 2016.

[24] D. Wang, D. Miao, and C. Xie, "Best basis-based wavelet packet entropy feature extraction and hierarchical eeg classification for epileptic detection," *Expert Systems with Applications*, vol. 38, no. 11, pp. 14 314–14 320, 2011.

[25] T. S. Kumar and V. Kanhangad, "Detection of electrocardiographic changes in partial epileptic patients using local binary pattern based composite feature," *Australasian Physical & Engineering Sciences in Medicine*, vol. 41, no. 1, pp. 209–216, Mar 2018. [Online]. Available: https://doi.org/10.1007/s13246-017-0605-8

[26] M. M. Goodwin, "The stft, sinusoidal models, and speech modification," in *Springer Handbook of Speech Processing*. Springer, 2008, pp. 229–258.

[27] Y. Romanyshyn and V. Hudym, "Wavelet transforms applications for speech signals processing," in *Experience of Designing and Applications of CAD Systems in Microelectronics. Proceedings of the VI-th International Conference. CADSM 2001 (IEEE Cat. No. 01 EX473)*. IEEE, 2001, pp. 297–298.

[28] A. Bhowmick and M. Chandra, "Speech enhancement using voiced speech probability based wavelet decomposition," *Computers & Electrical Engineering*, vol. 62, pp. 706–718, 2017.

[29] Y. T. Chan, *Wavelet basics*. Boston: Kluwer, 1995.

[30] MathWorks, "Wavedec," (Accessed on 3.5.2019). [Online]. Available: https://se.mathworks.com/help/wavelet/ref/wavedec.html

[31] P. Pedersen, "The mel scale," *Journal of Music Theory*, vol. 9, no. 2, pp. 295–308, 1965. [Online]. Available: http://www.jstor.org/stable/843164

[32] A. M. Noll, "Cepstrum pitch determination," *The journal of the acoustical society of America*, vol. 41, no. 2, pp. 293–309, 1967.

[33] A. C. Kelly and C. Gobl, "The effects of windowing on the calculation of mfccs for different types of speech sounds," in *International Conference on Nonlinear Speech Processing*. Springer, 2011, pp. 111–118.

[34] R. Palaniappan, K. Sundaraj, and N. U. Ahamed, "Machine learning in lung sound analysis: a systematic review," *Biocybernetics and Biomedical Engineering*, vol. 33, no. 3, pp. 129–135, 2013.

[35] B. Logan *et al.*, "Mel frequency cepstral coefficients for music modeling." in *ISMIR*, vol. 270, 2000, pp. 1–11.

[36] K.-L. Du and M. N. S. Swamy, *Neural Networks and Statistical Learning*, 2014th ed. London: Springer London, 2014, vol. 9781447155713.

[37] P. De Wilde, "Neural network models : theory and projects," (Accessed on 01.03.2019), Berlin, Germany ;, 1997.

[38] B. McFee, "Statistical methods for scene and event classification," in *Computational Analysis of Sound Scenes and Events*. Springer, 2018, pp. 103–146.

[39] D. A. Simovici, "The k -means clustering," in *Linear Algebra Tools For Data Mining*. World Scientific Publishing Co. Pte. Ltd., 2012, pp. 769–791.

[40] A. Quintal, V. Messier, R. Rabasa-Lhoret, and E. Racine, "A critical review and analysis of ethical issues associated with the artificial pancreas," *Diabetes & metabolism*, 2018.

[41] S. C. Christiansen, A. L. Fougner, Ø. Stavdahl, K. Kölle, R. Ellingsen, and S. M. Carlsen, "A review of the current challenges associated with the development of an artificial pancreas by a double subcutaneous approach," *Diabetes Therapy*, vol. 8, no. 3, pp. 489–506, 2017.

[42] Z. Mahmoudi, D. Boiroux, and J. B. Jϕrgensen, "Meal detection for type 1 diabetes using moving horizon estimation," in *2018 IEEE Conference on Control Technology and Applications (CCTA)*. IEEE, 2018, pp. 1674–1679.

[43] K. Kvello, "Detection of bowel sounds," Master's thesis, NTNU, 2018.

[44] "Audacity," (Accessed on 22.04.2019). [Online]. Available: https://www.audacityteam.org/

[45] "Sennheiser mke2," (Accessed on 22.05.2019). [Online]. Available: https://en-us.sennheiser.com/clip-on-lavalier-microphone-live-speech-instrument-mke-2

[46] MFCC, (Accessed on 15.02.2019). [Online]. Available: https://se.mathworks.com/help/audio/ref/mfcc.html#mw_01a71d82-bbc3-4f06-bbf7-02a6d6d0e833

NTNU
Norwegian University of
Science and Technology

ARTIFICIAL PANCREAS TRONDHEIM