Human presence detection using a particle filter on ultrasonic micro-Doppler measurements for assisting rescue work in large buildings

by

Tom van Groeningen

to obtain the degree of Master of Science
at the Delft University of Technology,
to be defended publicly on Friday July 6, 2018 at 14:00.

Student number:	4205847	
Project duration:	October 1, 2018 – July 6, 2018	
Thesis committee:	Prof. dr. Geurt Jongbloed,	TU Delft
	Dr. Jakob Söhl,	TU Delft
	Dr. Ir. Hans Driessen,	TU Delft
	Ir. Robert Voûte,	CGI Nederland

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

TU Delft CGI

Abstract

When the fire brigade arrives at a burning building, it is of vital importance that people who are still inside can quickly be found. In this thesis we contribute to an ultrasonic sound sensor for human presence detection in smoke-filled spaces. This type of sensor could assist the fire brigade when evacuating a large building by directing them to the places where their help is most needed. The advantage of ultrasonic sound over other sensors or cameras is that its signal is able to pierce through smoke, does not require badges or other wearable devices and introduces little privacy and security risks. In addition, ultrasonic sensors are very inexpensive making it possible to equip every room of a building with an ultrasonic presence detector. In this research an ultrasonic sensor was built for less than 20 Euros and it was found to be unaffected by the glycerine based smoke that it was tested in. Using a particle filter based on sequential importance resampling as well as a filter based on Gaussian approximation of the posterior density the resulting system was reliably able to detect when there was a single person walking in the sensor direction, even when other sources of movement such as doors and chairs were present.

Preface

This research was performed on behalf of CGI Netherlands as a graduation project for the degree of Master of Science at the TU Delft at the faculty of applied mathematics. During this thesis I have had the privilege to be guided by no less than three very committed supervisors. First of all, Jakob Söhl who assured the quality of the mathematics I performed and consistently spotted every little incorrect minus in my calculations. Secondly, Robert Voûte who came up with the idea for this project and always kept the grand final application in mind. And finally Hans Driessen, who provided very valuable knowledge and intuition into Doppler measurements and signal processing. These three people were always ready to help me with their specific expertise without ever contradicting each other, which made working on this project a very pleasant experience.

Contents

1 Introduction

In fires in large buildings, an important factor that determines the strategy of the fire brigade is the presence of people inside the building. Choices like where to start extinguishing and whether to risk going into the building to search for people depend on this knowledge. Usually, the only information there is comes from incomplete sources such as statements from the people who fled the building or lists of employees. The largest portion of fatalities in fire emergency situations is not caused by a direct result of the heat, but rather poisoning by the smoke generated by the fire [17]. People suffering from the results of smoke are often unconscious, but could still be saved if the firemen can reach them in time. For instance, carbon monoxide, the most common poisonous gas in smoke from an indoor fire, may incapacitate a person within 1 or 2 minutes at an air concentration of 6400 ppm, but actual death from carbon monoxide may not occur for up to 20 minutes [6]. During this time, quickly finding people is of vital importance. If an overview of the building, indicating where people are, could be given so that rescue actions can be quickly and efficiently coordinated then this could potentially save lives.

Of course there are many sensors that could provide the information to create such a map, for instance infra red sensors, cameras or even Bluetooth badges. However, these sensors either do not work when there is smoke or require people to wear something, making them prone to failure when their information is most required and possibly also privacy sensitive.

In a search for a better sensor that can assist in crisis management, ultrasonic sensors showed to have some clear advantages. First of all, sound can pass through smoke, which is a minimal requirement for the sensor. Secondly, ultrasonic sensors are very inexpensive, allowing these sensors to be placed in every room of a building. Finally, the sensors pose fewer privacy and security risks since their measurements are limited to distance and movement and hence cannot be used to obtain images of people.

For application in emergency management the fact that only distance and movement are readily available means that there will need to be some data processing in order to find where and how many people there are. When there is a lot of smoke, people may quickly faint, posing the requirement that the sensor can keep track of people even when they are not moving any more. This requirement can be partly satisfied by tracking people while they are moving using distance measurement and remembering their location when they have stopped moving. For this, it is important to recognise whether the movement is of human origin in order to prevent other moving objects, such as doors or chairs from being tracked. This can be done by analysing the characteristic movement signature, often called Doppler signature [9], that people cause in the signal. In addition, when the Doppler signature of a person is modelled, it might be possible to detect when there are multiple people whose Doppler signatures overlap.

The focus of our work is on analysing the Doppler signatures in an ultrasonic echo in order to see how much information we can extract from them. The Doppler effect is usually analysed using a spectrogram. A spectrogram shows how strongly every frequency is present in a signal and how this changes over time. Together all the small movements of a persons and their associated frequency changes in the signal cause the Doppler signature in the spectrogram,

which allows discriminating between types of movement based on the ultrasonic signal. Previously, this technique has been used for classifying gait [9]. In [7] a particle filter has been applied in combination with the Thalmann model [2] that describes human walking for classifying Doppler signatures from radar measurements. In this thesis we will use the same approach, but focussed on recognising human presence in an emergency situation. We will try to make the particle filter work on the sonar data that we will measure using a sensor device that we built. In addition, we will extend the particle filter in order to be able to detect multiple people and try to improve its performance by using Gaussian approximations of the posterior density.

We will start this thesis by providing a little bit more information about the measurements that can be made using an ultrasonic sensor in section 2. After that, we have enough properties of the ultrasonic sensor to be able to justify a more detailed research approach. In section 3 we will outline this approach and distil it into research questions. We will then specify our measuring device in section 4 and the measuring method in section 5. After this, we have solid knowledge of what the data we are using looks like and we are able te define the particle filter in section 6.2. We will try to improve the particle filter by using a more advanced sampling scheme in section 7. In order to get an idea of the performance of the algorithms in a real situation, we will test them on measurements in section 8. In this section, we will also comment on the performance in smoke conditions. Finally, we can draw our conclusions in section 9.

2 Sonar measurements

The relevant properties of a reflected sound wave for sonar applications are signal strength, time of flight and frequency shift. Signal strength is an determined by the the size and distance of a reflecting object. Time of flight is used to determine the distance to an object and frequency shift indicates that an object is moving relative to the sensor causing the Doppler effect. For our quantitative analysis of these three properties, we will use the spherically symmetric wave equation. In reality, the spread of the sound is spherically as the speed of sound is equal in all directions. However, as can be seen in figure 3 from the datasheet of the sensor we have used, the amplitude of the wave may not be exactly spherically symmetric, especially for high angles with the normal to the sensor surface. The solution to the spherically symmetric wave equation, giving the pressure change $p(r,t)$ due to the sound wave is [10]

$$p(r,t) = \frac{A}{r}\sin\left(2\pi f_s\left(t - \frac{r}{c}\right)\right), \tag{1}$$

where r is the radial distance to the sensor, c is the speed of sound in air and f_s is the sending frequency of the source. Using equation 1 we will analyse the three signal properties.

2.1 Movement measurement

When an object moves relative to the sensor, its location keeps changing while reflecting the sound wave. Because of this, the reflected sound wave will be either stretched out (lower frequency) when the object is moving away from the

sensor or pressed together (higher frequency) when the object is moving towards the sensor. This phenomenon is know as the Doppler effect. We can derive the frequency shift in the reflected sound wave due to a moving object as follows. When the wave reflects at some (changing) distance $r_{obj}(t)$ from the sensor, the direction will be reversed, but the phase will remain the same. Ignoring the amplitude for now, this means that the returned wave at location 0 will behave like

$$\sin\left(2\pi f_s\left(t - \frac{2r_{obj}(t)}{c}\right)\right). \tag{2}$$

When $r_{obj}(t)$ is changing with constant velocity v, we can rewrite this as

$$\sin\left(2\pi f_s\left(t - 2\frac{r_{obj}(0) + v \cdot t}{c}\right)\right). \tag{3}$$

If we rearrange this to collect all the time variables, we get

$$\sin\left(2\pi f_s\left(1 - 2\frac{v}{c}\right)t - 2\pi f_s \frac{2r_{obj}(0)}{c}\right) \tag{4}$$

and we see that the returned wave has frequency $f_r = f_s\left(1 - 2\frac{v}{c}\right)$ and phase shift $-2\pi f_s \frac{2r_{obj}(0)}{c}$. This means that we can calculate the velocity of an object in terms of the received frequency f_r as

$$v = \left(1 - \frac{f_r}{f_s}\right)\frac{c}{2}. \tag{5}$$

2.2 Distance measurement

Distance to an object can be derived from the time it takes a signal that has been sent out travel to the object and back again. This time is usually called the time of flight (ToF). Since sound has a constant velocity in air, the radial distance r_{obj} can simply be calculated from the time of flight t_{ToF} by

$$r_{obj} = c \cdot \frac{t_{ToF}}{2}. \tag{6}$$

The precise determination of the time of flight is the most difficult part. The simplest method to determine this time is by setting a threshold on the signal strength and measuring the time between sending the signal and the threshold being crossed. This method works well when there is not a lot of noise but as soon as noise is high or there is a lot of interference due to reflections by other objects, there are better ways to calculate the time of flight. The method that is generally used, often referred to as matched filtering, is finding the maximum of the cross correlation between the sent signal and the received signal[11]. This maximum then occurs at the time difference between the sent and received signal and is much less influenced by other signals as long as they are different from the sent signal. The calculation for the time of flight from the sent signal $S(t)$ and the received signal $R(t)$ then becomes

$$t_{ToF} = \arg\max_{t>0} \int_{\infty}^{\infty} S(\tau)R(\tau + t)\, d\tau. \tag{7}$$

2.3 The reflection strength

The strength of the reflection depends on the size, shape, material, orientation and distance of an object. Therefore it is difficult to derive one of these quantities from just the signal strength. When the distance r_{obj} is known, we can separate that from the equation and summarise the other four variables into a single variable a, which we will simply call the reflection coefficient of the object and define to be the ratio between the power density I of the wave reaching the object and the power density I_r of the signal reflected towards the sensor:

$$a = \frac{I(r_{obj})}{I_r(r_{obj})}. \tag{8}$$

The size, shape and material of an object usually do not change but we may sometimes let the reflection coefficient depend on the orientation of the object. For now we will assume that the orientation is constant.

If no losses would occur in the medium itself, the power passing trough a spherical area around the source would equal the power of the source itself because of the preservation of energy[10]. If the propagation of sound is spherically symmetric then this automatically implies that the power density at radius r must equal

$$I(r) = \frac{P_s}{4\pi r^2}. \tag{9}$$

We know from figure 3 that the energy distribution throughout the wave is not exactly spherically symmetric, but since the propagation itself is spherically symmetric, we can still use equation 9 to observe that the intensity is proportional to $1/r^2$. At least, that would be, if the air was a lossless medium. However, air is not a lossless medium and at typical indoor conditions of 50% humidity and temperature of 20 degrees Celsius, the attenuation of a sound wave at $41Khz$ is about $1.3db \cdot m^{-1}$ [18]. This means that the power density is in practice proportional to

$$I(r) \propto \frac{1}{r^2\, 10^{0.13r}}, \tag{10}$$

where r must be in meters. When we analyse a reflection, the same attenuation will happen on the way back. Hence the received reflection will have been attenuated by a factor of $r^{-4}\, 10^{-0.26r}$ due to the distance only.

3 Research approach

The main goal of this research is to assess whether an inexpensive ultrasonic sensor can be used for human presence detection in a building that is filled with smoke. Now that we have some information about the possible measurements that we may make with such a sensor, we can decide what approach to take. We will first provide a rationale to explain the choices we made for the setup of this project. After this we will summarise the set up of this thesis concisely in the form of a few research questions. These questions will lead the direction of this thesis.

3.1 Rationale

In order to filter the signal, we can choose between creating a model based filter, or taking a machine learning approach. We have decided to focus on a model based approach because it allows more understanding into when the filter will perform as required, and in which situations it will have difficulties finding people. If we would take a machine learning approach then this would depend mostly on the training data that we would give the model and it is possible that unexpected situations may give unexpected results. When we use the modelling approach then we can qualitatively estimate the performance of the filter by assessing how well the used model covers the situation at hand. In a system that will be used in emergencies, it is very important to know the shortcomings of the system before any emergency situation occurs and therefore the modelling approach is preferred.

Regarding the measurements that could possibly be made using an ultrasonic system, we have chosen to focus only on the signal strength and the Doppler effect. This means that it is not possible to extract distance information from the sensor as this would require the time of flight as well. It can be expected that using time of flight information in order to infer the distances of objects could increase the accuracy of the system. However, doing time of flight measurements requires sending more complex signals and performing a different analysis of the received signal. This would increase the size of the project too much and it is therefore not feasible to do both distance and movement measurements within this thesis. Since this thesis is performed for applied mathematics, we have chosen to use movement measurement since it can use a simpler measuring system, but requires more complex algorithms to extract information from the signal.

In order to asses whether an inexpensive ultrasonic sensor can be used for human presence detection in a building that is filled with smoke we will build a preliminary prototype of the sensor. Once we have the sensor, we are able to assess visually whether the presence of a person can be recognised from the signal. When this is the case, we will try to recognise human movement. Since we are taking the modelling approach for this research, we will need to restrict the kind of movement we want to filter for now because otherwise the model becomes to complex. we have chosen to filter walking movement since this is a very common type of movement in any building and has distinct characteristics in the signal. In order to test the limits of what the filter can recognise, we will also try to find the distinction in the signal between one and two persons.

The most realistic tests can be performed on real measurements. We will try to cover all situations that are described by the model and in addition, test how the filter responds to situations that are slightly different than the ones described by the model, for instance by measuring other types of movement. Finally, we will also test in smoke conditions, because being able to make the same measurements in smoke is a very important condition for this sensor to be usable in emergency situations.

3.2 Research questions

Guided by the above rationale, we have set up the following research questions for this project:

Main question: To what extend can an inexpensive ultrasonic sensor be used for human presence detection in a building that is filled with smoke based on Doppler measurements analysed with a particle filter?

- To what extend can a particle filter be used to estimate the presence of people walking through a room using the signal spectrogram as input?
  - How reliably can the particle filter estimate when a single person walks through the room?
  - To what extend is the particle filter able to "see" the difference between the signal of a single person and two persons?
  - To what extend is the particle filter able to discern walking from other common types of movement in an office environment such as opening doors and moving chairs?
  - How does approximation of the posterior density by Gaussian densities affect the performance of the particle filter?
- How does smoke affect the measurements we can make with the system?

4 Definition of the used device

We will shortly define the device we have built in order to experiment with ultrasound measurements. The device basically consists of an ultrasonic sender and receiver, both connected to an amplifiers. The generation of the sending signal and the analogue to digital conversion of the received signal are performed by the sound card of a laptop with a sampling frequency of 96KHz.

Figures 1 and 2 show the electrical circuit and the prototype breadboard layout. The same amplifiers are used for the input and output, with the difference being that the amplifier of the received signal has more small capacitors that suppress noise and the output of the sending amplifier is connected in series with a larger capacitor to the transmitter in order to filter out DC signals and let higher AC signals pass through to the transmitter.

The amplifiers used are LM386 integrated circuits created by Texas Instruments, having a default amplification of 20 times. The receiver and transmitter we used are the Murata MA40S4R and MA40S4S respectively. These devices are sensitive to sound in a narrow band around 40KHz. We found that the frequency with the highest output was 41KHz in the devices we obtained. The advantage of the MA40S4R/S is that the sender requires relatively low input voltages (maximum 20V) and the directivity of 80 degrees is quite wide. Figure 3 from the datasheet of the sensors shows how the sound intensity changes as a function of the angle away from the normal.

If we exclude the laptop used for the signal processing, the hardware is not expensive. The sender and receiver have a retail price of approximately 6 Euros each and the two audio amplifiers cost about 1 Euro per amplifier. Combined with some resistors and capacitors, the total price of all the components is less than 20 Euros.

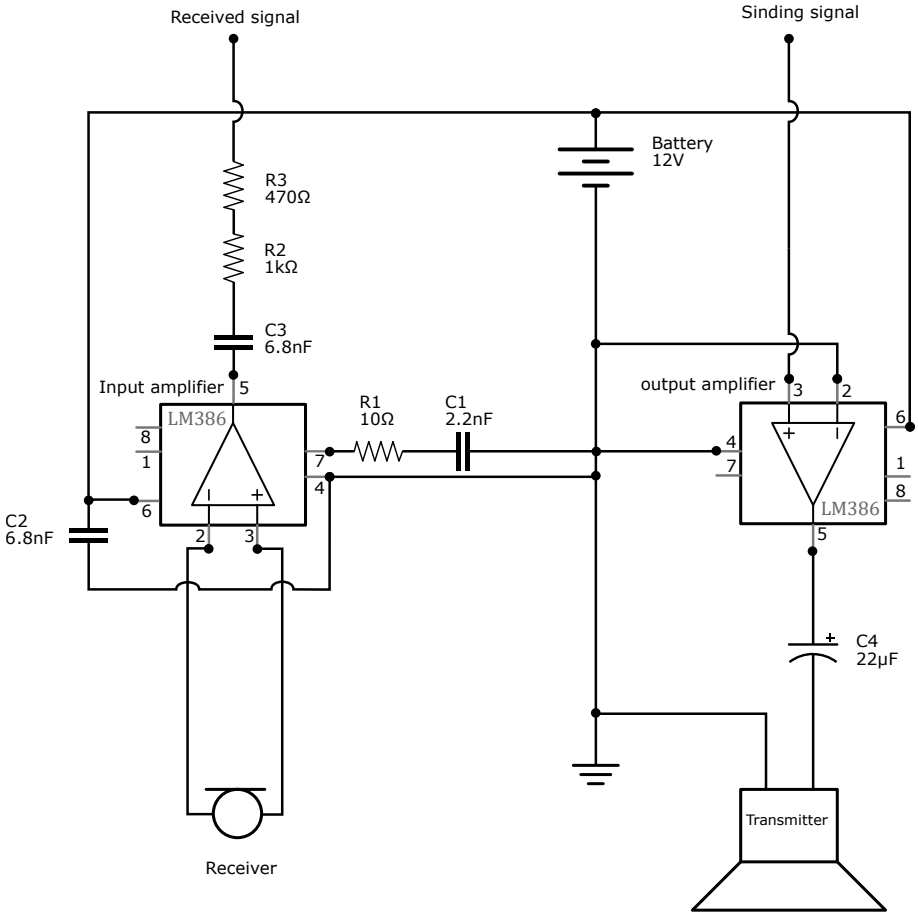


Figure 1: Schematic of the prototype measuring device used for this thesis.

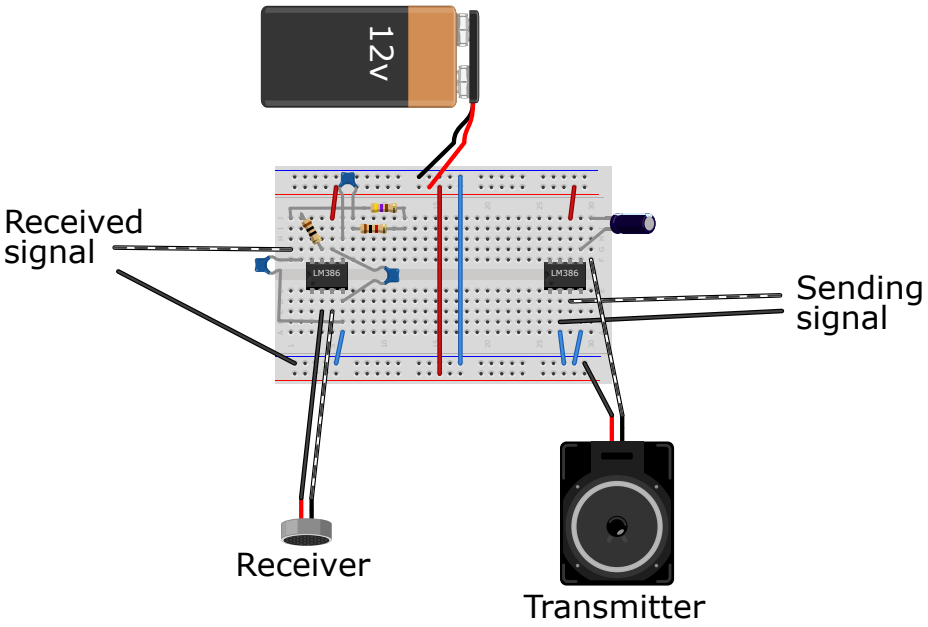


Figure 2: Overview of the prototype breadboard created using Fritzing.

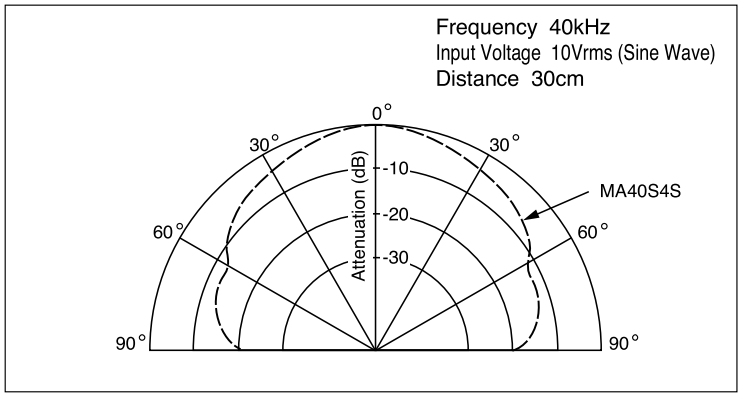


Figure 3: Figure 11 from the transmitter datasheet[15] shows that the sound intensity is approximately spherically symmetric when the angle from the normal is not too large.

5 Making measurements

5.1 Sampling

In order to obtain a digital representation of an acoustic signal, such as received from the ultrasonic sensor, we need to sample the signal. This means that we measure the signal at constant intervals and save the amplitudes of the signal that are measured over time. The well known sampling theorem by Harry Nyquist specifies that in order to save a signal without loosing any information, the sampling frequency should be at least twice the highest frequency that is present in the signal.

5.2 Analysing the signal

We want to be able to extract information about movement from the signal. This movement is shown by a frequency change in the signal. Therefore, we want to analyse the frequencies that are present in the signal over time. The frequencies in a signal are generally analysed using the Fourier transform of the signal. However, taking the Fourier transform leaves us only with frequency information and no time information. The way we can get both time and frequency information is by Fourier transforming short time intervals of signal so that we get frequency information for each interval. When we square the Fourier coefficients, the graph that this generates is usually called a spectrogram. When we use this approach, we have to make a choice between frequency resolution and time resolution because the frequency resolution of a discrete Fourier transformation is inversely proportional to the length of the time signal as $\Delta f = 1/\Delta t$. Therefore we have to choose an acceptable value for the time interval that captures the time changing nature of human movement as well as the velocity components in this movement. In addition, when the window size multiplied by the sampling frequency is a power of two, the Fourier transform can be performed very efficiently. By varying the time window, we have established that we can see most details in the Doppler signal when using a time window of 16384/96000 seconds, which is approximately 0.17 seconds.

Since the discrete Fourier transform is actually defined for an infinite time signal, by transforming a finite part of the signal, we implicitly assume that this signal is periodic with the period being the length of the time window we used. When we have a sine signal that fits an integer number of times into our time window, this assumption is perfectly fine and we will see a very narrow spike in the Fourier transform at exactly the frequency of the sine. However, as soon as the signal we analyse has a different period, or is not periodic at all, the assumed periodicity will cause discontinuities in the signal. The result of this is that the frequencies in the signal "leak" into other frequencies and cause artifacts in the spectrogram. This behaviour means that for any noisy real-world signal, it is beneficial to multiply the signal by a "window function" that makes the signal "more periodic" by letting the signal approach zero at the endpoints of the time window. There are many different window functions designed for this purpose which all have their own advantages and disadvantages but we will use one that is very generally applicable, called the "Hann" window defined by

$$w(n) = \frac{1}{2}\left(1 - \cos\left(\frac{2\pi n}{N-1}\right)\right), \tag{11}$$

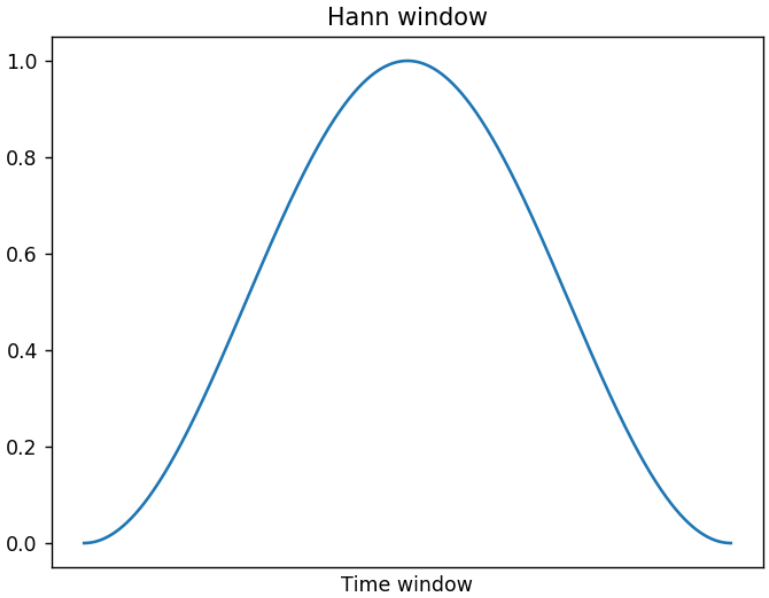


Figure 4: The Hann window

where N is the number of samples in the time window we choose.

Below two spectrograms of the same signal can be found, made using different window functions. The spectrogram in figure 5 was created without applying a window function, which is implicitly a rectangular window, and the spectrogram in figure 6 was made using the Hann window. In both spectrograms the frequency on the y axis has been transformed to velocity using equation 5. We can see that the Hann spectrogram contains much more detail and discriminates the signal much better from the background noise. The artifacts in the form of vertical lines are very well visible in the spectrogram that was made using the rectangular window.

A potential downside of using a window function is that some information is lost since the center part of the signal in a time window gets a higher weight in the Fourier transform than the edges. However, this can be resolved by taking overlapping time windows so that, in a way, each piece of the signal is used equally much. It turns out that this is very easily done using the Hann window by taking half a window overlap. We can see this by calculating the summed contribution of a sample at time n to the two windows in which it lies:

$$\frac{1}{2}\left(1-\cos\left(\frac{2\pi n}{N-1}\right)\right)+\frac{1}{2}\left(1-\cos\left(\frac{2\pi[n-(N-1)/2]}{N-1}\right)\right) \tag{12}$$

$$=\frac{1}{2}\left(2-\cos\left(\frac{2\pi n}{N-1}\right)+\cos\left(\frac{2\pi n}{N-1}\right)\right)=1 \tag{13}$$

independently of n and hence every sample gets exactly equal weight in the spectrogram.

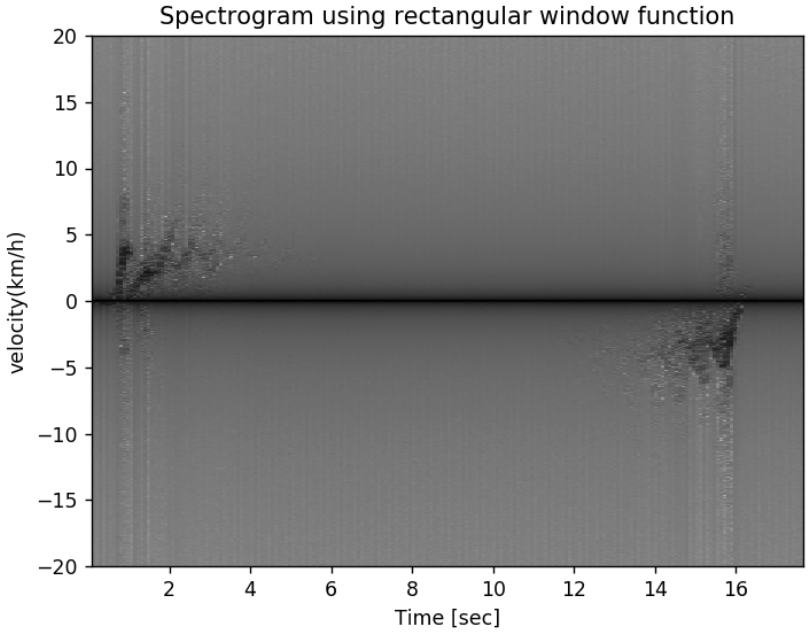


Figure 5: Spectrogram of a walking person calculated using rectangular window. The quality is clearly not so good.

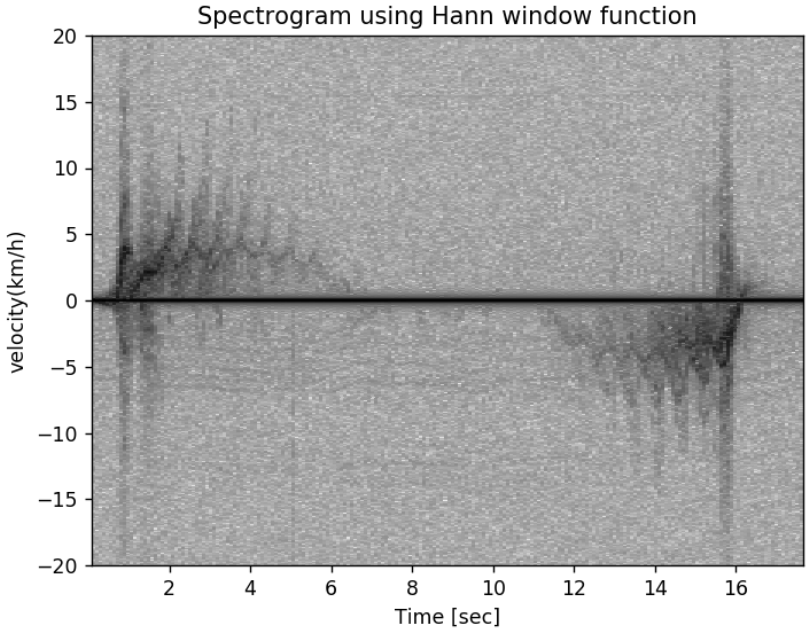


Figure 6: Spectrogram of a walking person calculated using Hann window.

5.3 The signal and the noise

We will describe now how exactly a recorded tone shows up in the Fourier transform. For each Fourier coefficient we can separate its complex value into an amplitude and complex argument. The amplitude reflects how strong a frequency is present in the signal and the argument gives the phase of the associated sinusoid. Since we are consecutively taking the Fourier transform of small pieces of signal in time, we will see that the phase keeps changing and hence the complex value rotating in the complex plane. Predicting the phase is nearly impossible in our set up because it depends on the length of the wave and there are usually many different reflectors of which we do not know their distance. Even if we knew the distance, a distance error of 1 cm would already present a phase error of more than a full period. Therefore we choose to ignore the phase information by analysing only the amplitude of the Fourier transform. When we analyse the amplitude of the Fourier coefficients, we will see that a single frequency always spreads out a bit even though we have chosen a window function to minimise this effect. However, we can easily predict how this spread will take place by calculating the Fourier transform of a sine multiplied with the window function we use. Figure 7 shows how this theoretical response relates to a measured response. Figure 7 also reveals that there is still some noise in

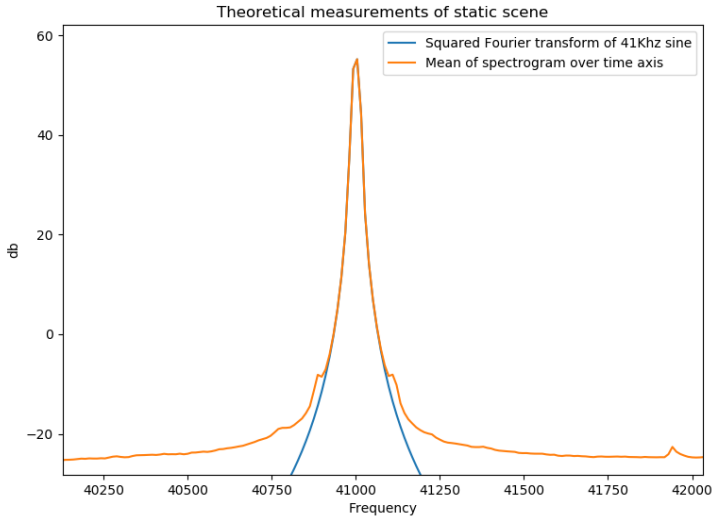


Figure 7: Squared Fourier coefficients (in db) of a 41KHz sine wave. The orange line is the mean of many measurements on a static scene and the blue line is the theoretical Fourier transformation scaled to match the height of the measured curve. The Hann window has been used in both cases.

the signal that we need to model. A reasonable assumption for the noise in the signal is additive Gaussian noise. When we translate this into the frequency domain, we get complex Gaussian noise added to the deterministic part of the Fourier coefficients caused by the signal. If we look at just the noise, then this has a two dimensional normal distribution with equal variance in the complex

and real planes and zero covariance. When we square the Fourier coefficients to remove the phase, we will have the sum of two independent squared normals which gives a chi square distribution, scaled by the variance (since chi square assumes standard normal). When there is a signal, the coefficients will also have a deterministic part and hence the mean of the two normals is not 0 any more. In this case, the measurement with noise can be modelled by the noncentral chi square distribution, which describes the squared sum of normals with variance 1 and nonzero mean [14]. Suppose that $C = X + iY$ is a complex Fourier coefficient with added Gaussian noise and that $\mu = \mu_x + i\mu_y$ is its true value so that

$$X \sim \mathcal{N}(\mu_x, \sigma^2) \tag{14}$$

$$Y \sim \mathcal{N}(\mu_y, \sigma^2). \tag{15}$$

Then $Z := |C|^2/\sigma^2$ has a noncentral chi square distribution. The pdf of this distribution is given by

$$f_Z(z; \delta, n) = \sum_{k=0}^{\infty} \frac{e^{-\delta/2}(\delta/2)^k}{k!} \frac{e^{-z/2} z^{(n+2k)/2-1}}{2^{(n+2k)/2}\Gamma((n+2k)/2)}, \tag{16}$$

where $\delta = (\mu_x^2 + \mu_y^2)/\sigma^2$ and n is the degrees of freedom, which is 2 in this case. Also,

$$\mathbb{E}(Z) = n + \delta \tag{17}$$

and

$$\mathrm{Var}(Z) = 2n + 4\delta. \tag{18}$$

6 Estimating the system parameters

6.1 The stochastic process

We will use a hidden Markov model with a continuous state space to describe our measurements. In this model, we have a system which is, at each time $k \geq 1$, defined by a set of P parameters $(X_{k,1}, ..., X_{k,P}) = X_k$. At each time step this system results in an observation Y_k which is available to us as a measurement. From these observations we would like to estimate the system parameters. For this we will use a Bayesian approach to sequentially update our belief in the parameters as we receive new measurements. The model is described stochastically as follows:

$$f_{X_1} \tag{19}$$

is the prior pdf of X when no data is available.

$$f_{X_k|X_{k-1}} \tag{20}$$

is the pdf that describes how X evolves over time. In addition, we assume that the Markov property holds:

$$f_{X_k|X_{1:k-1}} = f_{X_k|X_{k-1}}, \tag{21}$$

where we use the notation $X_{1:k-1} = X_1, ..., X_{k-1}$. Finally, we relate the system to the measurements by the density function of the measurements given the system parameters

$$f_{Y_k|X_k}. \tag{22}$$

Using these distributions we can use Bayes rule to do inference on the system. The posterior we want to obtain at each time step is

$$f_{X_{1:k}|Y_{1:k}}. \tag{23}$$

At the first measurement we can calculate this posterior using Bayes rule as

$$f_{X_1|Y_1}(x_1; y_1) = \frac{f_{X_1}(x_1) f_{Y_1|X_1}(y_1; x_1)}{\int f_{X_1}(x_1) f_{Y_1|X_1}(y_1; x_1)\, dx_1}. \tag{24}$$

Then we can update the posterior at each new measurement by first building the new prior based on the previous measurements

$$f_{X_{1:k}|Y_{1:k-1}}(x_{1:k}; y_{1:k-1}) = f_{X_{1:k-1}|Y_{1:k-1}}(x_{1:k-1}; y_{1:k-1}) f_{X_k|X_{k-1}}(x_k; x_{k-1}) \tag{25}$$

and then calculating the posterior as

$$f_{X_{1:k}|Y_{1:k}}(x_{1:k}; y_{1:k}) = \frac{f_{X_{1:k}|Y_{1:k-1}}(x_{1:k}; y_{1:k-1}) f_{Y_k|X_k}(y_k; x_k)}{\int f_{X_{1:k}|Y_{1:k-1}}(x_{1:k}; y_{1:k-1}) f_{Y_k|X_k}(y_k; x_k)\, dx_{1:k}}. \tag{26}$$

The nice thing about this way of updating is that we can proceed from $f_{X_{1:k-1}|Y_{1:k-1}}$ to the next posterior at each step when new data becomes available simply by multiplying with $f_{X_k|X_{k-1}}\, f_{Y_k|X_k}$ and normalising. Hence, ignoring the normalisation we have the easy relation

$$f_{X_{1:k}|Y_{1:k}} \propto f_{X_{1:k-1}|Y_{1:k-1}} \cdot f_{X_k|X_{k-1}}\, f_{Y_k|X_k} \tag{27}$$

In practice, this distribution quickly becomes too complex to do analytic calculations with so we will resort to Monte Carlo like methods. The algorithm that naturally arises in this case is often referred to as the particle filter.

6.2 The particle filter

6.2.1 Importance sampling

The particle filter, also described as sequential Monte Carlo filter or sequential importance sampling is based on importance sampling and comes from the following idea [3]. We would like to sample from the posterior in equation (26) in order to do inference. However, we cannot directly do this because the posterior has a difficult distribution. We can resolve this by using importance sampling to sample from an easier distribution and assign weights to the samples to correct for the discrepancy. An easy choice for an importance distribution is the unconditioned distribution that does not take into account the data,

$$f_{X_{1:k}}(x_{1:k}) = f_{X_1}(x_1) \cdot f_{X_2|X_1}(x_2; x_1) \cdot ... \cdot f_{X_k|X_{k-1}}(x_k; x_{k-1}). \tag{28}$$

From this distribution it is easy to update the importance samples $\{X_k^1, ..., X_k^N\}$ at each new time step by sampling recursively

$$X_k^i \sim f_{X_k|X_{k-1}}(\,\cdot\,; X_{k-1}^i). \tag{29}$$

The importance weights of the samples we generate in this way are then given by

$$\omega_k^i = \frac{f_{X_{1:k}|Y_{1:k}}(X_{1:k}^i; Y_{1:k})}{f_{X_1:k}(X_{1:k}^i)}. \tag{30}$$

Using the updating equations (27) and (28) we get that

$$\omega_1^i \propto \frac{f_{X_1}(X_1^i) f_{Y_1|X_1^i}(Y_1; X_1^i)}{f_{X_1}(X_1^i)} = f_{Y_1|X_1}(Y_1; X_1^i) \tag{31}$$

and can be updated like

$$\omega_k^i \propto \frac{f_{X_{1:k-1}|Y_{1:k-1}}(X_{k-1}^i; Y_{k-1}) \cdot f_{X_k|X_{k-1}}(X_k^i; X_{k-1}^i)\, f_{Y_k|X_k}(Y_k; X_k^i)}{f_{X_{1:k-1}}(X_{k-1}^i) \cdot f_{X_k|X_{k-1}}(X_k^i; X_{k-1}^i)} \tag{32}$$

$$= \omega_{k-1}^i \cdot f_{Y_k|X_k}(Y_k; X_k^i). \tag{33}$$

That gives us all the tools to do Monte Carlo importance sampling. However, it turns out that in practice, the use of the unconditioned distribution causes the samples to badly represent the true distribution when more measurements are received. Some of the weights will simply converge to 0 and add nothing to any estimate, but still consume computation time. A well known method to overcome this phenomenon, known as depletion, is to resample from the samples we drew [12] (also called particles in this context).

6.2.2 Resampling

Suppose we have drawn particles $\{x_k^1, ..., x_k^i\}$ from the importance distribution then we may use them together with their weights to estimate some expectation

$$\sum_{i=1}^{N} \omega_k^i \, g(x_k^i) \approx \mathbb{E}[g(X_k)], \tag{34}$$

where g is just some function we are interested in. Instead, we could also assign probabilities ω_k^i to the particles and then draw from the particles we already have. Suppose we draw one such a "new" sample, say $\tilde{X}_k$, then the expectation of the function g of this variable is

$$\mathbb{E}[g(\tilde{X}_k)] = \sum_{i=1}^{N} \omega_k^i \, g(x_k^i) \approx \mathbb{E}[g(X_k)]. \tag{35}$$

Hence we may average over many samples from $\tilde{X}^k$ to approximate the approximation of $\mathbb{E}[g(X_k)]$. The double approximation already reveals that the variance of this estimate will be theoretically higher, but the fact that only the particles that had a high likelihood under the measurements will "survive" the resampling compensates for the depletion and in practice gives better estimates. In a way (that can be made precise, see [13]), the particles and their weights now form an approximation to the distribution we want to sample from. The particle filter algorithm with the resampling step us usually called the sequential importance resampling filter, or SIR filter.

6.3 The particle filter algorithm

- Sample $\{x_1^1, ..., x_1^N\}$ from $\{X_1^1, ..., X_1^N\} \sim f_{X_1}$ i.i.d..
- Calculate the unnormalised weights of the particles using the measurement y_1 at time 1:
$$\tilde{\omega}_1^i := f_{Y_1|X_1}(y_1; x_1^i) \quad \text{for } i = 1, ..., N. \tag{36}$$
- Normalise the weights to get
$$\omega_1^i = \frac{\tilde{\omega}_1^i}{\sum_{j=1}^N \tilde{\omega}_1^j}. \tag{37}$$
- For time $k = 1, 2, ...$:
  - **Resampling**
    Draw $M = (M^1, ..., M^N)$ from a multinomial distribution with N trials and probabilites ω_{k-1}^i. M now represent the number of times we re-use each particle. Note that some particles may by discarded altogether by this process but we will always end up with a total of N particles.
  - **Importance sampling**
    Sample new particles $\{x_k^1, ..., x_k^N\}$ by drawing M^i samples from $f_{X_k|X_{k-1}}(\cdot\,; x_{k-1}^i)$ for each $i \in \{1, ..., N\}$.
  - Calculate the unnormalised weights of the new particles using the measurement y_k at time k:
$$\tilde{\omega}_k^i := f_{Y_k|X_k}(y_k; x_k^i) \quad \text{for } i = 1, ..., N. \tag{38}$$
  - Normalise the weights to get
$$\omega_k^i = \frac{\tilde{\omega}_k^i}{\sum_{j=1}^N \tilde{\omega}_k^j}. \tag{39}$$

  Then using the particles and their weights it is possible to estimate expectations at any time step.

6.4 The system model

The system we are trying to do inference on is a walking person that generates a Doppler signal. We will assume that the measured quantity is the spectrogram. In order to model how a walking person is reflected in the spectrogram, we calculate the velocities of the different body parts and translate them to Doppler shifts in the ultrasonic signal.

In order to calculate the velocities of the human body parts, we use the Thalmann model for human walking motion [2]. This model gives us the locations of the different body parts over time and hence allows us to calculate their velocity with respect to the sensor. The precise calculations of the Thalmann model can be found in appendix B. The amplitude of the signal these velocities give is then determined by the size of the body parts, or more general their reflection coefficients, and their distance. More information about this can be found in section 2. The velocities and amplitudes can then be calculated from the model using the following parameters.

- The walking velocity V.
- The persons height H.
- The walking phase ϕ.
- The radial distance r.
- The presence of a person p.

Where clearly, p being false makes the other parameters irrelevant. As can be deduced from equation (5) the relation between the velocity v of a sound reflecting object and the received frequency f_r of a signal that was sent with frequency f_s is

$$f_r = f_s \frac{c - v}{c + v}, \tag{40}$$

where c is the velocity of sound in air. In the spectrogram, we won't see this exact frequency due to a several reasons. First of all, we measure the spectrogram over a finite time interval. Therefore the velocity may change during this interval. Secondly, we describe the body parts as having a single velocity, but in reality the parts are rigid objects whose radial velocity with respect to the sensor may vary over their surface. Finally, the discrete Fourier transform itself causes each frequency to spread out a bit. More information about this last phenomenon can be found in section 5.2. In order to model this combination of "spread factors", we will add all velocities of the b body parts $v_i,\ i = 1, ..., b$ from the Thalmann model with amplitudes A_i to the spectrogram as

$$\sum_{i=1}^{b} A_i h(v - v_i), \tag{41}$$

where h is the spread function. We model h as a Gaussian with spread s like

$$h(v) = \exp(-\frac{v^2}{s^2}), \tag{42}$$

The Gaussian spread is a safe choice because there are many different phenomena affecting the distribution of the frequency in the spectrogram. The typical maximal acceleration for a person is $2ms^{-2}$ [8]. Multiplied by the length of a time window of about 0.17 seconds, this results in a maximum variation of about $0.34ms^{-1}$ within a time frame. We choose to set s to half this value:

$$s := 0.17 \tag{43}$$

The prior distribution we assume for the parameters when estimating a single persons is as follows:

$$V_0 \sim U([0.5, 1.7]) \tag{44}$$
$$H_0 \sim \mathcal{N}(1.7, 0.07^2) \tag{45}$$
$$\phi_0 \sim U([0, 1]) \tag{46}$$
$$r_0 \sim U([2, 6]) \tag{47}$$
$$p_0 \sim \text{Bernoulli}(0.5). \tag{48}$$

For the time evolution with a time step of Δt, we assume the following distributions on the parameters: $t_k = 1.346 \cdot \sqrt{0.53 H_k / V_k}$ is the period of the walking cycle according to the Thalmann model.

$$V_k \sim \mathcal{N}(V_{k-1}, (0.6\Delta t)^2) \tag{49}$$
$$H_k \sim \mathcal{N}(H_{k-1}, 0.05^2) \tag{50}$$
$$\phi_k \sim U(\Delta t / t_{k-1} + [-0.1, 0.1]) \tag{51}$$
$$r_k \sim \mathcal{N}(V_{k-1} \cdot \Delta t, 0.01^2) \tag{52}$$
$$p_k \sim \begin{cases} \text{Bernoulli}(0.9) & \text{if } p_{k-1} = 1 \\ \text{Bernoulli}(0.1) & \text{if } p_{k-1} = 0 \end{cases} \tag{53}$$

When we want to use the model for two persons, we simply add another independent copy of each of the stochastic parameters and sum the resulting evaluations of the Thalmann model on both parameter sets. The only parameter that we do not duplicate is p_k. In the two persons case, the prior distribution on p_k will be

$$U\{0, 1, 2\}, \tag{54}$$

a discrete uniform distribution over the numbers 0, 1 and 2. The evolution of p_k is then as follows: With 80% chance, p_k remains the same and with 10% and 10% chance p_k changes to one of the other two values.

Apart from the model for walking persons, we would also like to have a model that can capture other types of movement that are not human walking. We will call this 0 person state the null hypothesis. For this, we will use a very simple model consisting of constant Gaussian peaks in the spectrogram. When the parameter p_k is 0, we will use three additional parameters to calculate the theoretical spectrogram.

- S_k the spread of the peak.
- C_k the center of the peak.
- D_k the amplitude of the peak (in db).

The distribution of these three parameters will be as follows:

$$S_k \sim \mathcal{N}(0, 0.1) \tag{55}$$
$$C_k \sim \mathcal{N}(0, 0.3) \tag{56}$$
$$D_k \sim \mathcal{N}(-25, 20). \tag{57}$$

These parameters will be constant in time, therefore we do not need to provide the evolution equations for them. We calculate the theoretical signal in case of the null hypothesis as

$$T(v) = 10^{D_k/10} \exp\left(-\frac{(v - C_k)^2}{S_k^2}\right). \tag{58}$$

Now that we are able to calculate the theoretical measurement vector T_{X_k} for a given system state X_k, we can calculate the measurement likelihood. The pdf of the measurements given the system state, $f_{Y_k|X_k}$, is given by the noncentral chi squared distribution as described in section 5.3. The noise in the frequency

bins is assumed to be independent so that we can fill in $T^i_{X_k}$ - the theoretical amplitude in frequency bin i given the parameter set X_k - into equation (16) and then multiply over all I frequency bins to get the pdf of the full measurement vector as

$$f_{Y_k|X_k}(y_k;T_{X_k}) = \prod_{i=0}^{I}\sum_{j=0}^{\infty} \frac{exp\left(-\frac{T^i_{X_k}}{2\sigma^2}\right)\left(\frac{T^i_{X_k}}{2\sigma^2}\right)^j}{j!} \frac{exp\left(\frac{y^i_k}{2\sigma^2}\right) z^{(2+2j)/2-1}}{2^{(2+2j)/2}\Gamma((2+2j)/2)}, \quad (59)$$

The value of σ in this equation is determined empirically from the measurements of a part of a fragment of a signal where there are no people.

7 Gaussian approximation

We would ideally want to sample from the importance density. However, since that is not possible, we could improve the efficiency of the particle filter by trying to sample from an approximation of the posterior density. For convenience we repeat here the posterior density, calculated from the measurement y_k at time k and the previous state x_{k-1}

$$f_{X_k|Y_k,X_{k-1}}(x_k;y_k,x_{k-1}) = \frac{f_{X_k|X_{k-1}}(x_k;x_{k-1})f_{Y_k|X_k}(y_k;x_k)}{\int f_{X_k|X_{k-1}}(x;x_{k-1})f_{Y_k|X_k}(y_k;x)\,dx}. \quad (60)$$

We will approximate the log density using a Taylor expansion. This is very useful because it will turn out that this generates an approximate density that can be sampled from using a normal distribution [5]. For shorthand notation of the logarithm of the posterior we will define the notation

$$\ell(x_k) = \ln\left(f_{X_k|Y_k,X_{k-1}}(x_k;y_k,x_{k-1})\right). \quad (61)$$

Let ℓ' and ℓ'' be the first and second derivative of ℓ with respect to x_k. Note that since x_k is a vector, the first and second derivatives will be a vector and a matrix, respectively. The calculation of these derivatives will be performed by finite differences. We can use these derivatives to get the second order truncated Taylor expansion of ℓ around some parameter vector $\tilde{x}_k$ which gives

$$\ell(x_k) \approx \ell(\tilde{x}_k) + (x_k-\tilde{x}_k)^T\ell'(\tilde{x}_k) + \frac{1}{2}(x_k-\tilde{x}_k)^T\ell''(\tilde{x}_k)(x_k-\tilde{x}_k). \quad (62)$$

This approximation holds for x_k "close to" $\tilde{x}_k$. In section 7.2 we will make more precise when exactly this approximation is valid. A very useful observation about equation (62) is that we can transform it to the log density of a multivariate normal distribution if we set

$$\Sigma = -\left[\ell''(\tilde{x}_k)\right]^{-1} \quad (63)$$

$$\mu = \tilde{x}_k + \Sigma\ell'(\tilde{x}_k). \quad (64)$$

This observation only holds when $\ell''(\tilde{x}_k)$ is negative definite so that it is invertible and Σ becomes positive definite. We will discuss this further again in

section 7.2. For now, we assume that this holds. If we then substitute this into the normal log density, we get

$$-\frac{1}{2}(x_k-\mu)^T\Sigma^{-1}(x_k-\mu) \tag{65}$$

$$=-\frac{1}{2}\left(x_k-\tilde{x}_k-\Sigma\ell'(\tilde{x}_k)\right)^T\Sigma^{-1}(x_k-\tilde{x}_k-\Sigma\ell'(\tilde{x}_k)) \tag{66}$$

$$=-\frac{1}{2}(x_k-\tilde{x}_k)^T\Sigma^{-1}(x_k-\tilde{x}_k)+(x_k-\tilde{x}_k)^T\ell'(\tilde{x}_k)-\frac{1}{2}\ell'(\tilde{x}_k)^T\Sigma\ell'(\tilde{x}_k) \tag{67}$$

$$=const+\frac{1}{2}(x_k-\tilde{x}_k)^T[\ell''(\tilde{x}_k)](x_k-\tilde{x}_k)+(x_k-\tilde{x}_k)^T\ell'(\tilde{x}_k), \tag{68}$$

which is equal to equation (62) up to a constant addition. This added constant in the log density becomes a multiplicative constant in the density. However, in order to use the approximation we have obtained as an importance density, we need to normalise it anyway, so the constant will disappear and equations (62) and (68) represent the same probability distribution. This distribution is, due to the form of equation (65), Gaussian. The importance density $g(x_k;\tilde{x}_k,y_k)$ we have now derived is therefore

$$g(x_k,;\tilde{x}_k,x_{k-1},y_k)=\det(2\pi\Sigma)^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(x_k-\mu)^T\Sigma^{-1}(x_k-\mu)\right) \tag{69}$$

A general method for sampling from a multivariate normal distribution is to obtain any root $\Sigma^{\frac{1}{2}}$ of Σ, sample X from a standard multivariate distribution (or equivalently a vector of many standard scalar normals). Then the standard normal can be transformed by the property [14]

$$\Sigma^{\frac{1}{2}}X+\mu\sim\mathcal{N}(\mu,\Sigma). \tag{70}$$

The only thing we have not given attention to yet is how to choose around which $\tilde{x}_i$ to build the Taylor series. Ideally $\tilde{x}_i$ is in a place where the posterior has a high likelihood since this means that it is an important location in the posterior. The best estimation about the particles we have been able to make so far though, is by using the evolution equation $f_{X_k|X_{k-1}}$. We will use this equation now to make an initial estimate for $\tilde{x}_k$ by evolving x_{k-1}. This means that $\tilde{x}_k$ will be stochastic and is distributed according to

$$\tilde{X}_k\sim f_{X_k|X_{k-1}}(\tilde{x}_k;x_{k-1}). \tag{71}$$

By making this approximation for all particles the posterior distribution is represented by many Gaussians, which means that the posterior is now represented by a mixture of Gaussians.

7.1 Updating the weights

The importance weights should be calculated differently now that the importance distribution is different. The distribution of a sequence of samples generated using the approximated distributions posteriors $g(x_j;\tilde{x}_j,x_{j-1},y_j)$ for

$j = 1, ..., k$ is given by

$$p_{X_{0:k}|Y_{0:k}}(x_{0:k}, y_{1:k}) = f_{X_0}(x_0) \cdot f_{X_1|X_0}(\tilde{x}_1; x_0) \cdot g(x_1; x_0, \tilde{x}_1, y_1) \quad (72)$$
$$\times f_{X_2|X_1}(\tilde{x}_2; x_1) \cdot g(x_2; x_1, \tilde{x}_2, y_2) \quad (73)$$
$$\times \cdots \quad (74)$$
$$\times f_{X_k|X_{k-1}}(\tilde{x}_k; x_{k-1}) \cdot g(x_k; x_{k-1}, \tilde{x}_k, y_k). \quad (75)$$

Combined with equation (28), we can use this importance distribution in order to obtain an updating equation for the weights, analogous to equation (32):

$$\omega_k^i \propto \frac{f_{X_{1:k-1}|Y_{1:k-1}}(x_{k-1}^i; y_{k-1}) \cdot f_{X_k|X_{k-1}}(x_k^i; x_{k-1}^i)\, f_{Y_k|X_k}(y_k; x_k^i)}{p_{X_{1:k-1}|Y_{k-1}}(x_{k-1}^i; y_{k-1}) \cdot f_{X_k|X_{k-1}}(\tilde{x}_k^i; x_{k-1}^i) \cdot g(x_k; x_{k-1}, \tilde{x}_k, y_k)} \quad (76)$$
$$= \omega_{k-1}^i \cdot \frac{f_{X_k|X_{k-1}}(x_k^i; x_{k-1}^i)\, f_{Y_k|X_k}(y_k; x_k^i)}{f_{X_k|X_{k-1}}(\tilde{x}_k^i; x_{k-1}^i)\, g(x_k; x_{k-1}, \tilde{x}_k, y_k)}. \quad (77)$$

Note the subtle difference here between $f_{X_k|X_{k-1}}(x_k^i; x_{k-1}^i)$ in the numerator and $f_{X_k|X_{k-1}}(\tilde{x}_k^i; x_{k-1}^i)$ in the denominator. The distribution with $\tilde{x}_k^i$ is due to the stochastic choice of parameters around which to evaluate the Taylor expansion and the distribution with x_k^i is simply the likelihood of the parameter evolution.

7.2 Validity of the approximation

In making a Gaussian approximation of the posterior distribution, we have made two assumptions. First of all, we assumed that the Taylor series was a valid approximation and secondly, we assumed that the Hessian matrix of the posterior density function was always negative definite. In addition we are approximating the derivatives that are required for the Taylor expansion by finite differences. We will see in this section that neither of these assumptions is always true, but we will provide arguments to justify the approximation and finally we will estimate the finite difference error.

Before going into the details of the assumptions, please note that mathematically, any importance density that has the same or larger support as the posterior density is a valid importance density in the sense that it will give the correct asymptotic results [1]. Since the approximated posterior that we are proposing as importance density is a normal distribution, it has infinite support and is therefore always a valid importance density. The potential problem resulting from an approximation error of any kind is therefore restricted to a decrease in the sampling efficiency. Though this is theoretically not a problem, a bad efficiency may mean that it is computationally better to use a simpler importance distribution and simply increase the number of particles. Experimentation will show whether the proposed Gaussian approximation provides more efficiency to the estimator than it costs in terms of computation time.

7.2.1 Taylor approximation

The first approximation we will analyse is the approximation of the log-posterior by a Taylor series. We fill first provide some convenient notation for the higher order differentials we will need. Let α be a vector in $\mathbb{N}^n$ describing how many

times to take derivatives in each dimension and define

$$|\alpha| := \sum_{i=1}^{n} \alpha. \tag{78}$$

Then we define

$$D^{\alpha} := \frac{\partial^{|\alpha|}}{\partial^{\alpha_1} x_1 \cdot ... \cdot \partial^{\alpha_n} x_n}, \tag{79}$$

where $\partial^0 x_i$ should be interpreted as not taking any derivative in dimension i. Using this notation we can conveniently write down the following theorem which gives an estimation for the error of the Taylor series [4]:

Theorem 1. *Assume that ℓ is at least three times differentiable within some open ball B around $\tilde{x}$ with radius r, $B := \{x \in \mathbb{R}^n : |x - \tilde{x}| < r\}$. Let $\tilde{\ell}$ be the second order Taylor approximation of ℓ around $\tilde{x}$. Then the error $\epsilon(x) := |\tilde{\ell}(x) - \ell(x)|$ can be bounded by*

$$\epsilon(x) \leq \frac{1}{6} \max_{|\alpha|=3} \max_{y \in B} |D^{\alpha} \ell(y)| \tag{80}$$

Using theorem 1 we can get some indication of the error by analysing the third derivative of ℓ around $\tilde{x}$. Unfortunately, theorem 1 cannot always be used in our case. The reason for this is that the assumption that ℓ is three times differentiable in B does not always hold. In fact, the functions in the Thalmann model are often piecewise differentiable, as can be seen in for instance equation 94 and therefore the Taylor approximation may not hold any more when B covers multiple pieces. Since the expressions for the Thalmann model become very elaborate and due to all the interdependencies between the parameters, it is very difficult to find exactly where the resulting posterior will not be differentiable. Instead, we choose to analyse the approximation error in a more empirical way by plotting the posterior density around parameters that occur when running the particle filter on a real signal. This way at least some feeling can be obtained about the smoothness of the functions. Figures 8 and 9 show how the posterior varies in each dimension around a particle. In figure 8, the posterior is quite smooth and we can imagine that a Gaussian approximation could do quite well. In figure 9, the posterior is much less smooth, especially in the phase parameter. For the approximation this means that it's quality can vary quite a bit and that especially the estimation of the direction and phase may be difficult because the posterior can vary a lot in these directions.

7.2.2 Convexity of the posterior density

The second problem is the positive definiteness of the Hessian. When the Hessian matrix is negative definite in some point, this means that the function is concave in that point. The logarithm of the normal distribution is also concave since it is basically a scaled and shifted version of $-x^2$. Therefore it can be understood that the normal distribution could provide a reasonable approximation, at least locally. However, the posterior we obtain when using the particle filter is not always concave. Practically, this poses the problem that $\ell''(\tilde{x})$ can be positive definite, making $\Sigma = -[\ell''(\tilde{x})]^{-1}$ negative definite. When we then

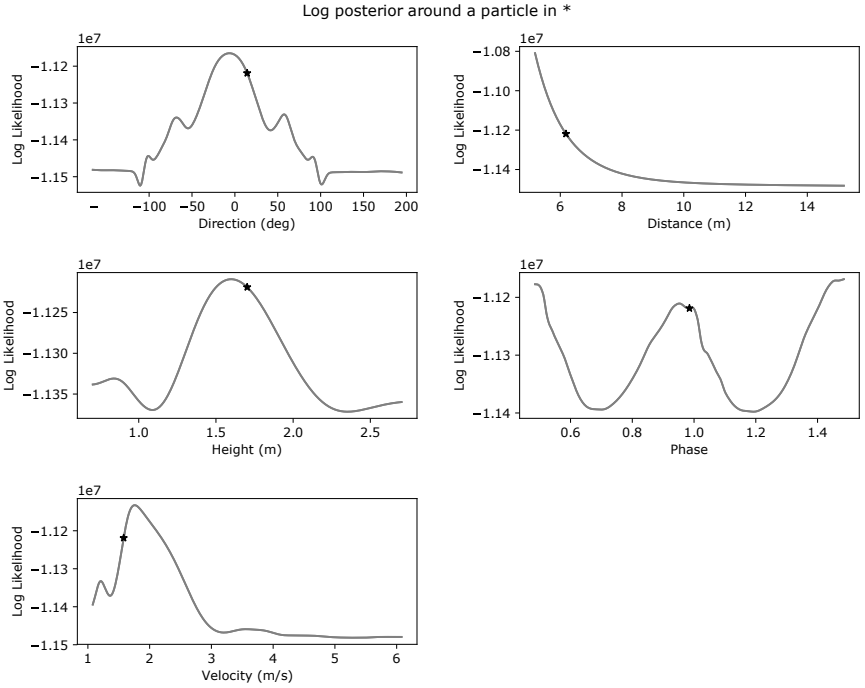


Figure 8: This part of the posterior density is relatively smooth with clearly concave regions around the maxima.

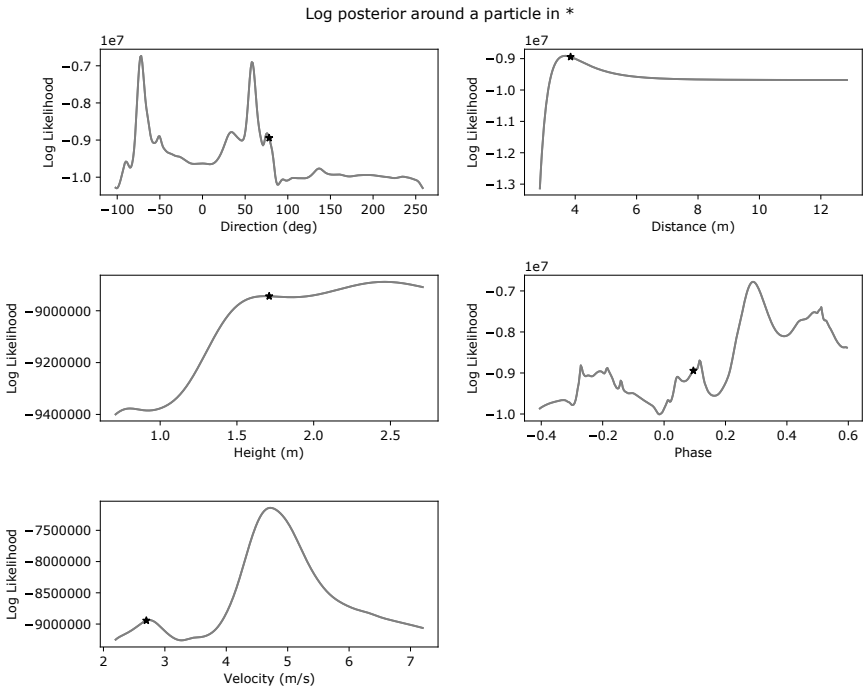


Figure 9: This part of the posterior density is much less smooth. We can see that the particle has ended up around a maximum that is actually only a local maximum in most dimensions.

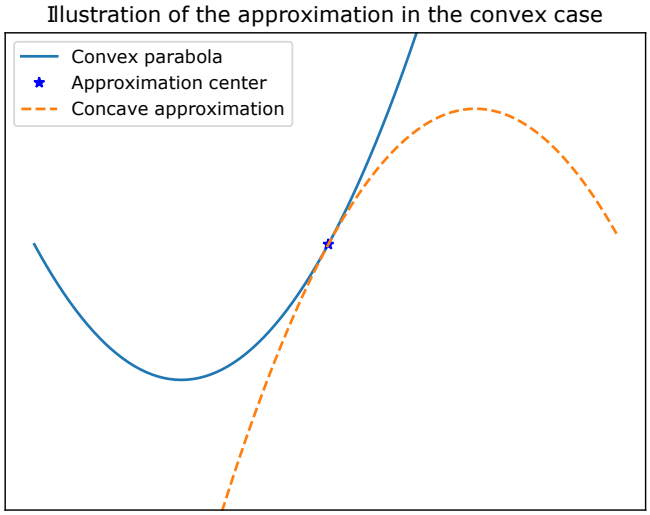


Figure 10: Approximation of a convex parabola by a concave parabola. The first order of the Taylor expansion still holds in this approximation.

want to sample from a normal distribution using equation 70 with this covariance matrix, we cannot find any real root $\Sigma^{\frac{1}{2}}$. Theoretically the matrix could never be a covariance matrix because one of the basic properties of covariance matrices is that they are always symmetric and positive semi-definite. If we still want to use a normal importance density, this means that we need to somehow change Σ so that it becomes positive semi-definite and remains symmetric, preferably in a way that still allows it to be a reasonable approximation of the posterior density, but does not require more computations. In order to make up a solution for this problem, we will assume that the posterior is locally parabolic so that a Taylor approximation could hold. The reason for this assumption is simply that when this does not hold, there is no way to know what the function looks like at all and we cannot find any solution at all. We will analyse scalar case first (meaning only one input parameter) and generalise the solution to multiple dimensions afterwards. When we calculate the Taylor approximation of a convex parabola, μ as calculated in equation 64 will be the minimum of this parabola. Using this as the mean of our normal approximation would mean that we will sample from a place that has lower posterior density than the location in which we calculated the derivative. Therefore, we will mirror the location of μ in the point in which we calculated the derivatives. Then we keep the same value for Taylor coefficient, but invert the sign so that the resulting parabola becomes concave and can be used for creating a normal density. The advantage of this method is that the resulting parabolic approximation still has the same first derivative in Taylor centre so that at least close to the Taylor centre, the approximation is not much worse than a convex approximation could be. Figure 10 illustrates how this approximation works.

In a multidimensional space, the function may only behave convexly in some directions. By doing a spectral decomposition of the Hessian matrix of the posterior density so that

$$\ell'' = V\Lambda V', \tag{81}$$

with V having the eigenvectors in its columns and Λ being a diagonal matrix containing the eigenvectors, we can decompose the parameter space into orthogonal eigenvectors (since the Hessian is symmetric) with their corresponding eigenvalues. Whenever these eigenvalues are positive, the function is convex in the one dimensional subspace, spanned by the eigenvector. When it is negative, the function is concave in this subspace. We may alter the Taylor approximation in equation 62 by replacing ℓ'' by $V(-|\Lambda|)V'$ to make the approximating polynomial concave without changing the first order Taylor approximation. In addition, due to the orthogonality of the eigenvectors, the second order approximation is still the same in the eigenvector directions in which the function was already concave. In addition, it can be seen from equation 64 that simply replacing ℓ'' this way and doing the same calculations results in the new μ being mirrored in $\tilde{x}_k$ in the directions of the eigenvectors whose eigenvalues were positive, just as we proposed.

7.2.3 Finite difference error

Finally, an error is introduced by the use of finite differences for the derivatives. This error depends on the step size we use for the finite differences. When the step size is too large, it will make an error due to the function that changes in between the steps. When the step size is too small, the error will rise again due to the limits of the floating point representation in the computer. Somewhere in between these two extremes, there will be an optimum for the step size where the error is minimal. Since we do not have an expression for the exact derivative, we cannot calculate the exact error of an approximation with a certain step size. Instead, we will calculate the relative difference between the derivatives approximated with two different step sizes. By doing this with step sizes $10^{-2}, 10^{-3}..., 10^{-9}$ we can find between which two step sizes the derivative changes fewest. Somewhere around here will be the optimal step size.

As we have seen in figures 8 and 9, the function we are trying to calculate the derivative of may be quite different in different places. In order to get a realistic idea of the error in practice, we will run the particle filter and approximate the derivatives for 10 time steps of signal for 50 particles. The figure 11 shows the mean of the resulting relative errors for the first and second derivative. It can be seen that the optimum for the first derivative lies between 10^{-6} and 10^{-7}. For the second derivative, this optimum lies between 10^{-4} and 10^{-5}. We choose to take 10^{-5} as our step size since the first derivative calculation benefits from a slightly smaller step size and it is close the optimal value for the second derivative.

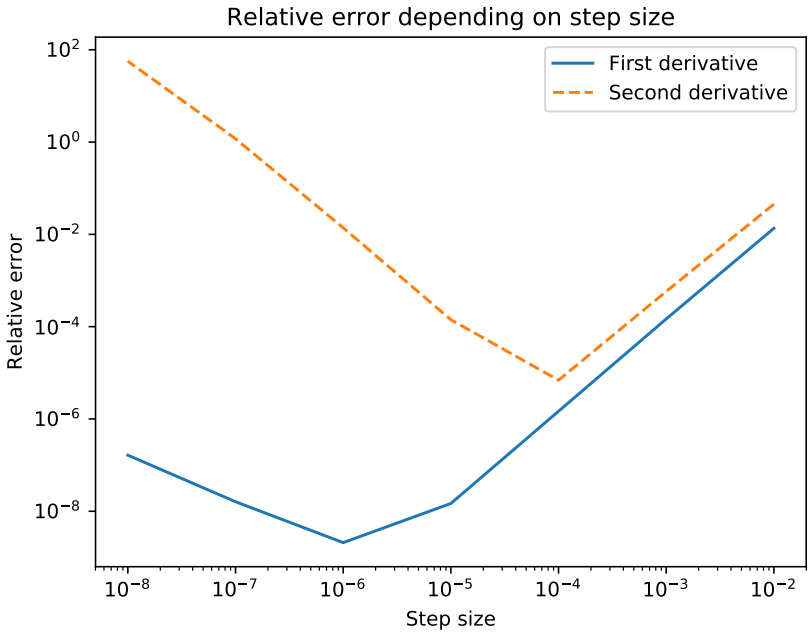


Figure 11: The relative difference between the derivative calculated using the step size given in the x-axis of the graph and a step size that is 10 times smaller.

7.3 The particle filter algorithm with Gaussian approximation

Now that we have all the calculations, we can write down the full particle filter algorithm with Gaussian approximation:

- Sample $\{x_0^1, ..., x_0^N\}$ from $\{X_0^1, ..., X_0^N\} \sim f_{X_0}$ i.i.d.
- Now for time $k = 1, 2, ...$:
  - **Resampling** (for $k > 1$)
    Draw $M = (M^1, ..., M^N)$ from a multinomial distribution with N trials and probabilites ω_{k-1}^i. M now represents the number of times we re-use each particle. Now relabel the superscripts of $\{x_{k-1}^1, ..., x_{k-1}^N\}$ so that each x_{k-1}^i occurs M^i times. Note that some particles may be discarded altogether by this process but we will always end up with a total of N particles.
  - **Gaussian approximation**
    Sample the first approximation $\{\tilde{x}_k^1, ..., \tilde{x}_k^N\}$ by drawing from $\tilde{X}_k^i \sim f_{X_k|X_{k-1}}(\cdot\,; x_{k-1}^i)$ for $i = 1, ..., N$.
  - Calculate $\ell'(\tilde{x}_k^i)$ and $\ell''(\tilde{x}_k^i)$ as defined by equation (61) using finite differences.
  - Calculate the eigenvalue decompositions of $\ell''(\tilde{x}_k^i)$ so that $\ell''(\tilde{x}_k^i) = V^i \Lambda^i [V^i]'$.

– Set

$$\Sigma^i = -(V^i(-|\Lambda^i|)[V^i]')^{-1} = V^i|\Lambda^i|^{-1}[V^i]' \tag{82}$$

$$\mu^i = \tilde{x}_k^i + \Sigma^i \ell'(\tilde{x}_k^i) \tag{83}$$

and let $X_k^i \sim \mathcal{N}(\mu^i, \Sigma^i)$ for $i = 1, ..., N$.

– **Importance sampling**
Now sample $\{x_k^1, ..., x_k^N\}$ from $\{X_k^1, ..., X_k^N\}$.

– Calculate the unnormalised weights of the new particles using the measurement y_k at time k:

$$\tilde{\omega}_k^i := \frac{f_{X_k|X_{k-1}}(x_k^i; x_{k-1}^i)\, f_{Y_k|X_k}(y_k; x_k^i)}{f_{X_k|X_{k-1}}(\tilde{x}_k^i; x_{k-1}^i)\, g(x_k; x_{k-1}, \tilde{x}_k, y_k)} \quad \text{for } i = 1, ..., N. \tag{84}$$

Where g is defined as in equation (69).

– Finally, normalise the weights to get

$$\omega_k^i = \frac{\tilde{\omega}_k^i}{\sum_{j=1}^N \tilde{\omega}_k^j}. \tag{85}$$

Then using the particles and their weights it is possible to estimate expectations at any time step.

8 Results

8.1 Signal and noise

The quality of the measurements than can be obtained by the device described in section 4 depends on the so called signal to noise ratio (SNR), which is the ratio between to the power in the signal and in the noise. Clearly the SNR depends on the signal strength and therefore it will vary depending on the distance to the object we measure. We will therefore plot how the SNR changes over time when a person walks away from the sensor to get some feeling for it. In all signals we will have a very large peak at 0 velocity caused by all static objects in the scene. We do not want to count this peak as signal since in contains no Doppler information. We therefore calculate the SNR by dividing the total energy in the frequency bins corresponding to velocities between 1.5 and 20 km/h by the energy in these bins when there is no Doppler signal. To be precise, let y_k^m be the measurement in frequency bin m and time k when the scene is empty and z_k^m the signal when there is a person walking. Suppose for notational simplicity that frequency bins 1 until M correspond to the velocities between 1.5 and 20 km/h. Then we first calculate an estimate for the mean energy in the noise

$$E_{noise} = \frac{1}{K} \sum_{k=1}^{K} \sum_{m=1}^{M} (y_k^m)^2. \tag{86}$$

Then we can calculate the signal energy relative to this noise floor in dB as

$$SNR(k) = 10 \log \left(\frac{\sum_{m=1}^{M} (z_k^m)^2}{E_{noise}} \right). \tag{87}$$

It should be noted that z actually contains signal as well as noise so that the resulting SNR is actually the signal to noise ratio plus one. For large ratios this difference negligible. For smaller ratios, it means that the minimal value the SNR can take is 0 dB.

The noise level in our system is determined by the receiver and the amplification circuit. These component are currently fixed, but it is still possible to increase the power of the sender in order to increase the signal to noise ratio. While measuring with increased power, unexpected artefacts started to occur in the signal. Figure 12 shows a spectrogram of a single walking person that was measured with a higher power. The signal of the person can be seen in the positive velocities, however, this signal also appears mirrored in the zero velocity line in a slightly weaker form. We have not studied the exact reason for this, but our hypothesis is that it is caused by the sound following a secondary path, reflecting onto the surroundings behind the walking person and then arriving from the opposite direction at the person. This explains why the frequency change is exactly opposite and also why the secondary signal has a lower amplitude (because it travels a longer path). Practically, it means that the maximum usable signal power is bounded by this effect. Figure 13 shows a measurement that was recorded with slightly lower sending power. In this spectrogram the effect has diminished enough to be visible only when the person is very close (less than two meters) to the sensor. We will use this power level for all subsequent measurements or otherwise state this explicitly.

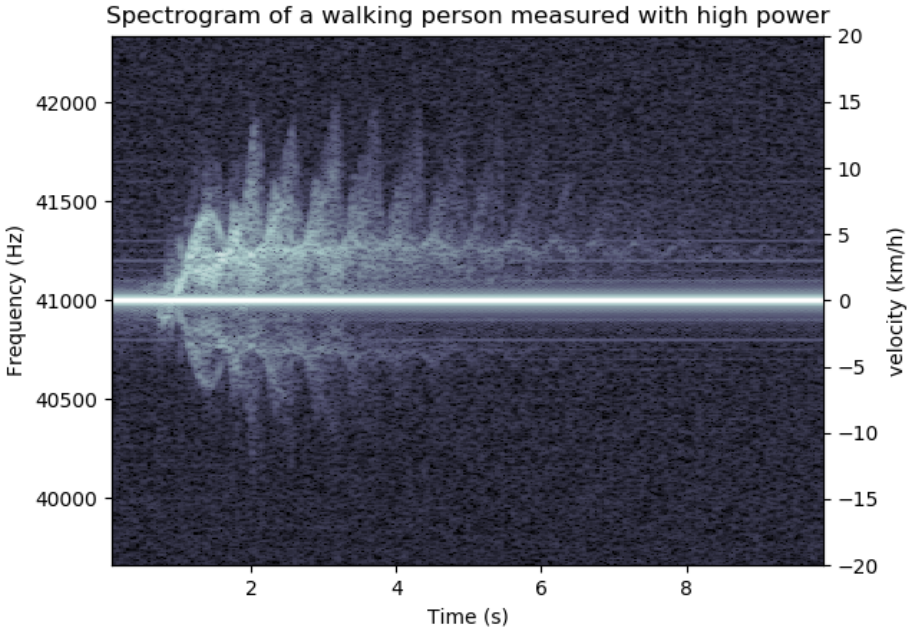


Figure 12: Example of a spectrogram measured with high sending power, showing a copy of the original signal mirrored in the zero velocity line.

The graph containing the SNR of the two signals can be found in figure

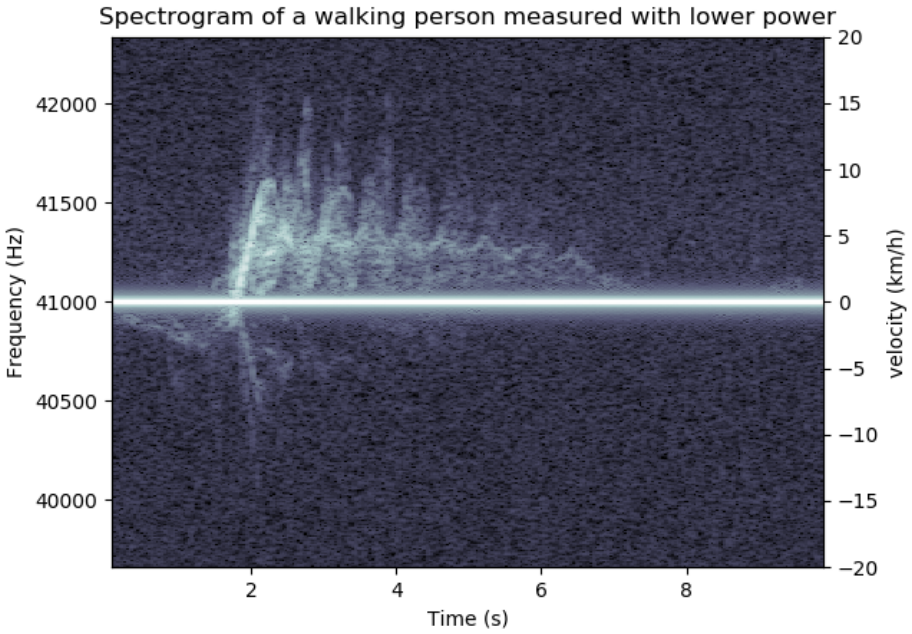


Figure 13: Example of a spectrogram measured with a lower sending power than the one in figure 13, the secondary signal occurs very slightly at the very beginning but is quickly too low to see.

14. In this graph the theoretical attenuation of the signal power (described in equation 10 has also been plotted based on a person walking at 4 km/h. It is a bit difficult to compare the two lines because there will be small variations in the way the person walked, causing the SNR to decrease faster when the person walks faster. At least we can see that the theoretical line has the same characteristics as the measured lines and that the high power SNR remains higher for a longer time, confirming the expectation that the high power measurement provides a larger range.

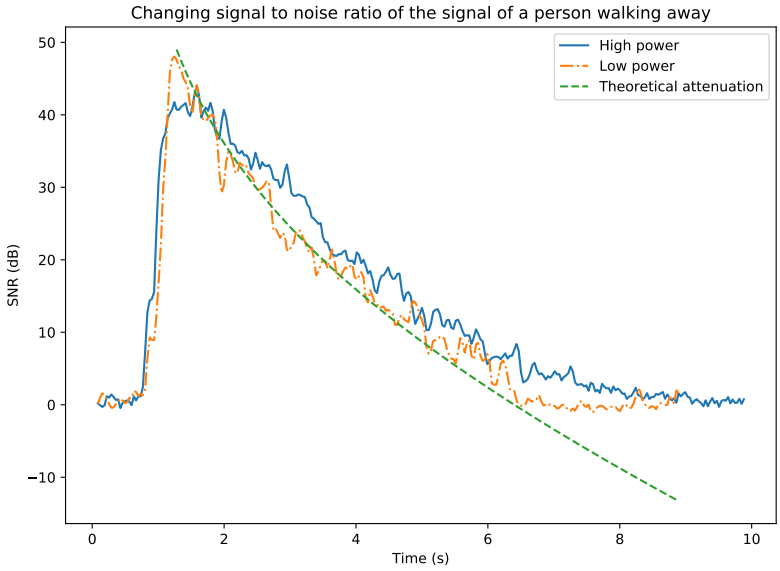


Figure 14: Signal to noise ratio changing over time when a person walks away calculated from the signal in figure 12 for high power and figure 13 for low power. In addition the theoretical attenuation of a person walking at 4 km/h has been plotted.

8.2 Computational complexity

In the end, the model will need to be run in real time. The computation of a single measurement should therefore not take any longer than the time between the measurements. At each time step in the spectrogram is calculated from a segment of approximately 0.17 seconds. Since, in addition, there is a 50% overlap between the time windows, the filter needs to be able to process a measurement within 0.085 seconds. We will see that this speed is currently not reached for all the algorithms. That does not mean that it is not possible at all though. In the recommendations at the end of this thesis we will provide some suggestions to make the algorithm more efficient. One of the things we should state already is that in the current implementation of the algorithm, all calculations are performed for all three models (0, 1 and 2 persons) for each particle, regardless the state of the number of persons parameter. The theoretical advantage of this is that the computation time is constant as opposed to depending upon the measurements. Practically it means that the times we will give can be seen as an upper boundary on the computation time.

We will compare the time it takes the SIR filter and the Gaussian approximation filter to run 12 consecutive measurements. 12 measurements cover a time exactly equal to 1.024 seconds. We divide by this time to get the ratio between the calculation and the covered signal time. This test was run on an notebook with a dual core Intel Core I7-3540M processor running at 3.00 Ghz. While running the filters, the CPU utilisation was 100%. Figure 15 shows the

results of this test. It can be seen that the Gaussian approximation filter is not able to run in real time. The SIR filter for one person can run in real time for less than approximately 140 particles and the SIR filter for two persons runs stays below the real time boundary when the number of particles is less than approximately 50. Of course these numbers depend very much on the hardware used.

We can see that the relative difference between the one person and two person versions of the two filters are quite constant. For the SIR filter, the computations for the two person version take, on average, 1.7 times longer than the single person version. The Gaussian approximation takes, on average, 7.8 times as long to calculate for two persons. This can be explained as follows. In the ordinary SIR filter, the calculation of the Thalmann model take most of the time. Since it then needs to be calculated for twice as many persons, the calculations of the full filter will take approximately twice as long. From this argument it follows that the computational complexity of the SIR filter fill grow linearly with the number of people. In the Gaussian approximation filter, a lot of computation time is spent on calculating the eigenvalue decomposition of the Hessian matrix. This operation has a computational complexity of $\mathcal{O}(n^3)$ [16], which explains why doubling the number of persons increases the computation time of the filter by a factor of approximately 8. Unfortunately, this also means that the computational complexity of the filter will grow cubically with the number of persons that the filter can detect.

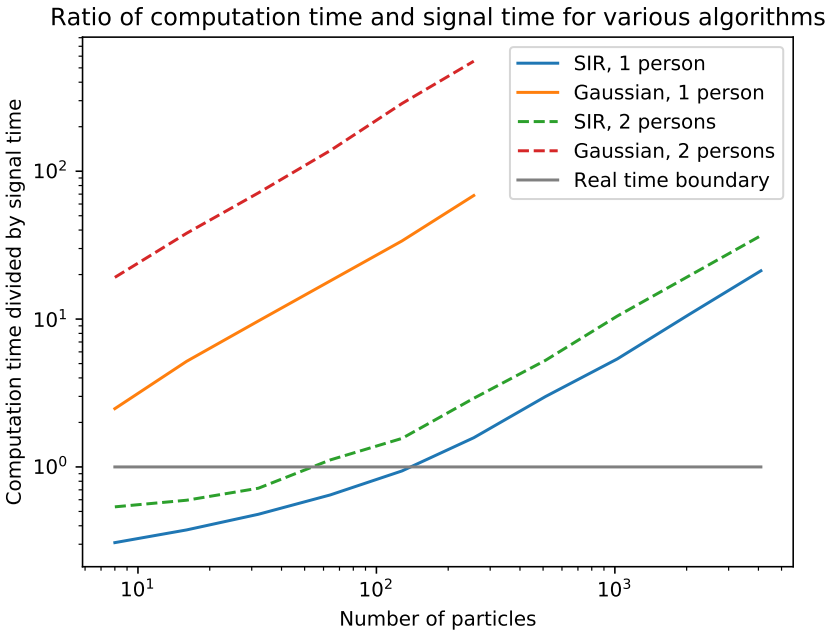


Figure 15: Computation time of the four variations of the particle filter on a log-log scale. It can be seen that the computation time increases linearly with the number of particles, apart from some overhead at low numbers of particles. Only the ordinary SIR filters manage to run in real time. Due to the very long computation time, the Gaussian filters have not been run with more than 256 particles.

8.3 Filtering behaviour

In order to analyse the behaviour of the filters in some real situations, we have made measurements in an office environment. We have tested how the filter performs for one and two persons walking in various directions. In addition, we have tested how the filter responds to other types of movement that are common in an indoor environment, namely opening doors and moving chairs. This way we can assess whether the filter correctly characterises when movement is not human. We will start with the situation that gives the most distinct Doppler signature, namely, a single person walking either directly towards or away from the sensor.

Due to the stochastic nature of the filter, it is difficult to draw conclusions from a single evaluation of the filter on a piece of signal. Figure 16, 17, 18 and 19 should be viewed merely as examples of what results the filters could provide. Still it is also important to look at these examples because we need to judge if they would provide useful information when used in a real situation. One thing that can be seen from these graphs, is that the estimation of the number of persons seems to take on integer values only. The cause of this is that the noise in the signal is relatively low, therefore the measurement likelihood function is very steep, causing the likelihood difference between the particles to be very

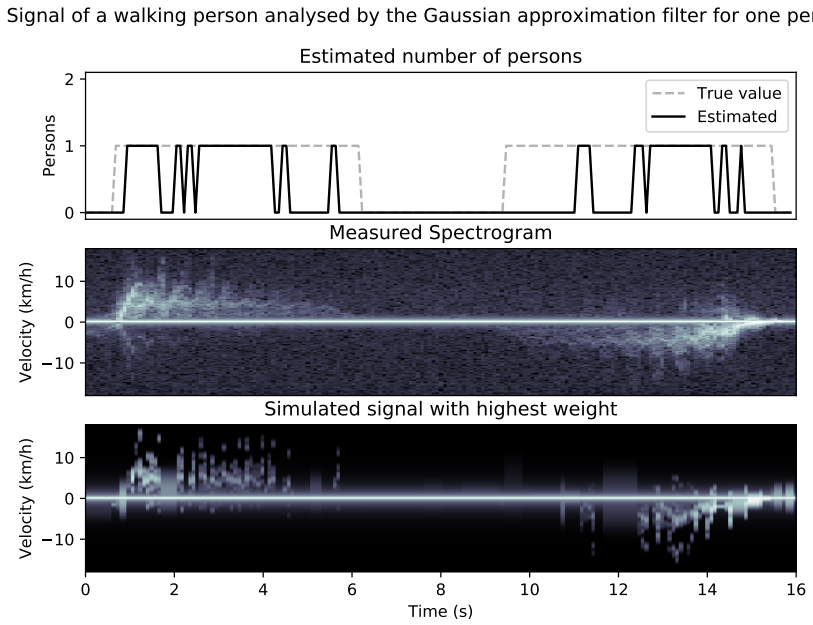


Figure 16: Example of the result of the Gaussian approximation filter for one person run with 50 particles.

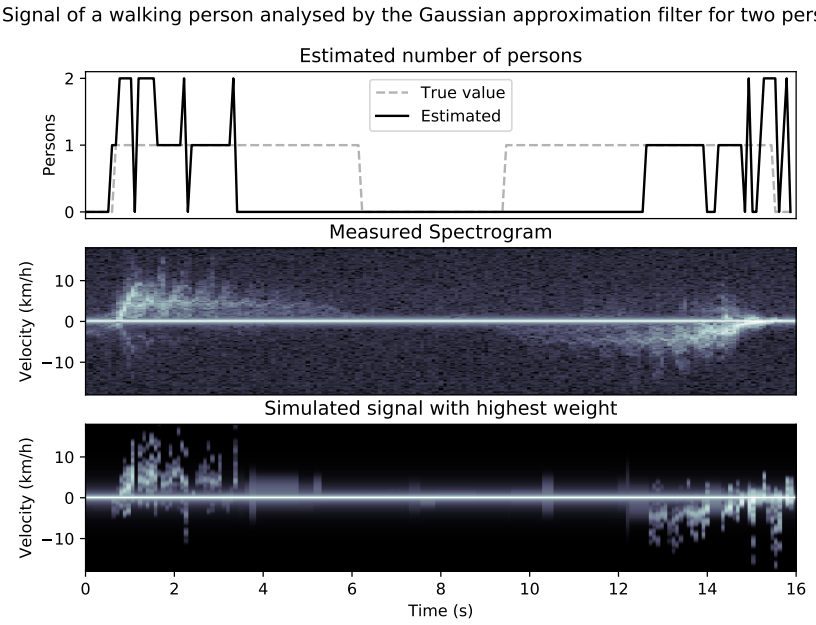


Figure 17: Example of the result of the Gaussian approximation filter for two person run with 50 particles.

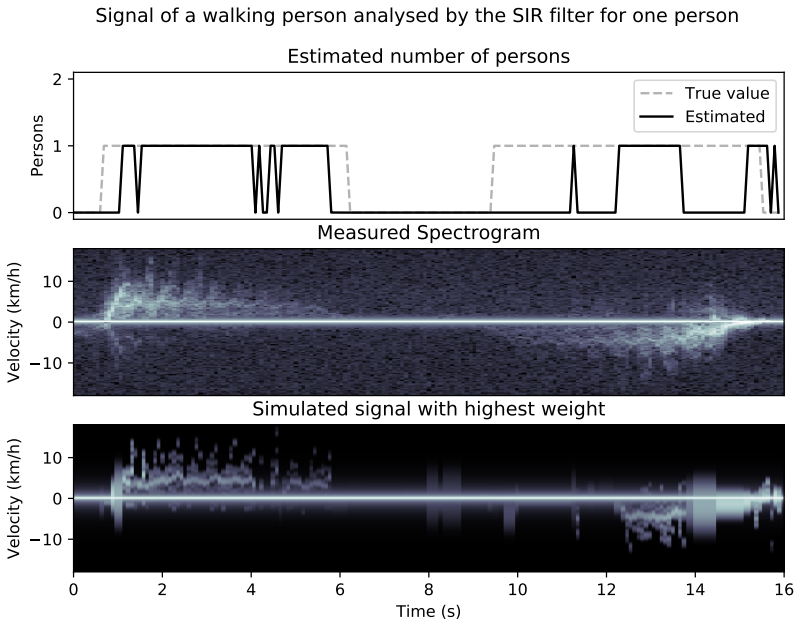


Figure 18: Example of the result of the SIR filter for one person run with 50 particles.

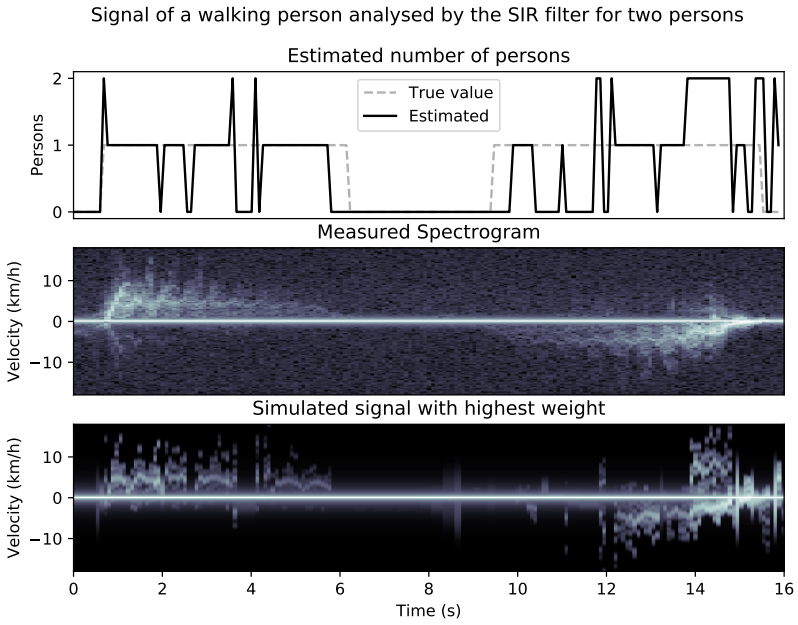


Figure 19: Example of the result of the SIR filter for two person run with 50 particles.

high even when their difference is quite small. In practice, this means that a single particle gets all the weight and the estimation is more like a maximum likelihood estimator than an expectation. Another thing that can be seen in these results, is that the estimations often quickly switch from 1 person to 0 and back again when there is still a signal. This is generally caused by the null hypothesis, accidentally providing a signal that fits the measurement better than the Thalmann model. Since the signals in the null hypothesis are constant peaks, the dynamic nature of walking means that the Thalmann model will quickly fit better again. In figure 18 we can see this happen several times. Starting from 14 seconds, this happens twice in a row, leaving 0 persons in a large part of the estimation.

For a more representative analysis of the algorithms, we have evaluated them 50 times on the same piece of signal and plotted how the mean and error of the estimations behave over time. Figures 20 until 35 show the results of these experiments on various measurements. We have used the shorthand labelling SIR 1 and SIR 2 for the one person versions and two person version of the ordinary SIR algorithm and Gaussian 1 and Gaussian 2 for the algorithms with the Gaussian approximation of the posterior for one and two persons.

We will start our analysis on the most clear measurement: a persons walking away from the sensor and back. The results of this can be found in figure 20. It can be seen that in general the filters for a single person have smaller error than the ones for two persons. This makes sense because the estimate can vary more if the filter can choose between an extra state. If we compare the SIR filter and the Gaussian approximation filter for one person, we can see that they perform approximately the same, with the SIR filter having a slightly lower error is most cases. Looking at how the decreasing signal strength affects the filters, we can see that the error of the one person filter is close to zero up to about 5 seconds, after which it increases quickly. At this point the signal is too low to convey the specific details of the Doppler signature and is therefore not seen as a walking person any more. When the person walks back, the filter is less quick in detecting the person. This can be explained by the mechanism of the particle filter. When the signal start out high, it is quite clear that there is a person and many particles representing a persons will survive the resampling step and have some time to "adapt" to the specific Doppler signature before its strength starts to decrease. However, when the signal start out low, it remains uncertain if the signal is caused by a person and it is more difficult to select the right parameters to follow the signal. Therefore persons walking away from the sensor can generally be detected earlier than persons walking towards the sensor. Another final feature caused by signal strength that can be observed from figure 20 is that the filters for two persons more often tend towards detecting two persons when the signal strength is higher. This is not unexpected as multiple persons can also increase the signal strength and it is not possible for the filter to distinguish a persons who is close to the sensor from two persons walking in the same pace a bit further away. A much more difficult case for the sensor is when a person walks perpendicular to the measuring direction. In principle, the sensor can only measure radial velocities so in theory a person moving in a perfect circle around the sensor would be invisible. In practice people walk in straight lines so that there will generally be some radial component relative to the sensor in their velocity. When a person walks perpendicular to the sensor, this radial component will suddenly change sign when the person crosses the normal line of

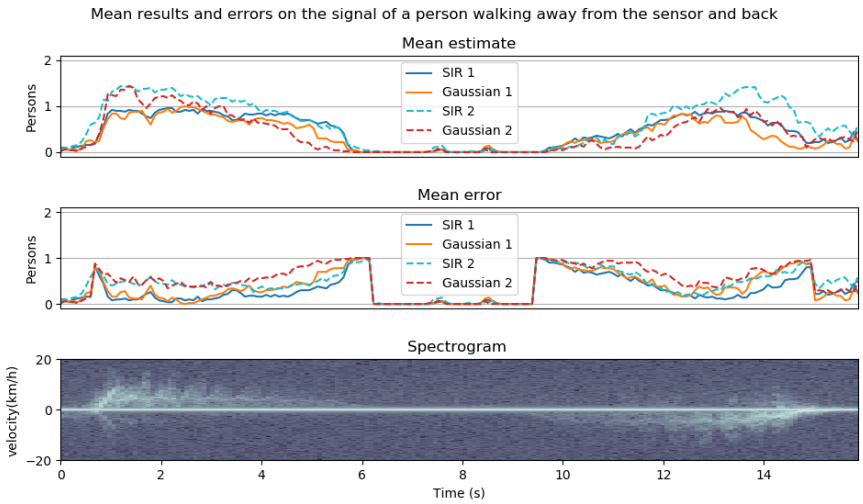


Figure 20: The mean estimate and error of the four different filters on the signal of a persons walking away from the sensor and back. The spectrogram of the analysed signal has also been plotted.

the sensor. Since it is not possible to measure at which angle from the normal a person is, this moment comes quite unexpected for the particle filter and we can see in the resulting analysis in figure 21 that at around 3.6 seconds, this happens and also that the error of the particle filter increases. In general, we can see that the error is much higher than the error on the previous measurement. We can see in the signal that there are much fewer details to observe making it much harder to distinguish human movement. In fact, the mean estimated number of persons never rises above 0.5, indicating that the filter is wrong more than 50% of the time. We must conclude that a person walking perpendicular to the sensor cannot be very reliably detected.

When we turn to the detection of multiple persons, the easiest case is when the persons are walking in opposite directions. This way their signals are clearly separated by positive and negative velocity relative to the sensor. However, the performance is not great. Of course the one signal increases in strength while the other decreases. This can also be seen in the error which is lowest in the centre of the signal where both signals are approximately equal. At around 5 seconds the person who is walking towards the sensor is already gone and there is just one person in the signal. Here we can see that the mean estimated value clearly drops, indicating that the filter is able to make the distinction between one and two persons, but just not very well. Since the parameter space has become considerably larger when a second person is added, it might be the case that the filter just needs more particles in order to function more reliably. For the Gaussian approximation, the calculation time is just too large for this to be feasible, but for the SIR filter we can try to increase the number of particles. To test this hypothesis we have also run the algorithm with 2500 particles. The results of this can be found in figure 23. We can see that the estimation increases a little bit, but the mean is still just about 1.5 where it should be close to 2.

When we try to run the filters for two people who are walking in opposite

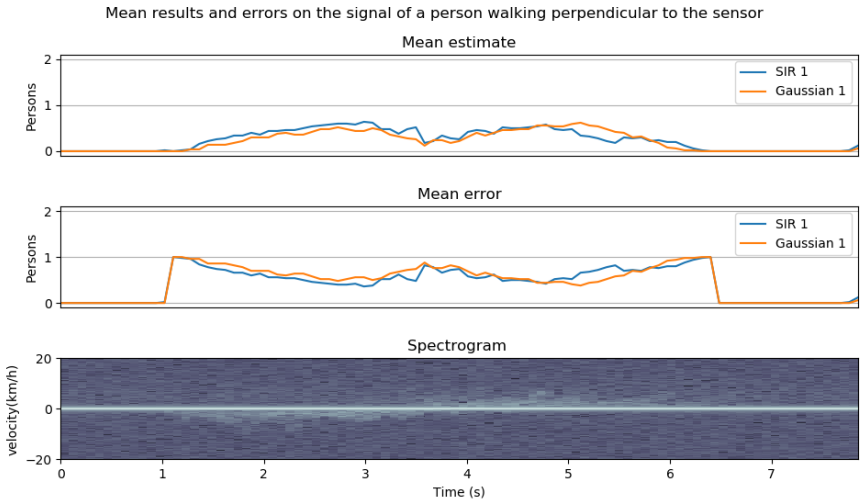


Figure 21: The mean estimate and error of the two one-person versions of the filters on the signal of a persons walking perpendicular to the sensor. The spectrogram of the analysed signal has also been plotted.

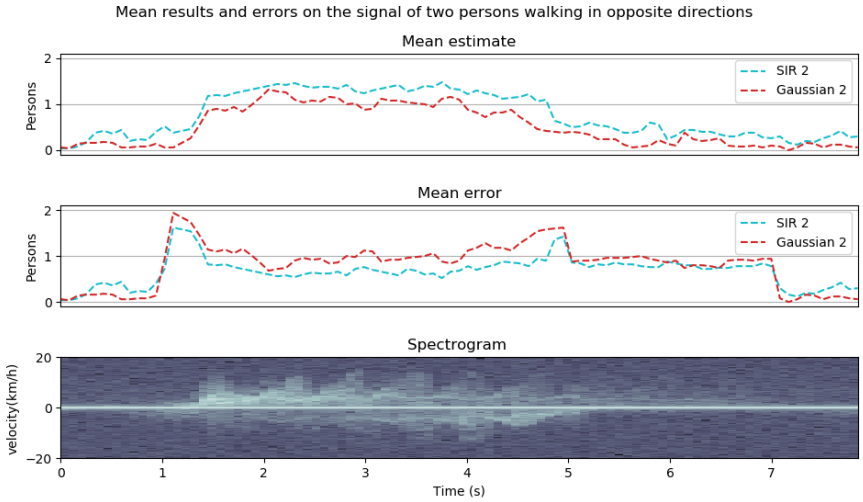


Figure 22: The mean estimate and error of the two two-person versions of the filters on the signal of two persons walking in opposite directions. Above 5 seconds, there is just one person left.

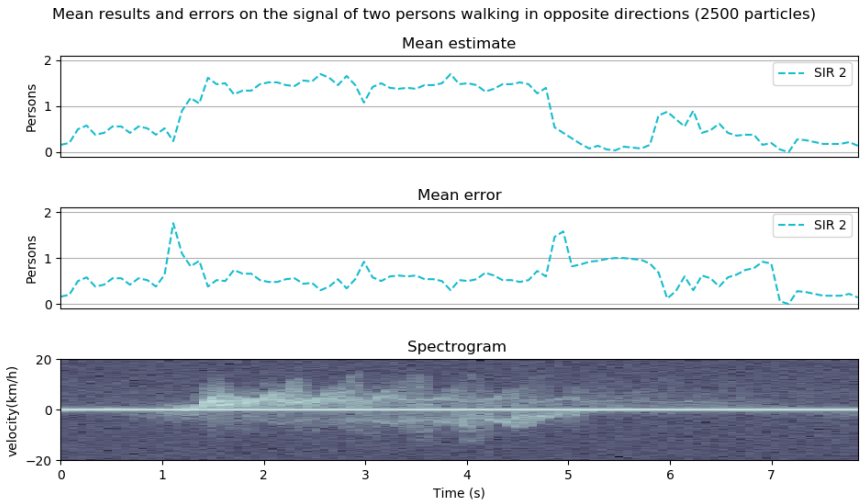


Figure 23: The same measurement as in figure 22, but now the SIR filter was run with 2500 particles.

directions, we obtain about similar performance of the two filters as we saw in the signal of one person in figure 20. This can be understood by looking at the signal which looks very much like the signal of one person. When we look at the video footage that was recorded with the Doppler signal in figure 29, we can see that the two persons have taken on the same pace, making them virtually indistinguishable for the system. Finally, we would like to see if the filters can make the distinction between human walking movement and other types of movement. We will test this on an opening door and a person moving around a chair. Figures 34 and 35 show the results of these tests. We can see that the filters generally respond less to the door and chair motion than to walking motion of similar signal strength such as seen in figure 20. However, we can also see that in about half of the times, the filter still classifies the motion as a walking person. In order to make the filter more reliable, we could try to make a more elaborate model for the null hypothesis that better captures dynamic signals than the constant peaks in the spectrogram we have used now. In addition, we could try to smooth the signal and set a threshold at for instance 80% certainty before detecting a person.

Figure 24: Mean estimate and error of two persons walking in the same direction. The estimates are similar to the estimates on the signal of a single person.

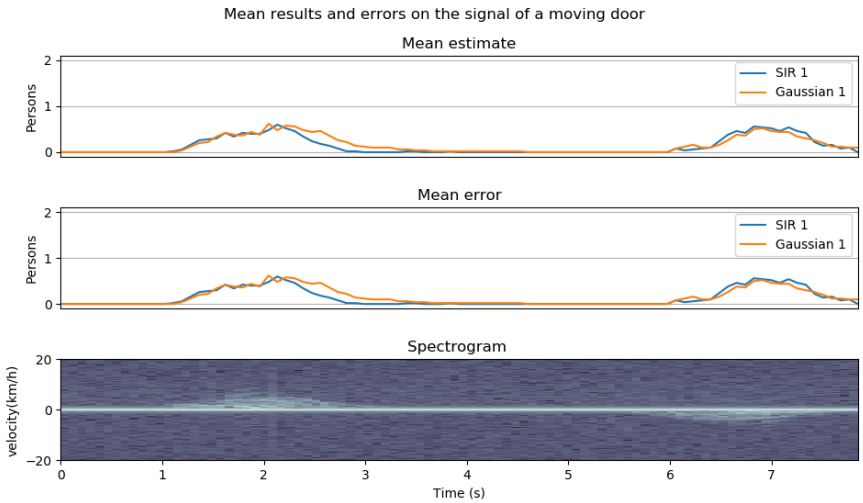


Figure 25: Mean estimate and error of estimates on the signal of a door opening, standing still for a while and then closing.

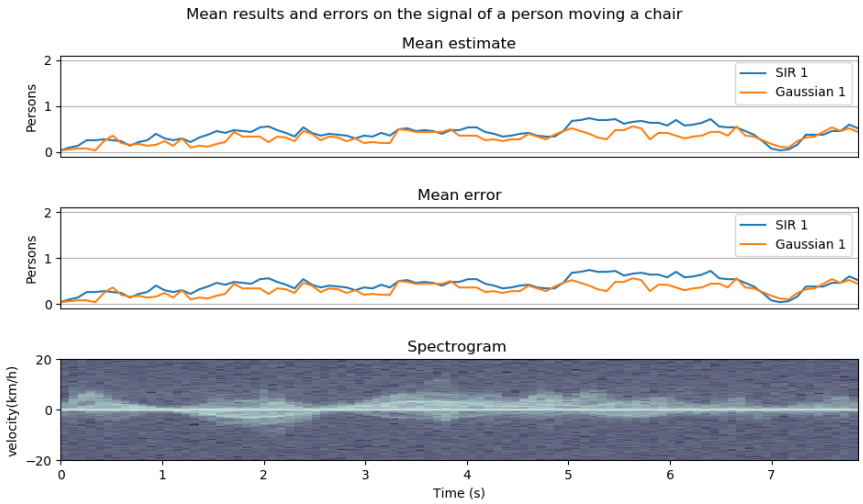


Figure 26: Mean estimate and error of estimates on the signal of a chair that is moved around in various directions.

8.4 Testing in smoke

A very important requirement for the sensor is that it continues working in smoke conditions. In order to test this, we were allowed to measure on a training location of the fire brigade at the Spinel Safety Center in Dordrecht, The Netherlands. At this training facility, a house containing some common furniture like a couch and some closets and a small kitchen was filled with glycerine based smoke, so dense that it was not possible to distinguish objects from even half a meter distance. This type of smoke is sometimes also called "cold smoke" because it does not require heating of the room in order to generate it. This way we were able to test our prototype that has not been designed to withstand high temperatures yet. The sensor was placed inside the house next to its entrance. Figure 27 shows the inside of the house before it was filled with smoke.

During the smoke measurement, we measured the signal strength in the room with and without smoke when the room was empty (no moving objects). The difference in signal strength between these two measurements was smaller than the natural changes in the signal strength in a static scene, showing that the sensor can measure equally well in smoke as in normal conditions. Despite this, when looking at the spectrogram that was measured on a person walking in the smoke filled room in figure 28, it is much harder to recognise the characteristic Thalmann Doppler signature. Part of the reason for this is that the reflections in the opposite velocity occur which were described in section 8.1. Unfortunately this is the only measurement we currently have and therefore we will not run the particle filter on the signal as it would not give a representative result. However, this is not the only thing that can be noticed in the signal. Clearly the characteristic waves that can be observed in walking motion (such as in for instance figure 13) are less present. Starting from about 18 seconds they can be found, but still at a much lower velocity level than in normal conditions. Since the sensor itself is not affected by the smoke, the only conclusion that can be

Figure 27: A picture of the inside of the testing house before it was filled with smoke. The picture is taken from the entrance. The test subject walked the room from the right side to the left and back.

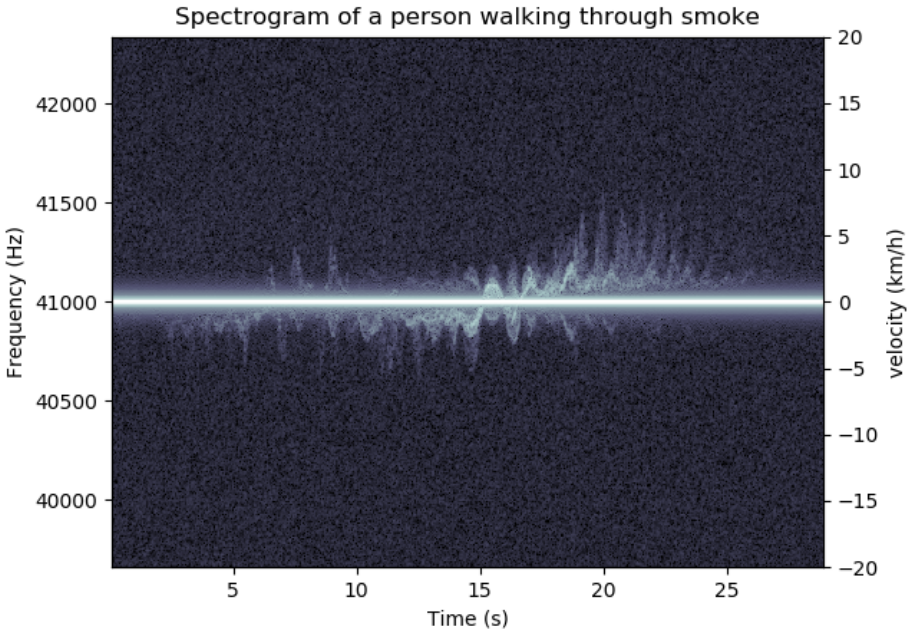


Figure 28: This spectrogram was made while a person walked towards the sensor and then back again through smoke.

drawn from this difference, is that the person himself is behaving differently. This conclusion was indeed confirmed by the person on whom the measurement was performed. When walking in the smoke filled room, the test subject was not able to see anything and therefore very withholding while walking out of fear of bumping into something. This can especially be seen in the signal before 15s where the person first walked towards the sensor. After this the person knew the room a little bit better and was able to walk back in his normal fashion, but still much more slowly than he would normally walk.

9 Conclusion

I this thesis, we have built and tested an inexpensive sonar sensor in combination with a particle filter in order to estimate when there are people walking through a room. Based on the tests performed in various different situations, we are now able to answer the research questions we set in section 3.2. In order to find out to what extend a particle filter can be used to estimate the presence of people walking through a room from the measured spectrograms, we have performed tests on a single person walking in different directions, on multiple persons and on moving doors and chairs. We performed this test using an ordinary SIR filter as well as a filter that we tried to improve by sampling from a Gaussian approximation of the posterior density. Since these two filters turned out to perform approximately the same, we will first draw our general conclusions for the filter performance in the different scenes and afterwards comment on the difference between the two filter implementations.

From the test on a single person we can conclude that all four particle filters we defined can detect a person walking in the sensor direction reliably. Provided that the person is close enough to the sensor so that the details of its Doppler signature can be distinguished from the signal, the accuracy is close to 100%. The filters that were made for detecting two persons have a significantly larger error because they have a tendency to over estimate the number of people when the signal strength is high. When a person is walking perpendicular to the sensor, the filter does not make reliable estimates, being wrong in more than 50% of the filter evaluations at each time step. We must therefore conclude that reliable detection is possible with the filters we defined, but only if the movement is radial to the sensor.

When testing with multiple persons, we found that the filter can make the distinction between no person at all or more than zero persons very well, just like in the 1 person case. However, in both the easier case when the two persons are walking in opposite directions and the harder case when the persons are walking in the same direction, the filters have a tendency to estimate one persons rather than two, resulting in the mean estimate not rising above 1.5 at any time, even when the number of particles is increased by a fifty fold. We therefore conclude that the detection of any person at all in the presence of multiple persons is very reliable, but actually counting their number is not accurately possible with the filters we defined.

We also tested the filters on measurements performed on moving doors and chairs. The mean number of persons detected during these kinds of movement never raised above 0.6. This is much lower than in the one and two persons case so that we can conclude that the filter is clearly more sensitive to walking movement than to the movement of chairs and doors and it would be possible to select for only human walking movement by classifying a piece of signal as walking movement only when the certainty rises above a certain threshold.

Combining these three different tests, we can assess to what extend a particle filter can be used to estimate the presence of people walking through a room from the measured spectrograms. When the people are walking in the sensor direction, the particle filter can reliably serve as a zero-one indicator for the presence of a walking person, even when other types of movement such as opening doors or people moving chairs are present. For counting the number of persons, the filters we have defined are not reliable enough.

To comment on the difference between the filter based on Gaussian approximation and the ordinary SIR filter, we have observed that the SIR filter is able to run in real time and its computing time scales linearly with the number of people it is programmed to detect. The Gaussian approximation filter for one person should still see a speed up of 10 times before it could run in real time with 50 particles and additionally scales cubically with the number of persons it can detect. In addition, throughout all our tests, the SIR filter consistently has equal or lower error. We must therefore conclude that the Gaussian approximation does not improve the particle filter, showing that after all, the posterior density we acquire for this problem cannot be well approximated by the Gaussian density.

Finally, to answer how smoke affects the measurements, we have performed measurements in a room filled with glycerine based smoke. This test showed that the sensor that we have built is not affected by this smoke at all. It should be noted though that the persons that we are trying to measure clearly are affected by the smoke so that their movement is different. In order to be able to really detect people in an emergency situation, this difference should be taken into account.

We can summarise our work and answer the main research question as follows: Using the spectrogram generated from the Doppler measurements of an inexpensive ultrasonic sensor, a SIR particle filter can provide a reliable indication of the presence of a walking person, also in smoke conditions, even when there are also other sources of movement, provided that the person exhibits ordinary walking motion and the sensor is directed in the walking direction.

10 Discussion

We acknowledge that the restrictions that we named in the conclusion as requirements for the system we built to be reliable are too strict to be usable in a real emergency situation. However, this thesis should be regarded as the very first test for the feasibility of using ultrasonic sensors for assistance of the fire brigade in emergencies. Within this framework, we have contributed only a very small part of what would eventually be the total solution. In our view, the total solution requires many sensors working together in order to track people while they are in a room so that when at some moment during an emergency they become unconscious, their locations can be remembered. In order to find the locations of people, distance measurement may be used with multiple sensors in order to be able to use trilateration to estimate locations. However, when locations are saved after people have stopped moving, it is very important to distinguish human movement from other sources of movement. Otherwise any door that has been moved at some point would be registered as an unconscious person by the system. This is exactly the point where our research can provide a valuable addition to the sensor system we just described.

Of course even for this specific feature, our research is just the start. We have only tested on at most two different persons in very controlled situations. Before productising the system we have proposed, extensive testing in realistic situations should be performed to be sure that the system will perform as expected in a real emergency situation. In addition to more testing, there are several other interesting directions for performing future research.

One thing that has the potential to greatly improve the reliability of the detection of people is to incorporate distance measurements. This would require sending a pulsed signal and processing it accordingly, possibly by extending the particle filter. When the distance to a moving object is known, it is possible to separate multiple objects in space and to estimate their size from the signal strength they return. When multiple sensors performing distance measurements are used together, it is even possible to perform trilateration in order to find the locations of the observed objects.

A more hardware related type of research would be to build a more robust ultrasonic sensor that is able to withstand higher temperatures and has additional components to be able to send its measurements to some central control unit that combines all the measurements. The prototype we have used worked in glycerine based smoke at low temperatures but would not be able to withstand temperatures above 80 °C and may be affected by smoke that contains corrosive chemical compounds.

Finally, if the above two research topics provide promising results, it is still necessary to do research into a good way to gather all the data produced by the sensors when there are many of them in a building and to process them. In addition, a very important question to answer would be how the information the sensors can obtain can be presented to the people from the fire brigade in a way that allows them to get a quick overview of the situation and make the right decisions.

References

[1] Søren Asmussen and Peter W. Glynn. *Stochastic Simulation: Algorithms and Analysis*. Springer New York, 2007. DOI: 10.1007/978-0-387-69033-9.

[2] Ronan Boulic, Nadia Magnenat Thalmann, and Daniel Thalmann. "A global human walking model with real-time kinematic personification". In: *The Visual Computer* 6.6 (Nov. 1990), pp. 344–358. DOI: 10.1007/bf01901021.

[3] Olivier Cappe, Simon J. Godsill, and Eric Moulines. "An Overview of Existing Methods and Recent Advances in Sequential Monte Carlo". In: *Proceedings of the IEEE* 95.5 (May 2007), pp. 899–924. DOI: 10.1109/jproc.2007.893250.

[4] Richard Courant and Fritz John. *Introduction to Calculus and Analysis Volume II 1 Chapters 1 4*. 2009. ISBN: 9783540665694. URL: https://www.amazon.com/Introduction-Calculus-Analysis-II-Chapters/dp/B0095HARQQ?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=B0095HARQQ.

[5] Arnaud Doucet, Simon Godsill, and Christophe Andrieu. "On sequential Monte Carlo sampling methods for Bayesian filtering". In: *Statistics and Computing* 10.3 (2000), pp. 197–208. DOI: 10.1023/a:1008935410038.

[6] Mark Goldstein. "Carbon Monoxide Poisoning". In: *Journal of Emergency Nursing* 34.6 (Dec. 2008), pp. 538–542. DOI: 10.1016/j.jen.2007.11.014.

[7] Stephan Groot et al. "Human motion classification using a particle filter approach: multiple model particle filtering applied to the micro-Doppler spectrum". In: *International Journal of Microwave and Wireless Technologies* 5.03 (Apr. 2013), pp. 391–399. DOI: 10.1017/s1759078713000366.

[8] Sevgi Z. Gürbüz, William L. Melvin, and Douglas B. Williams. "Detection and identification of human targets in radar data". In: *Signal Processing, Sensor Fusion, and Target Recognition XVI*. Ed. by Ivan Kadar. SPIE, Apr. 2007. DOI: 10.1117/12.718974.

[9] Kaustubh Kalgaonkar and Bhiksha Raj. "Acoustic Doppler sonar for gait recogination". In: *2007 IEEE Conference on Advanced Video and Signal Based Surveillance*. IEEE, Sept. 2007. DOI: 10.1109/avss.2007.4425281.

[10] Lawrence E. Kinsler et al. *Fundamentals of Acoustics*. Wiley, 1999. ISBN: 0471847895.

[11] D. Marioli et al. "Digital time-of-flight measurement for ultrasonic sensors". In: *IEEE Transactions on Instrumentation and Measurement* 41.1 (1992), pp. 93–97. DOI: 10.1109/19.126639.

[12] S. Maskell. "A tutorial on particle filters for on-line nonlinear/non-Gaussian Bayesian tracking". In: *IEE International Seminar Target Tracking: Algorithms and Applications*. IEE, 2001. DOI: 10.1049/ic:20010246.

[13] Pierre Del Moral. "Nonlinear filtering: Interacting particle resolution". In: *Comptes Rendus de l'Académie des Sciences - Series I - Mathematics* 325.6 (Sept. 1997), pp. 653–658. DOI: 10.1016/s0764-4442(97)84778-7.

[14] Robb J. Muirhead. *Aspects of Multivariate Statistical Theory*. John Wiley & Sons, Inc., Mar. 1982. ISBN: 9780471094425. DOI: 10.1002/9780470316559.

[15] Murata. *Murata MA40S4R Datasheet*. Oct. 31, 2008. URL: http://www.produktinfo.conrad.com/datenblaetter/500000-524999/506193-da-01-en-ULTRASCHALL_SENSOR_MURATA_MA40S4R.pdf (visited on 02/01/2018).

[16] Victor Y. Pan and Zhao Q. Chen. "The complexity of the matrix eigenproblem". In: *Proceedings of the thirty-first annual ACM symposium on Theory of computing - STOC '99*. ACM Press, 1999. DOI: 10.1145/301250.301389.

[17] Anna A. Stec. "Fire toxicity – The elephant in the room?" In: *Fire Safety Journal* 91 (July 2017), pp. 79–90. DOI: 10.1016/j.firesaf.2017.05.003.

[18] D A Waters. "Echolocation in air: Biological systems, technical challenges, and transducer design". In: *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 221.10 (Oct. 2007), pp. 1165–1175. DOI: 10.1243/09544062jmes504.

A Measurements that were used with stills from the scene

In this appendix, the measurements that were used in the analysis of the filters have been collected. Their spectrograms have been plotted and in addition, we have included images from the video that was taken while measuring. We have included one image per second, taken in between every full second (0.5 s, 1.5 s, 2.5 s, ...). The videos were recorded separately from the sound so even though we have carefully tried to align the video and audio a well a possible, one should refrain from drawing conclusions that are based on very time specific features in the video. The images give a good indication of the scene and the type of movement though.

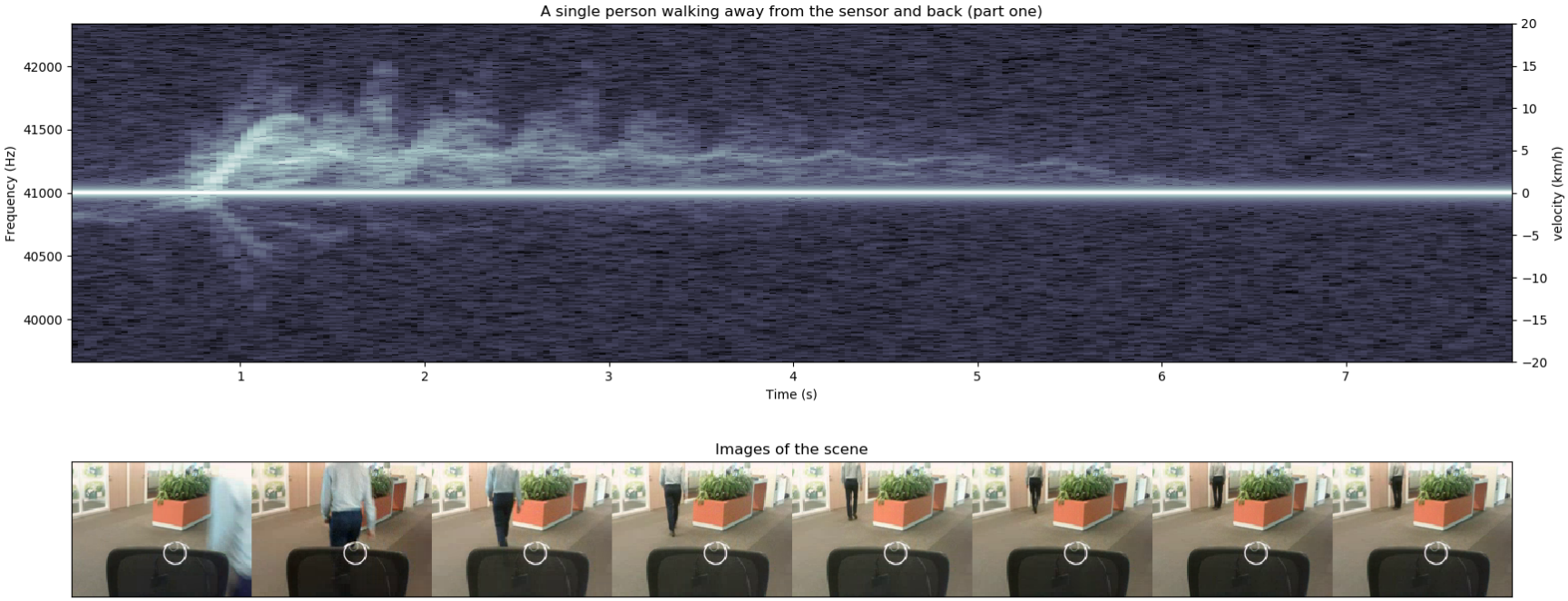


Figure 29: In this measurement a single person walks away from the sensor. This is the first part of a longer measurement.

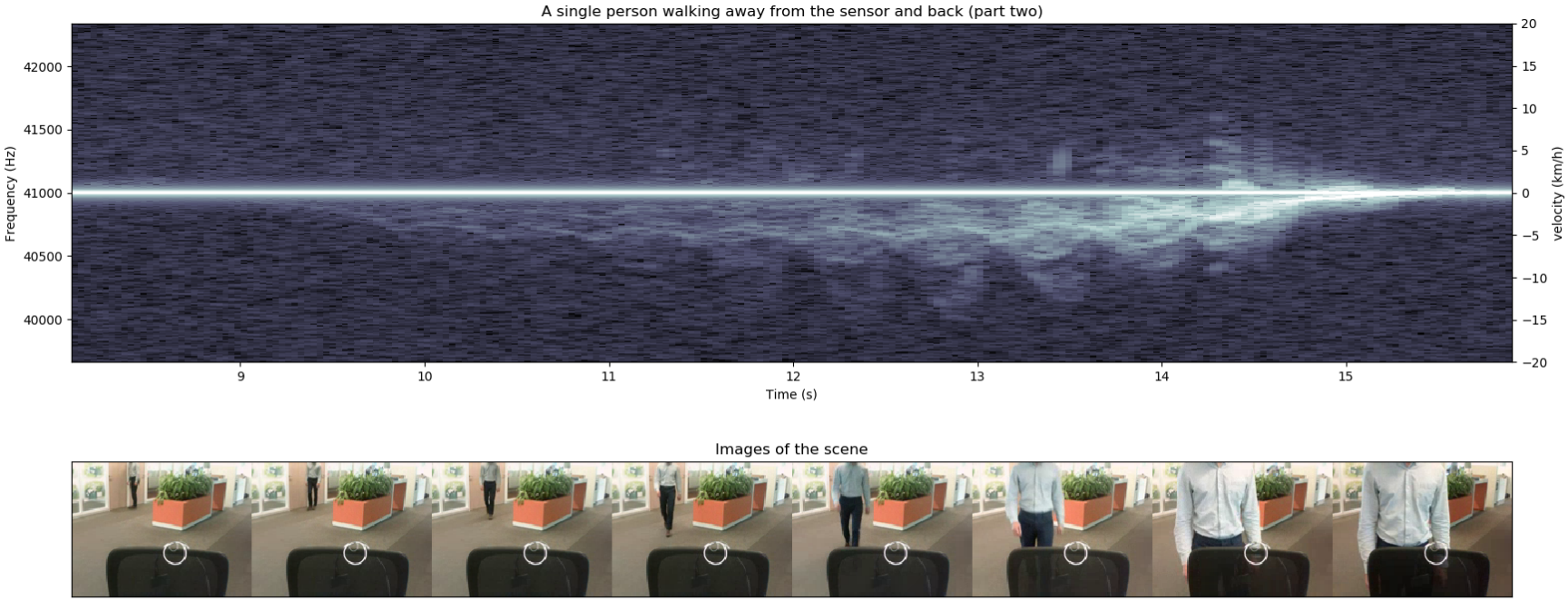


Figure 30: In this measurement a single person walks towards the sensor. This is the second part of a longer measurement.

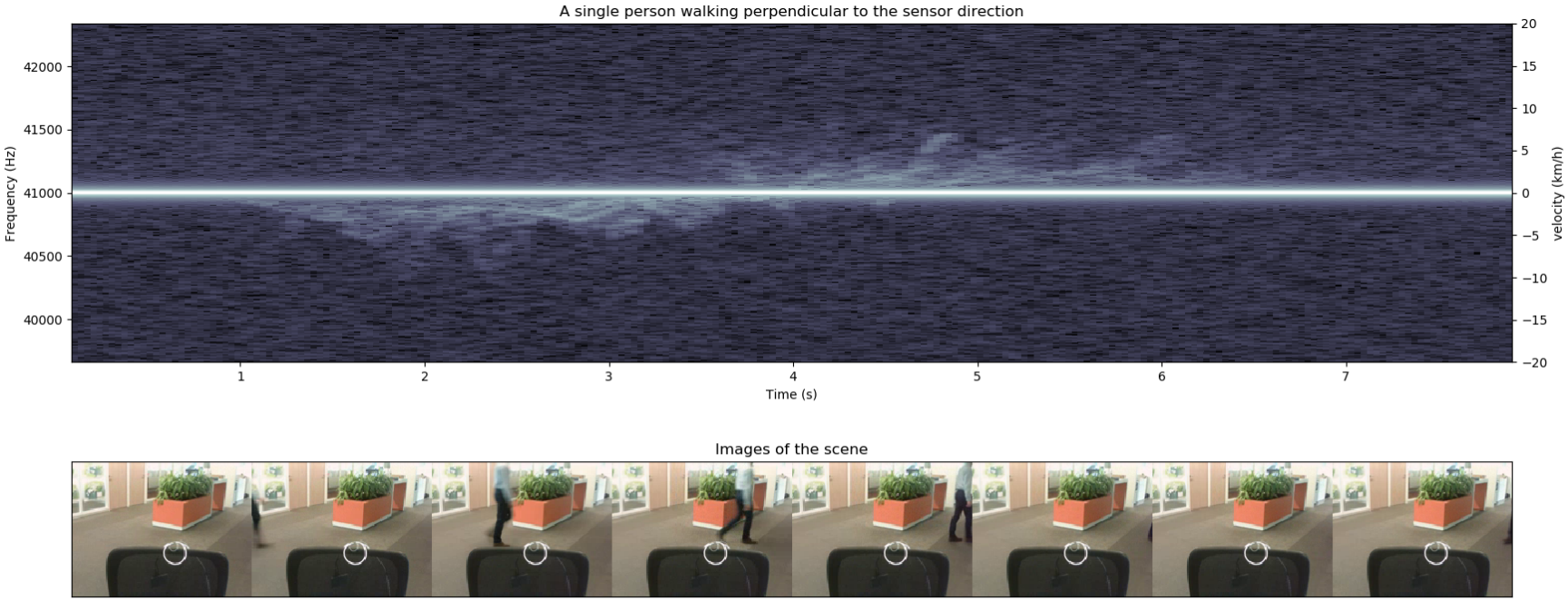


Figure 31: In this measurement a single person walks perpendicular to the direction of the sensor.

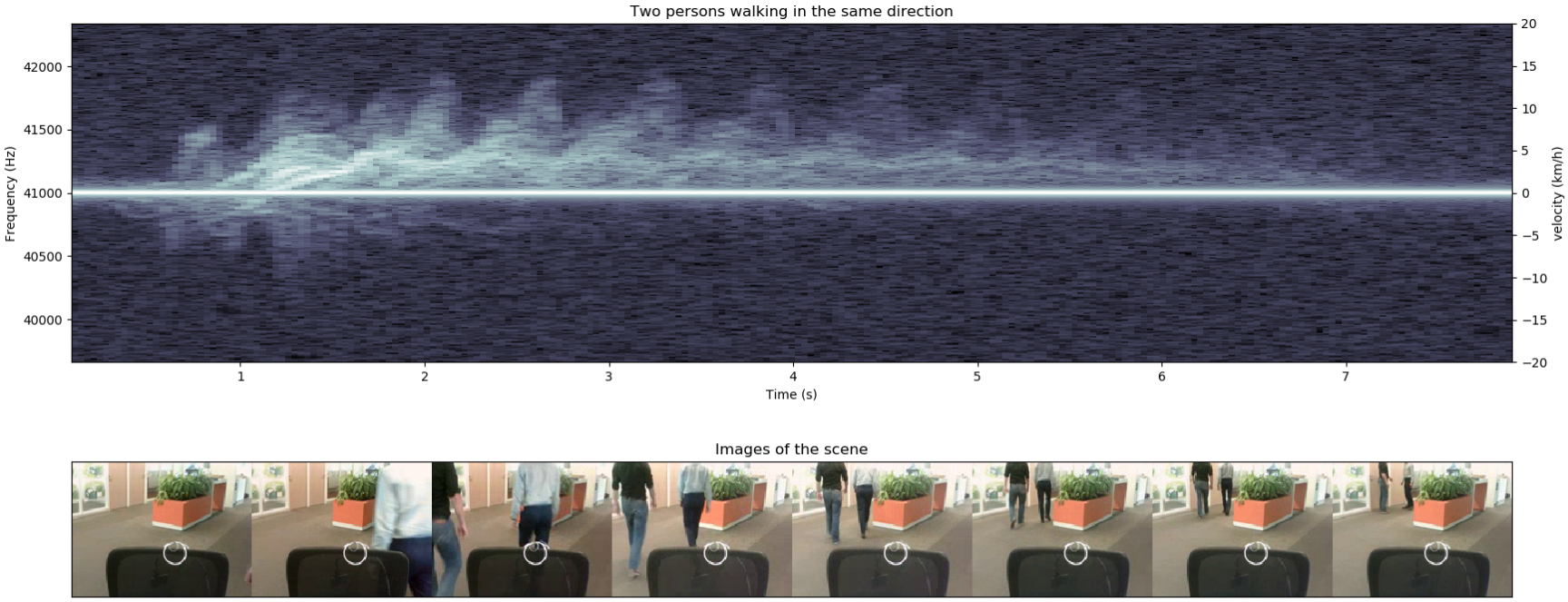


Figure 32: In this measurement two persons walk away from the sensor together. Looking closely at their legs, we can see that their walking phases are aligned.

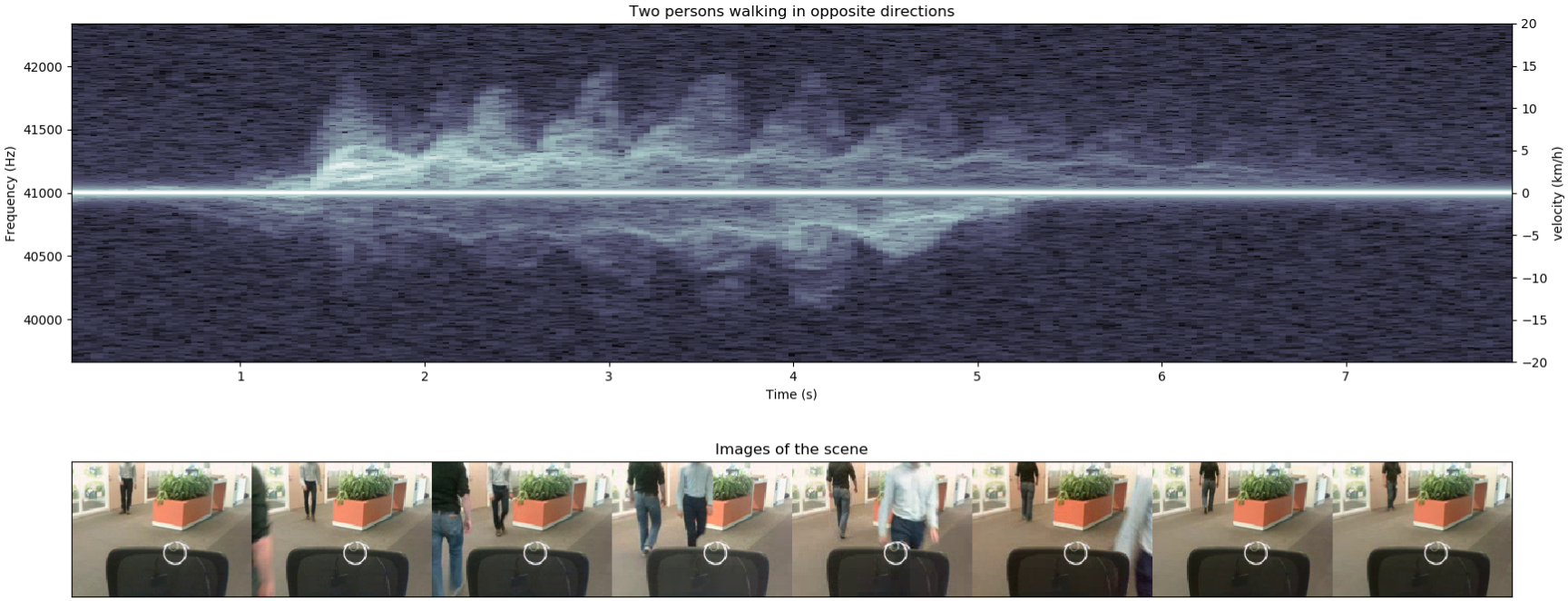


Figure 33: In this measurement two persons are walking in opposite directions.

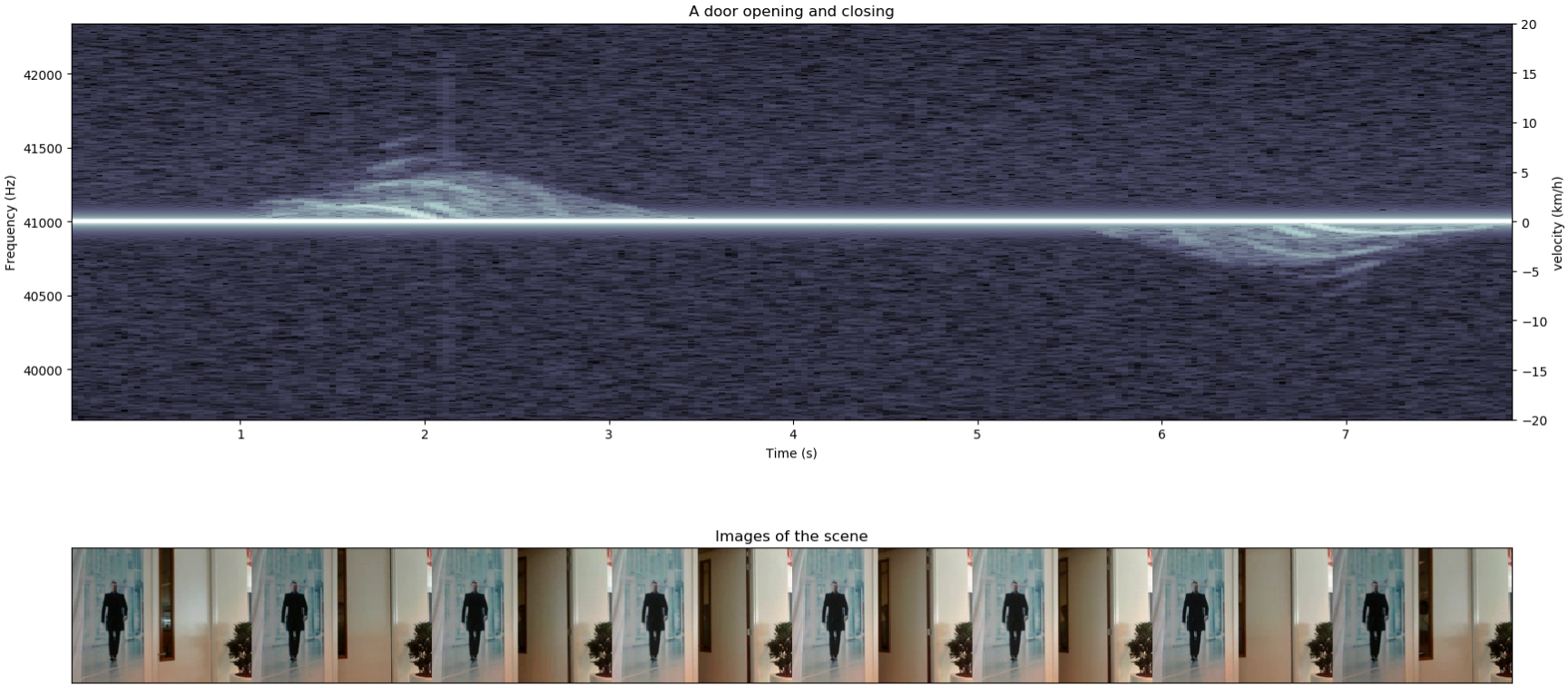


Figure 34: In this measurement, a door is opened and closed. The persons moving the door is always behind the door so the measurement only captures the door itself. Please note that the walking man on the left in each image is just a picture on the wall.

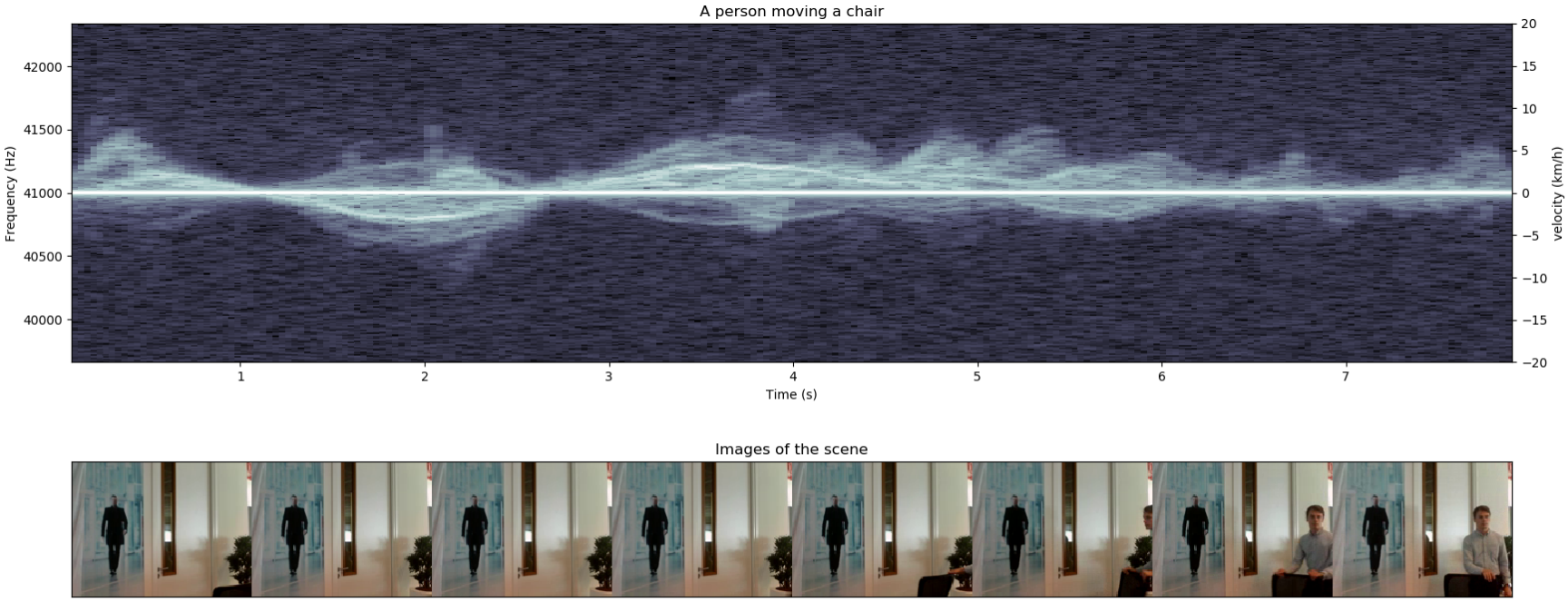
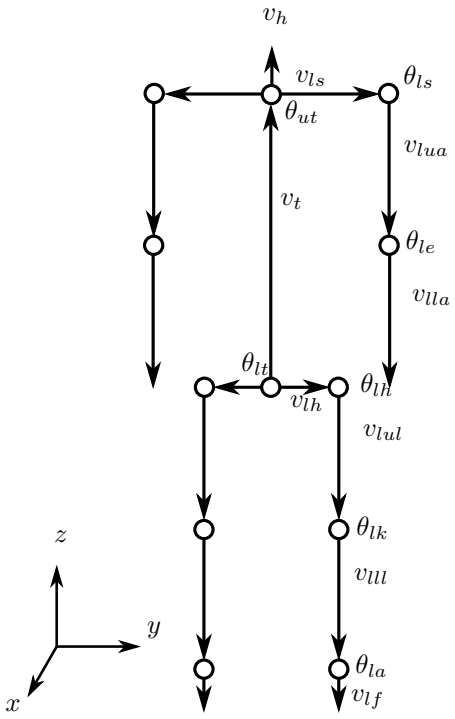


Figure 35: In this measurement is moving a chair. The person himself is also moving his body to move the chair, but he is not walking. Please note that the walking man on the left in each image is just a picture on the wall.

B The Thalmann model

The body is defined as follows.



A hierarchical model is used to calculate the positions of the joints when they have moved from the starting position. In this model, the joint in the lower thorax is used as the starting point. In other words, we use this joint to fix our coordinate system to so that it is always in the point $(0, 0, 0)$ We illustrate how to find a position by means of an example.

Suppose we want to calculate the position of the hand. Then we start off with the vector that is at the end of the arm, the lower arm. Rotate it by the elbow rotation, add the vector of the upper arm, rotate the result by the shoulder rotation, add the shoulder vector, rotate by the thorax rotation, add the thorax vector and finally rotate by the forward hip rotation. If we set $R_{joint\ name}$ for the rotation matrix corresponding to the joint $\theta_{joint\ name}$, then we can write this operation as

$$R_{lt}(v_t + R_{ut}(v_{rs} + R_{rs}(v_{rua} + (R_{re}v_{rla})))).$$

When we remove the brackets, we get

$$R_{lt}v_t + R_{lt}R_{ut}v_{rs} + R_{lt}R_{ut}R_{rs}v_{rua} + R_{lt}R_{ut}R_{rs}R_{re}v_{rla}. \tag{88}$$

Hence we see that we can simply apply the rotations of all joints to which a vector is connected sequentially to obtain the correctly rotated version of this vector.

In the simulation we use this property to calculate the location of each of the joints over time. We define the body as a tree of nodes and recurse trough the tree, starting from θ_{lt}. At each node we calculate its location using the location of its parent plus the rotated version of the displacement from its parent, v_{name}.

For the movement, we assign functions of time to the joints that describe their rotation (in degrees, positive orientation) around either the x, y or the z axis. We will define the movement functions below. We only specify only the functions for the left body parts. The functions for the joints on the right side of the body can be found by substituting t_r by $t_r - 0.5$. All function that are defined here are periodic in t_r on the interval $[0, 1)$ First we define these expressions for convenience:

$$V_r = v/(0.53H) \tag{89}$$
$$R_c = 1.346 \cdot \sqrt{V_r} \tag{90}$$
$$T_c = R_c/V_r \tag{91}$$
$$T_s = 0.752 * T_c - 0.143 \tag{92}$$
$$t_r = t/T_c \tag{93}$$

The forward-backward moving of the torso (around the y axis) at each step.

$$a_{TorsoY} = \begin{cases} -8V_r^2 + 8V_r & \text{for } V_r < 1/2 \\ 2 & \text{for } V_r \geq 1/2 \end{cases} \tag{94}$$

$$\theta_{TorsoY} = a_{TorsoY} - a_{TorsoY} \cdot \sin(2\pi \cdot (2t_r - 0.1)) \tag{95}$$

The up and down movement of the hips that makes the hips move down at the standing leg.

$$a_{PelvisX} = 1.66 \cdot v_r \tag{96}$$

$$\theta_{PelvisX} = a_{PelvisX} \cdot \begin{cases} -1 + \cos(2\pi \cdot 10 \cdot t_r/3) & \text{for } 0 \leq t_r \bmod 1 < 0.15 \\ -1 - \cos(2\pi \cdot 10 \cdot (t_r - 0.15)/7) & \text{for } 0.15 \leq t_r \bmod 1 < 0.5 \\ 1 - \cos(2\pi \cdot 10 \cdot (t_r - 0.5)/3) & \text{for } 0.5 \leq t_r \bmod 1 < 0.65 \\ 1 + \cos(2\pi \cdot 10 \cdot (t_r - 0.65)/7) & \text{for } 0.65 \leq t_r \bmod 1 < 1 \end{cases} \tag{97}$$

Rotation of the shoulder to move the arm (around the y axis).

$$\theta_{Shoulder} = -3 \cdot 9.88 \cdot v_r \cdot \cos(2\pi \cdot t_r) \tag{98}$$

Rotation of the elbow to move the lower arm relative to the upper arm (around y axis).

$$\theta_{Elbow} = -23 + 11 \cdot v_r \cdot \cos(2\pi \cdot t_r) \tag{99}$$

Some of the functions are defined as cubic spines through control points. In table 1 an overview of the control points can be found. In section B.1 the method for building the splines from the control points can be found. Besides expression for the rotation of the joints, there are also some linear movement

The rotation of the hip around the y-axis θ_{HipY}:

θ_{HipY}	-15	25
t_r	0.5	0.9

The rotation of the knee around the Y axis θ_{kneeY}:

θ_{kneeY}	3	20	70	3
t_r	0.18	0.4	0.75	0.99

The rotation of the ankle around the Y axis θ_{ankleY}:

θ_{ankleY}	-3	-18	8	-20	-3
t_r	0	0.08	0.5	0.65	0.85

The rotation of the torso around the z axis θ_{torsoZ}:

θ_{torsoZ}	1.4	-1.5	-1.4	1.5
t_r	0.1	0.4	0.6	0.9

Table 1: List of the control points that define the splines used for these rotation functions.

expressions that describe how the body accelerates and decelerates during each step. These movement expressions are all movements are small deviations from the general walking velocity of the person such that their total displacement within one cycle remains 0. We will call these translations of the body in the x, y and z direction T_x, T_y and T_z.

$$T_z = 0.015 \cdot V_r \cdot (-1 + \sin(2\pi(2t_r - 0.35)) \tag{100}$$

$$T_y = \begin{cases} (0.128V_r^2 - 0.12V_r) \cdot \sin(2\pi(t_r - 0.1)) & \text{for } V_r < 0.5 \\ 0.032 \cdot \sin(2\pi(t_r - 0.1)) & \text{for } V_r \geq 0.5 \end{cases} \tag{101}$$

$$T_x = \begin{cases} (0.084 \cdot V_r^2 - 0.084 \cdot V_r) \cdot \sin(2\pi(2t_r + 2(0.625 - T_s))) & \text{for } V_r < 0.5 \\ 0.021 \cdot \sin(2\pi(2t_r + 2(0.625 - T_s))) & \text{for } V_r \geq 0.5 \end{cases} \tag{102}$$

B.1 Splines

For the some rotation functions, cubic splines through control points are used. Splines are piecewise polynomials that connect the control points in such a way that the value of their 0th until 2nd derivative in the control points are equal. Suppose we have control points (x_i, y_i) for $i \in \{0, ..., n-1\}$. Then the piecewise polynomials are

$$\begin{aligned} &S_i(x) = a_i + b_i(x - x_i) + c_i(x - x_i)^2 + d_i(x - x_i)^3 \\ &\text{for } x \in (x_{i-1}, x_i] \text{ when } i \in \{0, ..., n-1\}. \end{aligned} \tag{103}$$

where we define

$$x_{-1} = x_{n-1} - 1. \tag{104}$$

The reason for this is the following: We want to make the spline such that it is periodic on $[0, 1)$ while preserving the differentiability. Therefore we repeat the point x_{n-1} below zero by adding $(x_{-1}, y_{-1}) = (x_{n-1} - 1, y_{n-1})$ to the control

points to make S_0 connect to it smoothly when we generate the spline. Then we add

$$S_n(x) = S_0(x-1) \quad \text{for } x \in (x_{n-1}, 1) \tag{105}$$

to smoothly finish the support in $[0, 1)$.

In order to find the coefficients for the splines, we use the equations for the differentials in the control points. The easiest relation is

$$y_i = S_i(x_i) = a_i + b_i(x_i - x_i) + c_i(x_i - x_i)^2 + d_i(x_i - x_i)^3 = a_i \tag{106}$$

which immediately shows that

$$a_i = y_i \quad \text{for } i \in \{0, 1, ..., n-1\}. \tag{107}$$

Then using the equality of adjacent polynomials at the control points gives the following three relations for $i \in \{1, ..., n-1\}$.

$$S_i(x_i) = S_{i-1}(x_i) \tag{108}$$
$$\implies a_i = a_{i-1} + b_i(x_i - x_{i-1}) + c_{i-1}(x_i - x_{i-1})^2 + d_{i-1}(x_i - x_{i-1})^3 \tag{109}$$
$$\implies a_i - a_{i-1} = b_{i-1}(x_i - x_{i-1}) + c_{i-1}(x_i - x_{i-1})^2 + d_{i-1}(x_i - x_{i-1})^3 \tag{110}$$

$$S_i'(x_i) = S_{i-1}'(x_i) \tag{111}$$
$$\implies b_i = b_{i-1} + 2c_{i-1}(x_i - x_{i-1}) + 3d_{i-1}(x_i - x_{i-1})^2 \tag{112}$$
$$\implies 0 = -b_i + b_{i-1} + 2c_{i-1}(x_i - x_{i-1}) + 3d_{i-1}(x_i - x_{i-1})^2 \tag{113}$$

$$S_i''(x_i) = S_{i-1}''(x_i) \tag{114}$$
$$\implies 2c_i = 2c_{i-1} + +6d_{i-1}(x_i - x_{i-1}) \tag{115}$$
$$\implies 0 = -2c_i + 2c_{i-1} + 6d_{i-1}(x_i - x_{i-1}) \tag{116}$$

For $i = 0$ the relations reduce to

$$a_0 - a_{n-1} = b_{n-1}(x_0 - x_{-1}) + c_{n-1}(x_0 - x_{-1})^2 + d_{n-1}(x_0 - x_{-1})^3 \tag{117}$$
$$0 = -b_0 + b_{n-1} + 2c_{n-1}(x_0 - x_{-1}) + 3d_{n-1}(x_0 - x_{-1})^2 \tag{118}$$
$$0 = -2c_0 + 2c_{-1} + 6d_{-1}(x_0 - x_{-1}). \tag{119}$$

We can turn this into a matrix-vector equation by putting all the unknown a_i, b_i and c_i in a vector. First define

$$A = \begin{pmatrix} 0 & 0 & 0 \\ -1 & 0 & 0 \\ 0 & -2 & 0 \end{pmatrix} \tag{120}$$

$$B_k = \begin{pmatrix} (x_k - x_{k-1}) & (x_k - x_{k-1})^2 & (x_k - x_{k-1}) \\ 1 & 2(x_k - x_{k-1}) & 3(x_k - x_{k-1})^2 \\ 0 & 2 & 6(x_k - x_{k-1}) \end{pmatrix}. \tag{121}$$

Then

$$\begin{pmatrix} A & & & B_0 \\ B_1 & A & & \\ & \ddots & \ddots & \\ & & B_{n-1} & A \end{pmatrix} \begin{pmatrix} b_0 \\ c_0 \\ d_0 \\ \vdots \\ b_{n-1} \\ c_{n-1} \\ d_{n-1} \end{pmatrix} = \begin{pmatrix} a_0 - a_{n-1} \\ 0 \\ 0 \\ a_1 - a_0 \\ 0 \\ 0 \\ \vdots \\ a_{n-1} - a_{n-2} \\ 0 \\ 0 \end{pmatrix} \tag{122}$$

When we solve this equation we get the coefficients.