



DEGREE PROJECT IN ELECTRICAL ENGINEERING,
SECOND CYCLE, 30 CREDITS
STOCKHOLM, SWEDEN 2021

FMCW mmWave Radar for Detection of Pulse, Breathing and Fall within Home Care

AXEL TRANGE

Abstract

Countless of elderly people fall and get hurt within their homes, worldwide, every year, and as they can not always reach out for help themselves, they end up helplessly waiting for someone to notice what has occurred.

Throughout this work, it is investigated if remote sensing of the mmWave FMCW radar IWR6843AOPEVM can be adopted to detect the incident of falls, and also detect the vital signs of the human subject. The goal is to prove that this is possible in a home care environment.

By locating the sub-range resolution oscillatory motions, caused by breathing and heartbeats, and unwrapping consecutive phase measurements of multiple range bins and multiple virtual antennas, human heart rate for seated position is estimated with a RMS-Error of 3.49 beats/min (at 1.0 m range with 120° FoV) for a 130 second time sequence. Analogously, breathing rate is estimated for a sitting person with a RMS-Error of 0.29 breaths/min (at 4.0 m range with 120° FoV) for a 100 second time sequence. Different estimation methods are evaluated, such as Fourier transform (FFT), Chirp Z-transform (CZT) and autocorrelation peak-finding, where the CZT approach is deemed to provide the best estimations. Methods are presented to minimize spectral leakage, improve spectral resolution and reduce breathing harmonics. The measurements were performed in a home care environment and the heart rate ones were compared to measurements of the FDA approved pulse oximeter CMS50D+.

729 own recordings of range-doppler-time data was collected for the fall detection, which was fed into a convolutional neural network to extract image features. These features were then used as training and softmax classified by a LSTM recurrent neural network for multi-label classification. Promising results on separate test data showed a balanced accuracy for fall detection as 92% with a direct 2% false positive rate and 15% false negative rate. The area under the ROC curve for the falls was close to 1, namely 0.99, illustrating that the false negative rate may be chosen as lower at the cost of slightly more false alarms. A sweet spot in the ROC curve suggested that fall detection was possible with a 3.5% false positive rate and 6% false negative rate.

Sammanfattning

Många äldre människor faller och skadas inom sina hem, världen över, varje år, och eftersom de inte alltid kan hämta hjälp själva, får de hjälplöst vänta på att någon ska upptäcka vad som har skett.

I detta arbete undersöks det om mmWave FMCW-radarn IWR6843AOPEVM kan användas för att upptäcka fall inom hemmet, samt även puls och andning, med syfte att kunna larma om någonting går snett. Målet är att visa att detektion är möjligt i en hemvårdsMiljö.

Genom att lokalisera ett föremål som oscillerar med avstånd mindre än radarns avståndsupplösning, och kombinera på varandra följande fasmätningar inom flera avstånd, lyckades puls bestämmas för en sittande människa med en RMSE av 3.49 slag/min (på 1.0 m avstånd med 120° FoV) för en sekvens av 130 sekunder. På liknande sätt bestämdes andningsfrekvensen av en sittande person till en noggrannhet motsvarande RMSE av 0.29 andetag/min (på 4.0 m avstånd med 120° FoV) för en sekvens av 100 sekunder. Olika uppskattningsmetoder av frekvenserna utvärderas, så som Fourier transform (FFT), Chirp Z-transform (CZT) och en autokorrelations-metod, där CZT-metoden bedöms ge bäst resultat. Metoder presenteras för att minska störningar i frekvensdomän, förbättra upplösningen i frekvensdomän och för att reducera övertoner från andnings-signalen. Mätningarna utfördes i en hemvårdsMiljö och uppskattningarna för hjärtfrekvens jämfördes med den mätta frekvensen av den FDA-godkända pulsoximetern CMS50D+.

729 egna inspelningar av avstånd-doppler-tid-data samlades in för falldetekteringen, som matades in i ett neuralt faltnings-nätverk för att extrahera egenskaper från bilder. Dessa egenskaper användes sedan som träning med softmax klassificering i ett återkopplat neuralt nätverk med LSTM celler för klassificering av flertal gester. Lovande resultat på skild testdata fick den viktade noggrannheten för falldetektering till 92% med 2% falska positiva och 15% falska negativa förutsägelser. Arean under ROC-kurvan för fallen var nära 1, nämligen 0.99, vilken visar att mängden falska negativa förutsägelser kan minska på bekostnad av att acceptera fler falsklarm. Förslagsvis kunde en implementering uppfylla 3.5% falsklarm och 6% missade fall.

Contents

1	Introduction	1
1.1	Conventional Approaches	3
1.2	Project Aim	3
2	Background	5
2.1	The Radar Principle	5
2.2	FMCW Radar Design and Processing	6
2.2.1	Transmitter- and Receiver Chain	7
2.2.2	Range-FFT and Pulse-Doppler Processing	9
2.2.3	In-Phase Quadrature Demodulation	12
2.2.4	MIMO Multiplexing and Virtual Antennas	13
2.3	Signal Characteristics of Heartbeats and Breaths	14
2.3.1	Breathing- and Cardiac Cycle	14
2.3.2	Breathing Harmonics	16
2.4	Gesture Recognition	17
2.4.1	Data Domains	17
2.4.2	Recurrent Neural Network	18
2.4.3	Long Short-Term Memory	20
3	Problem Formulation	22
3.1	Modelling the Vital Signs Problem	22
3.2	Formulating the Fall Detection Problem	24
4	The IWR6843AOPEVM Radar	27
4.1	Overview	27
4.2	Flashing Firmware to the Board	29
4.3	Configuring the Radar Subsystem	29
4.4	Communicating with the Radar	31

5 Methods	33
5.1 Finding & Estimating the Vital Signs	34
5.1.1 Vital Signs Setup	34
5.1.2 IQ Signal DC Correction	35
5.1.3 Vibrating Target Localization	37
5.1.4 Phase Unwrapping	38
5.1.5 Improving SNR for MIMO Architecture	39
5.1.6 Vital Signs Processing Chain	42
5.1.7 Suppressing Breathing Harmonics	44
5.2 Implementing Fall Detection	45
5.2.1 Recording Data	46
5.2.2 Preprocessing	49
5.2.3 Video Classification RNN Model	51
5.2.4 Training and Classification	53
6 Results	54
6.1 Vital Signs Results	54
6.2 Fall Detection Results	59
7 Discussion	62
7.1 Evaluation of Vital Signs Results	62
7.2 Evaluation of Fall Detection Results	63
7.3 Fields of Application	64
7.4 Future Work	65
7.4.1 Linking Harmonics for Improved Estimation Performance	65
7.4.2 Cascading of Radars	65
7.4.3 Machine Learning for Estimating Vital Signs	66
8 Conclusions	67
Acknowledgments	68
Bibliography	69
Appendix	72

Abbreviations and Acronyms

ADC Analog to Digital Converter.

AoA Angle of Arrival.

API Application Programming Interface.

BPM Beats Per Minute.

BR Breathing Rate.

CFAR Constant False Alarm Rate.

CPI Coherent Pulse Interval.

CW Continuous Wave.

DFT Discrete Fourier Transform.

DPU Data Processing Unit.

DSP Digital Signal Processing.

EM Electromagnetic.

FM Frequency Modulated.

FMCW Frequency Modulated Continous Wave.

FoV Field of View.

GPU Graphics Processing Unit.

HR Heart Rate.

HRV Heart Rate Variability.

IF Intermediate Frequency.

IQ In-Phase Quadrature.

LFM Linear Frequency Modulated.

LFMCW Linear Frequency Modulated Continous Wave.

LNA Low Noise Amplifier.

LO Local Oscillator.

LOS Line of Sight.

LSTM Long Short-Term Memory.

MCU Micro-Controller Unit.

MIMO Multiple Input Multiple Output.

MTD Moving Target Detector.

MTI Moving Target Indication.

NN Neural Network.

PA Power Amplifier.

PRF Pulse Repetition Frequency.

PRT Pulse Repetition Time.

RMSE Root Mean Square Error.

RNN Recurrent Neural Network.

ROC Receiver Operating Characteristic.

Rx Receiver.

SDK Software Development Kit.

SNR Signal-to-Noise Ratio.

TDM Time Division Multiplexing.

TLV Type Length Value.

Tx Transmitter.

UART Universal Asynchronous Receiver Transmitter.

Chapter 1

Introduction

Technical healthcare solutions are valuable to the society, for both humane and economical reasons. One of the most fundamental solutions is to through different means monitor and determine pulse or breathing of a living being, also known as the vital signs. There is benefit in studying these signs for multiple reasons.

Study may conclude presence or absence of instantaneous heart- and breathing rate, which could be indications of myocardial infarction or sudden cardiac arrest for the former case, and sleep apnea for the latter. Furthermore, long term study of the signs may be used to reveal underlying health issues. For instance, Heart Rate Variability (HRV), being the variation in period and amplitude of heart beats, can be measured. Evaluating this health measure through vital signs study could provide information and premonition to prevent future cardiac issues.

Great detection and monitoring devices are available at hospitals, but rarely elsewhere. However, most incidents occur at home in the households, and many of the affected are elderly people. Falling is the most common incident among the elderly and as many as one out of three over the age of 65 fall every year [1], resulting in injuries and declined quality of life. If no help is available, the outcome might even be death. Furthermore, it is also an expense for a society in terms of assistance and medical care that may cost up to US\$3,500 per month [1].

With this in mind, it seems pleasing to provide a single solution to both problems, both vital signs detection and fall detection. As the issue of falling and benefits of vital signs study is known, some solutions already exist. These solutions are for instance smartwatches, personal alarms, wristband sensors and video cameras, illustrated in Figure 1.1. However, many of these solutions

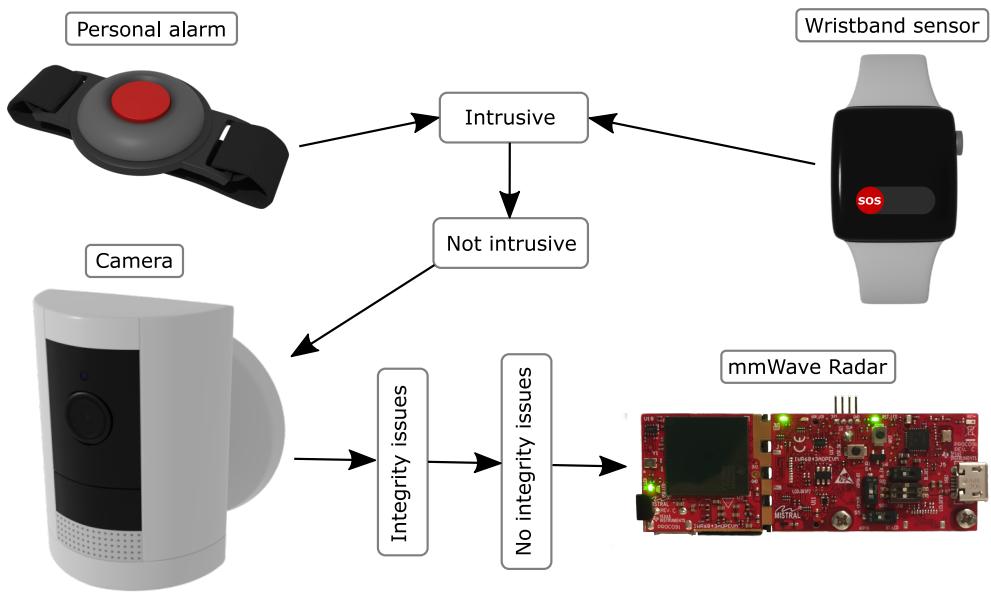


Figure 1.1: Different solutions for home care monitoring. Both wristbands have a problem of intrusiveness, and while the camera solves that problem it yet has an integrity issue. The radar, however, solves both the underlying issues.

have drawbacks. It has reportedly been shown that elderly dislike wearing attached sensors or wristbands because it has an element of intrusiveness. As a result, some people choose to not equip them, and there is always a risk of forgetting to.

This brings us to remote sensing. Here, one solution is using a camera to analyze a stream of images, and with the help of machine learning determine if a fall incident occurred or not. However, people are skeptical towards installing cameras inside their homes, in their bedrooms and bathrooms, because there are underlying integrity issues with it. The data captured by the camera could potentially be used in ways that people disapprove. Even if the manufacturers state that the data is not stored, people remain sceptic. There are also complaints of people feeling like they are being watched.

Hence, a new idea of solving this problem has emerged. The idea is to use radars for the home care monitoring. This solves the first issue since it is a form of remote sensing and requires nothing to be worn. It also solves the second issue since a) the radar can not tell exactly what you are doing and b) the radar does not reveal your identity by resolving your facial features like a camera would. Another advantage making radars attractive for fall detection is their sensing capabilities, that at certain frequencies may penetrate walls [2],

thus enabling robust detection.

1.1 Conventional Approaches

To conclude, the radar solves the issues of intrusiveness and integrity, while also providing a great foundation for measurements. Nonetheless, conventional approaches will here refer to how pulse, breathing and fall detection conventionally is determined with a radar. Starting with the fall detection, this has traditionally been investigated utilizing Continuous Wave radars, where one classifier has managed a result corresponding to a 0.974 area under the Receiver Operating Characteristic (ROC) curve [3]. Here an area of 1 corresponds to zero false alarms and 100% of predicted gestures being true. However, the CW radars can not determine distance and hence misses a dimension in recognizing gesture signatures. Hence there is a potential to achieve better results with FMCW radars.

For the vital signs detection, it has been shown that heart rate and breathing rate can be detected with FMCW radars. In [4] a 80 GHz radar was used, where the performance of the rates were estimated using Pearson correlation r_{XY} (more on this later) between reference signal and estimated signal yielding $r_{XY} = 0.80$ for the heart signal and $r_{XY} = 0.94$ for the breathing signal. In [5], a 9.6 GHz radar analogously resulted in $r_{XY} = 0.552$ for the heart rate, and in [6] $r_{XY} = 0.872$ was found for the heart rate and $r_{XY} = 0.91$ for the breathing rate.

1.2 Project Aim

The hardware that was available for this task was a radar from Texas Instruments, the IWR6843AOPEVM radar. Consequently throughout this work, it will be studied:

1. If the FMCW radar IWR6843AOPEVM can be used in standalone mode to detect gestures, especially falls, utilizing machine learning.
2. If the radar is suitable for detection of breathing and estimation of breathing rate.
3. If the radar suits the task of detecting heartbeats and estimation of heart rate.

Due to limited time and resources, the points above will be restricted to a single human target within the radar line of sight. Moreover, a couple of goals were set for the project, namely:

- The radar should be able to detect falls at ranges of up to 10 m.
- The fall detection should have an accuracy greater than 70%.
- Breathing rate estimation should be possible up to 5 m distance.
- Heart rate estimation should be possible up to 2 m distance.

Chapter 2

Background

To be able to detect vital signs and falls using a Frequency Modulated Continuous Wave (FMCW) radar, it is important to understand a couple of concepts. Firstly, it is necessary to understand how a FMCW radar operates, what it can measure and what its assets and limitations are. Secondly, it is essential to realize the characteristics of the motion being measured; such as its magnitude, velocity and periodicity. Lastly, since the fall detection part it greatly aided by some form of machine learning, some theory on neural networks shall also be covered. Throughout this chapter, a theoretical background will be given of the points explained above.

2.1 The Radar Principle

Radar stands for Radio Detection and Ranging. In simple terms, it consists of an antenna configured for both transmission and reception. This antenna, when powered and transmitting, sends an electromagnetic wave through space. Eventually, the transmitted wave collides with an object and scatters. As the scattered wave travels back to the antenna, now ready for reception, currents are induced in the antenna because of the time varying fields and signals are received. This concludes the Radio Detection part of the abbreviation. Note that what is sent from the antenna is an electromagnetic wave, hence traveling at the speed of light. Now, recall the very basic formula for the relation between speed (s), distance (d) and time (t):

$$d = s \cdot t. \quad (2.1)$$

Let us say that R is the distance to the scattering object and replace the speed with light speed, c . Keep in mind that the wave travels back and forth, giving

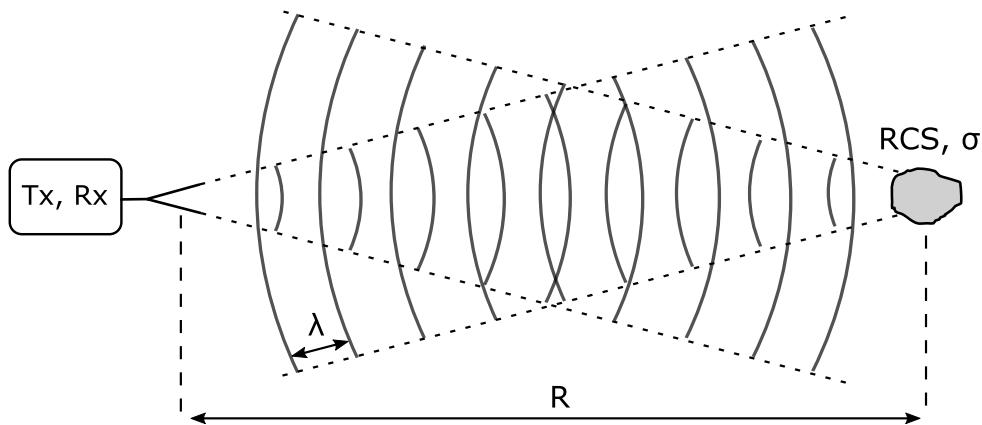


Figure 2.1: Antenna transmitting electromagnetic waves that are scattered on a target.

the total traveled distance as $2R$. If we summarize this, a familiar expression is obtained for the range:

$$R = \frac{c \cdot t}{2}. \quad (2.2)$$

Obtaining the time mentioned above obviously requires a more complex setup than *just* an antenna, and if a continuous wave is to be employed a way of differentiating between transmitted waves is required as well.

In Figure 2.1 it can be seen how a radar transmits and receives a scattered signal. The power that the radar receives is given by the following equation [7]

$$P_R = P_T \frac{G_T G_R \lambda^2}{(4\pi)^3 R^4} \sigma, \quad (2.3)$$

where P_T is the transmitted power, G_T and G_R are the gains of the transmitting and receiving antennas, respectively, σ is the radar cross section of the target, R is the distance to the target and λ is the wavelength of the transmitted signal.

2.2 FMCW Radar Design and Processing

A Frequency Modulated Continuous Wave (FMCW) radar is a certain type of radar, like Continuous Wave (CW) radar and pulse radar. What is special about a FMCW radar is that it can distinguish range, and does so by frequency modulating a continuous signal. With a procedure called IQ demodulation and by adopting multiple chirps it can even determine range, velocity and phase all at the same time for multiple targets. To conclude, this makes the FMCW radar a powerful sensor.

Below, the fundamental design of a FMCW radar will be explained. It will also be explained how it can operate with a sawtooth waveform and a pulse-doppler processing scheme to measure range, velocity and phase. General concepts such as radar Multiple Input Multiple Output (MIMO) and virtual antennas will also be covered.

2.2.1 Transmitter- and Receiver Chain

The design of a FMCW radar is composed of two blocks; a transmitter chain and a receiver chain. A simple block diagram of this can be seen in Figure 2.2. In the transmitter chain, it all starts with some form of *oscillator*. The oscillator's purpose is to generate a steady reference with set frequency and low phase noise. This can for instance be a *crystal oscillator*. However, this oscillator can rarely generate the desired frequency all by itself. Hence the signal is often modulated to desirable frequency modulation using a *synthesizer* [7], which may be a phased lock loop. For a Linear Frequency Modulated Continuous Wave (LFMCW) the frequency is linearly ramped by the synthesizer from a starting frequency f_0 to an end frequency $f_0 + B$, where B is the bandwidth. This is commonly known as a *chirp*, giving the instantaneous

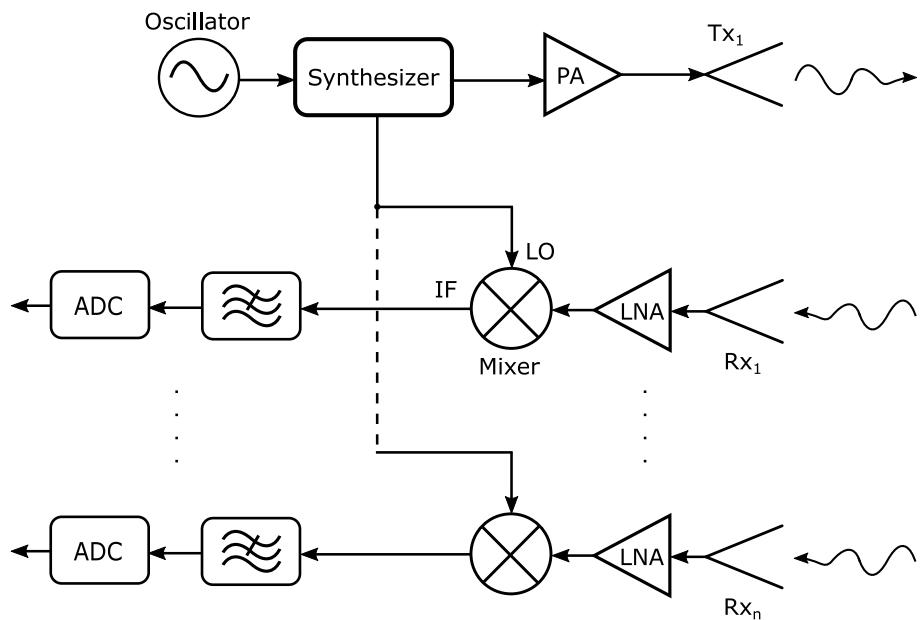


Figure 2.2: Design of a multiple (heterodyne) receiver FMCW radar system. Here the Transmitter (Tx) chain is illustrated at top and the n Receiver (Rx) chains at bottom.

frequency

$$f_i(t) = f_0 + St, \quad 0 \leq t \leq \frac{B}{S}, \quad (2.4)$$

where S is the rate of change of the chirp. Furthermore, the output of the synthesizer becomes of the form

$$x_T(t) = A_T \sin [\omega_T(t)t + \Phi_T(t)], \quad (2.5)$$

in which A_T is the amplitude of the signal, $\Phi_T(t)$ is the phase and $\omega_T(t) = 2\pi f_i(t)$ is the angular frequency of the chirp. Furthermore, this frequency modulated signal then passes through a Power Amplifier (PA) and often also a phase shifter. In the case for a Multiple Input Multiple Output (MIMO) radar with multiple transmitting- and receiving antennas, the phase shifter may be used either for beam steering analogous to traditional antenna arrays or for binary phase modulation.

Once the signal reaches the transmitting antenna, Electromagnetic (EM) fields start propagating through space. When these fields are backscattered from arbitrary target and reaches the receiving antenna, after a round-trip delay time t_d , voltage is induced in the receiver chain. This voltage, or signal, is typically of low power as can be understood from Equation (2.3). Moreover, received radar signals tend to have a low Signal-to-Noise Ratio (SNR). Hence noise suppression is critical [7] and the signal is amplified by a Low Noise Amplifier (LNA).

Next, the received signal is fed to a *mixer*, and so is the transmitting signal, or the Local Oscillator (LO) signal. The purpose of the mixer is to combine (multiply) the signals and downconvert them to a manageable and processable frequency, referred to as Intermediate Frequency (IF). This process is known as *heterodyning*, and one of the resulting terms contain the frequency and phase difference between the LO signal and received signal

$$x_{IF}(t) = A_R \sin [(\omega_T(t) - \omega_R(t))t + \Phi_T(t) - \Phi_R(t)] + U(t), \quad (2.6)$$

with $\omega_R(t) = \omega_T(t - t_d)$ being the received angular frequency, A_R the amplitude and $\Phi_R(t)$ the received phase of the EM wave. Once the IF signal is obtained, filters are used to reject unwanted mixer products and out-of-band signals [8]. The high frequency output terms of the mixer, $U(t)$, will be omitted in further discussion as the signal is low-pass filtered. Furthermore, the signal is sampled to a digital signal through an Analog to Digital Converter (ADC), giving the digital *baseband* signal

$$x_B[n] = A_R \sin [(\omega_T(nt_s) - \omega_R(nt_s))nt_s + \Phi_T(nt_s) - \Phi_R(nt_s)], \quad (2.7)$$

in which $t_s = 1/F_s$ is the ADC sampling period and $n = 0, 1, \dots, N_D - 1$ are the total samples limited by the total number of ADC samples N_D . After this step, the signal is typically processed with some digital radar processing to obtain measurement results.

2.2.2 Range-FFT and Pulse-Doppler Processing

There exists many different waveforms to use as frequency modulation for a FMCW signal. One that is used in many current mmWave radars is the linear sawtooth waveform. The processing of a filtered and sampled IF signal, originating from a sawtooth waveform, to obtain measurement results such as range and velocity will be explained here.

Figure 2.3 shows the transmitted LFM sawtooth signal (solid line) and a scattered received signal (dashed line) from a single object that is received after a round-trip delay time t_d . The bottom graph shows the result after mixing these signals and obtaining the IF signal. The frequency that shows up across the chirps is known as the *beat frequency*, and is for the static case directly proportional to the range. This is easily understood as when the range to an object increases, so does the round-trip delay time (t_d), and since the frequency of the

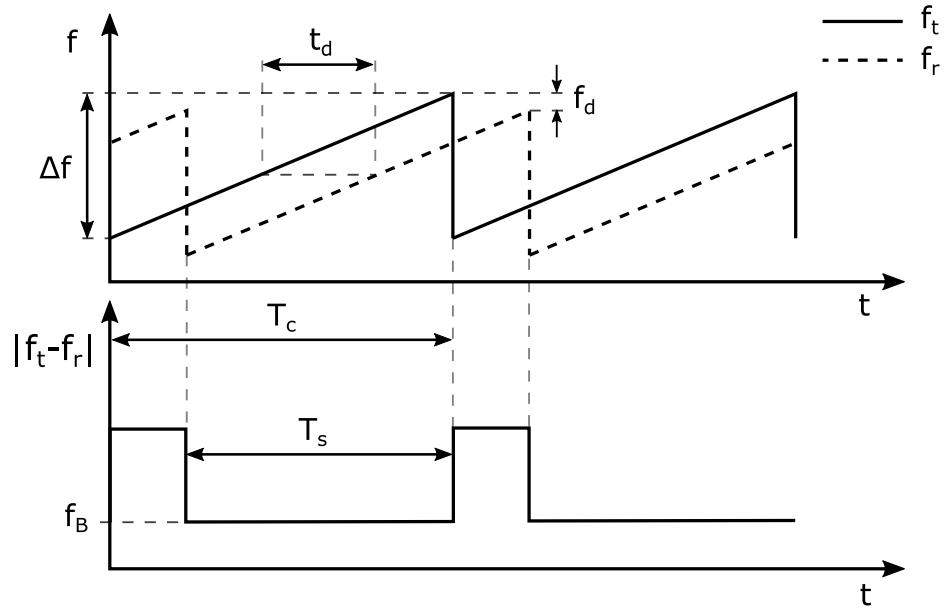


Figure 2.3: Transmitted and received sawtooth signal (top) and its corresponding intermediate frequency (IF) signal after mixing (bottom). f_t is the transmitted signal and f_r is the received.

received signal is a t_d time delayed replica of the transmitted chirping signal, the difference in frequency increases. This difference in frequency is the frequency of the IF signal, and hence its frequency increases too. Analogously if the object range decreases, the frequency of the IF signal decreases. For simplicity the IF signal is here illustrated to contain one frequency corresponding to a single target, but in reality the signal is composed of many different frequencies corresponding to scattered and received waves from multiple targets. Since the frequency of the sampled and filtered time domain IF signal, $x_B[n]$, is proportional to distance, the radial measured range of a single chirp can be found using a Discrete Fourier Transform (DFT), known as *range DFT*

$$X_k = \sum_{n=0}^{N_D-1} x_B[n] e^{-\frac{j2\pi}{N_D} kn}, \quad (2.8)$$

where N_D denotes the number of ADC samples, the index $k = 0, 1, \dots, N_D - 1$ corresponds to the range bin and the magnitude of X_k is analogous to the received echo strength. A clever way to write Equation (2.8) is by making k a function of range, such that multiples of the range resolution $r_{RES} = \frac{c}{2B}$ gives integers of k . This is achieved if $k = B \frac{2r}{c}$, giving [9]

$$X(r) = \sum_{n=0}^{N_D-1} x_B[n] e^{-\frac{j2\pi}{N_D} B \frac{2r}{c} n}. \quad (2.9)$$

However, in practical digital processing the DFT is often approximated using a fast fourier transform (FFT).

If there is a too large velocity induced doppler shift present, the saw-tooth frequency-modulated waveform can not accurately distinguish range in

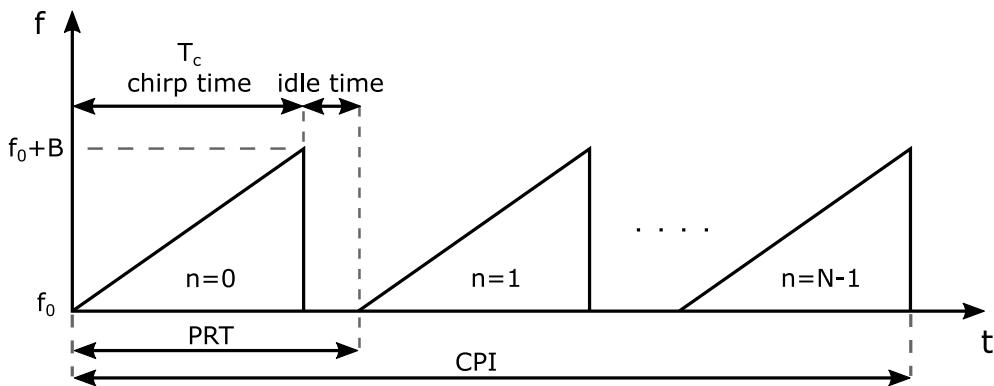


Figure 2.4: Multiple consecutive transmitted chirps over a Coherent Pulse Interval (CPI).

a straightforward manner, due to the offset beat frequency [10]. However, for fast-FMCW radars this effect is usually small and may be easily corrected after processing of the doppler-FFT [11]. The purpose of the doppler-FFT is to resolve different velocities, and it may be implemented by transmitting and measuring the phase of multiple consecutive chirps, similar to pulse-doppler processing. It may seem counter-intuitive to use pulse-doppler processing for a FMCW radar, since the words *continuous* and *pulse* in terms of radar may be contradicting. Nonetheless, as previously explored, pulse-doppler or MTI processing can in fact be adopted for a FMCW radar [12].

The idea of the pulse-doppler approach is to transmit a signal for N number of chirps over a Coherent Pulse Interval (CPI) with the same time between each pulse, also known as Pulse Repetition Time (PRT), here denoted T_{PRT} . See Figure 2.4 for illustration. Furthermore, once the received IF signals have been obtained for all the chirps of the CPI, a discrete Fourier transform may be carried out for measured phase differences to determine the velocity

$$V_m = \frac{\lambda}{4\pi T_{PRT}} \sum_{n=0}^{N-1} (\varphi[n+1] - \varphi[n]) e^{-\frac{j2\pi}{N} mn}, \quad N \geq 1, \quad (2.10)$$

where λ is the wavelength and the phase measurements $\varphi[n]$ and $\varphi[n+1]$ are given by arctangent demodulation of the I and Q signals. This will be explained

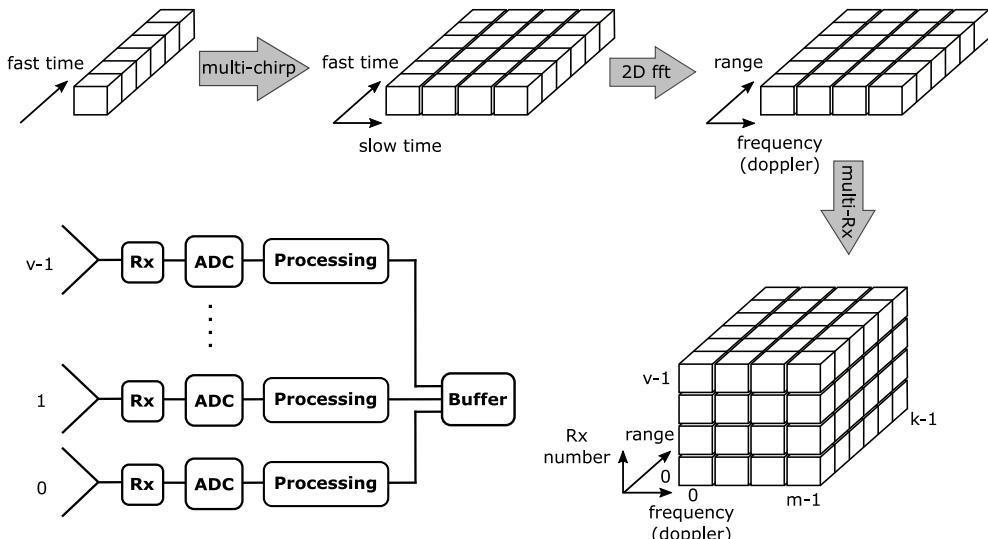


Figure 2.5: Illustration of how measurements are obtained. Multiple chirps, one for each PRT, build up a 2D space of fast time and slow time, after which a 2D-FFT yield range and doppler frequency. Combining the measurements of all receivers as a dimension results in the stored data; the radar cube data.

in Section 2.2.3. In addition, the index $m = 0, 1, \dots, N - 1$ corresponds to the doppler bins; convertible to velocity by the maximum unambiguous velocity $v_{max} = \frac{\lambda}{4T_{PRT}}$ and velocity resolution $v_{res} = \frac{\lambda}{2T_{CPI}}$.

Both range and velocity measurements may be combined, a process summarized as shown in Figure 2.5. Fast time (chirp) and slow time (pulses) build up a 2D space that, through a 2D-FFT across fast time and slow time, provides the range and doppler frequency (or velocity) measurements. For a MIMO radar this procedure is carried out for all the receiver elements (or virtual antennas), giving a data structure known as *radar datacube* [13].

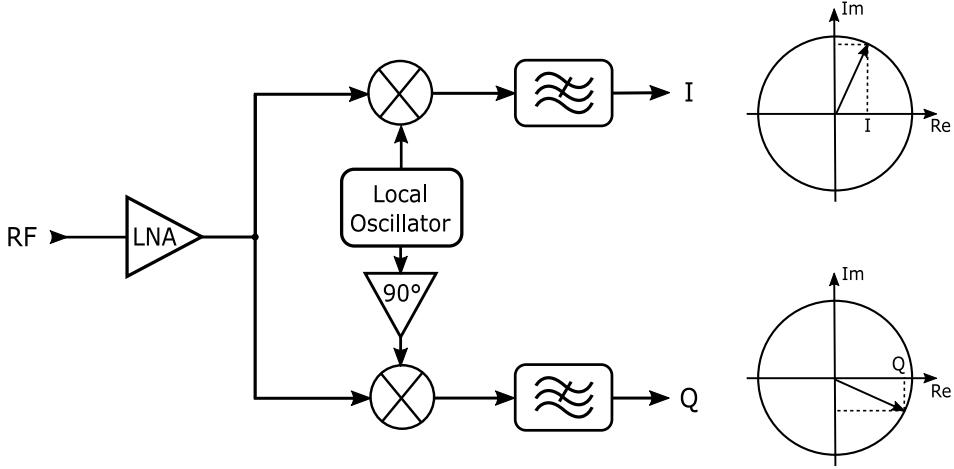


Figure 2.6: IQ demodulation procedure in the receiver chain.

2.2.3 In-Phase Quadrature Demodulation

In-Phase Quadrature (IQ) demodulation is a procedure that allows the radar to capture and distinguish phase information of a signal. It is achieved by splitting the output of the Local Oscillator (LO) into two parts, an in phase (I) part and a quadrature (Q) part. Here the Q signal is phase-shifted by 90 degrees (see Figure 2.6). After mixing both signals with the received signal down to IF, it can be interpreted as a signal split into a real and imaginary part

$$I(t) = A_R \cos \left[(\omega_T(t) - \omega_R(t))t + \Phi_T(t) - \Phi_R(t) \right], \quad (2.11)$$

$$Q(t) = A_R j \sin \left[(\omega_T(t) - \omega_R(t))t + \Phi_T(t) - \Phi_R(t) \right], \quad (2.12)$$

where $\omega_T(t)$ and $\omega_R(t)$ are the transmitted and received angular frequencies and $\Phi_T(t)$ and $\Phi_R(t) = \frac{4\pi R(t)}{\lambda}$ are the transmitted and received phases. Since

$\Phi_T(t)$ often is known or constant, phase difference in $\Phi_T(t) - \Phi_R(t)$ often corresponds to a change in $\Phi_R(t)$. It is worth noting that a complex representation is essential to obtain an unambiguous phase measurement; without it, there is no way to distinguish what part of the cycle the phase is in, since $\cos \Phi = \cos(-\Phi)$. Thus the IQ demodulation allows distinguishing between positive and negative phase shifts [8], with the phase of

$$\varphi(t) = \tan^{-1} \left[\frac{\text{Im}\{Q(t)\}}{\text{Re}\{I(t)\}} \right]. \quad (2.13)$$

2.2.4 MIMO Multiplexing and Virtual Antennas

In terms of a radar, Multiple Input Multiple Output (MIMO) simply means that there are multiple transmitters (input) and multiple receivers (output). This is interesting because a specific spatial setup of transmitters and receivers can give further independent information than a similar setup with more antennas. The reasoning behind this is that the former can form a greater number of *virtual antennas*.

However, to take advantage of multiple transmitters to form extra virtual antennas, it is essential for the transmitted signals to be mutually orthogonal to not interfere with another. Many schemes exist to achieve this, but a simple one is Time Division Multiplexing (TDM) [14]. With this scheme *only one* transmitting antenna is active at once in time. Moreover, if the receivers are continuously active and receiving signals from all transmitters, a matched filter can be performed between each individual transmitted waveform and the

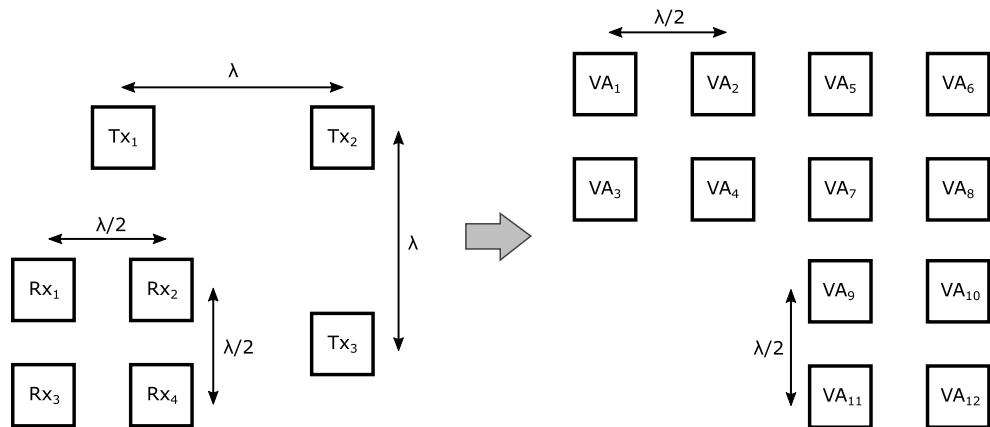


Figure 2.7: Example of resulting virtual antennas of 3 transmitters and 4 receivers. Here the transmitters have one wavelength spacing and receivers half wavelength spacing.

receivers to separate the signals into channels [15], where each channel corresponds to a virtual antenna.

An example of a MIMO setup can be seen in Figure 2.7. Here 4 receivers are carefully spaced $\lambda/2$, where λ is the wavelength, and the transmitters spaced with distance λ . With this spacing, the number of virtual antennas is given as

$$N_{VA} = N_{Tx}N_{Rx}, \quad (2.14)$$

where N_{Tx} and N_{Rx} is the number of transmitters and receivers, respectively. For the mentioned figure, $N_{Tx} \cdot N_{Rx} = 12$ virtual antennas are obtained. Hence this setup of 7 antennas provide equal data as a similar setup of 12 linearly spaced receivers and one transmitter would have.

2.3 Signal Characteristics of Heartbeats and Breaths

In this section a basic connection will be made between the physiological events of breaths and heartbeats, to their radar measured counterparts. Hence, here the signals will correspond to what a mm-wave radar can see; primarily the spatial displacement of the chest region due to its expansions and compressions over time.

2.3.1 Breathing- and Cardiac Cycle

The largest and most simple motion to detect is breathing, whose cycle consists of inhaling and exhaling. As air is sucked into the lungs the chest expands,

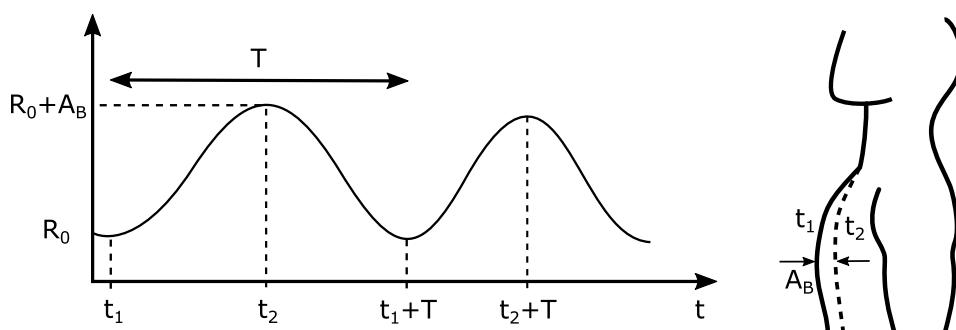


Figure 2.8: Illustration of breathing cycle as would be measured by a radar in front of a person. Here R_0 is the static distance to the radar, T denotes the period of a breath and A_B the peak-to-peak amplitude of the chest expansion.

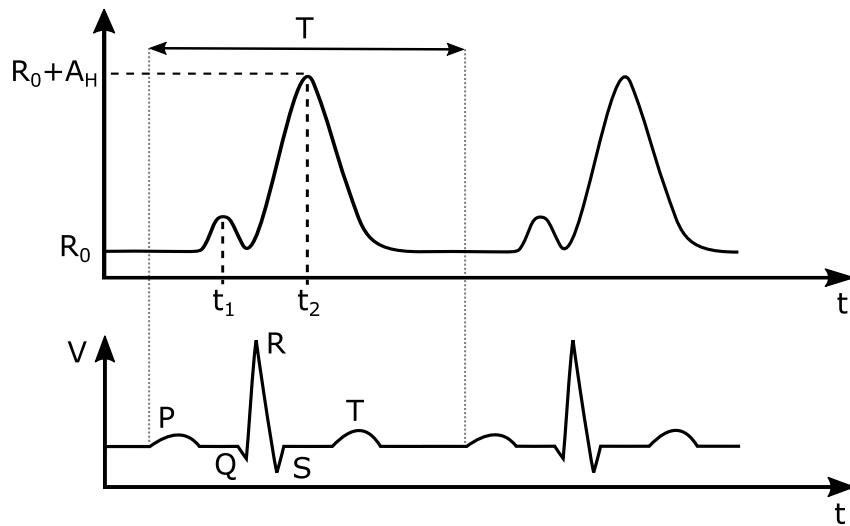


Figure 2.9: Illustration of cardiac cycle, as seen by the radar from generated vibrations (top) of the atrial- and ventricular systole. Here T denotes the period of a heartbeat and A_H the peak amplitude of the chest expansion. A QRS complex as would be seen on an ECG is shown as reference; indicating that atrial systole occurs post P-wave and ventricular systole post R-peak.

causing a measured displacement in the order of a few millimeters to a few centimeters, depending on the person [5] and the body's angle in relation to the radar. A typical breathing cycle can be seen in Figure 2.8. Here t_1 denotes the time of inhale and t_2 the time of exhale. Moreover, the total period T of a breath is typically between 2 s and 10 s, causing a breathing rate of 0.1-0.5 Hz (6-30 breaths/min). It is worth noting that albeit the motion may appear sinusoidal, it is not perfectly so. This causes harmonics, which will be explained more carefully in Section 2.3.2.

The measured waveform of heartbeats is a bit different. Due to its double beating nature the waveform for a period is bimodal, as seen in Figure 2.9. More specifically, the first peak, at time t_1 in the graph, corresponds to atrial systole. The aforementioned contraction causes a vibration located to the thorax region whose phase can be measured by the radar. The second and greatest vibration, at time t_2 , is caused by ventricular systole, occurring shortly after the ECG R-peak [16]. This vibration induced displacement can certainly be measured at the thorax region, but is not bound to it. In fact, as the blood rushes through the aorta and pulmonary arteries and enters the arteries throughout the entire body, then the entire body vibrates and pulses too, only with lower magnitude. These displacements are generally in the order of 0.1-

1.0 mm [5]. Furthermore, typical heart rate frequencies for people can range between 0.8-3.0 Hz [17], alternatively 48-180 beats/min.

Since the radar measures the total displacement of the body and chest region, the measured displacement caused by the breathing and heartbeats is a superposition of the two signals. In time domain it often looks like the signal in the top right corner of Figure 2.10.

2.3.2 Breathing Harmonics

In reality, the chest displacement due to breathing is not perfectly sinusoidal. This is due to irregularities in the breathing motion itself and the phase modulation of the radar. What this means is that the signal of the chest displacement cannot be explained by a single sinusoid with a set frequency. As can be recalled from Fourier analysis and specifically Fourier series, any signal can be constructed of a set of harmonically related sinusoids. With this in mind, by decomposing a time domain breathing signal into frequency domain, this should give a fundamental breathing harmonic, but also possibly an infinite set of harmonics. A typical example of this is shown in Figure 2.10, that shows the Fourier transform of a breathing signal and the first three breathing harmonics.

The part where this may become an issue, is when there is overlap between any of the breathing harmonics with the heartbeat fundamental harmonic. If

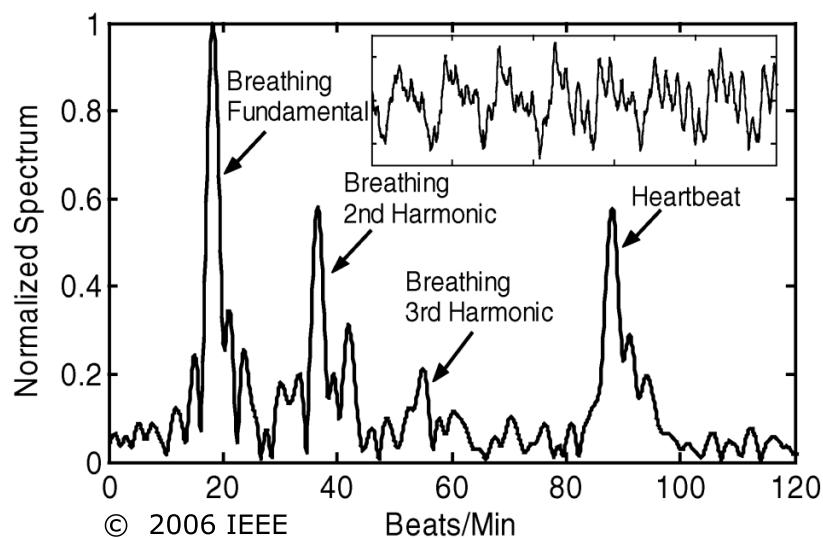


Figure 2.10: Frequency spectrum of chest movements measured by a 24 GHz doppler radar from front at 2 m distance. By [18].

the breathing harmonics are in the same frequency range as typical heartbeats and especially if they are of higher amplitude, then it may be difficult to discriminate the two signals. A mathematical model and explanation of the breathing harmonics, by Fourier series and Bessel functions, is further given in Section 3.1.

Of course the signal of the heartbeat has harmonics as well, but since the harmonics of solely heartbeats are at frequencies that are multiples of the fundamental heartbeat frequency, these harmonics do not interfere with any other fundamental harmonics.

2.4 Gesture Recognition

To interpret human gestures using mathematical algorithms is called gesture recognition. Just as this can be adopted to register human facial expressions or body signs, it may also be used to determine whether a person is falling or not. In this section it will be explained what useful signals a radar can record of a falling incident, and what neural networks might be useful for training a model on that recorded data.

2.4.1 Data Domains

Many different types of data may be captured by modern radars, for instance azimuthal- and elevation angles, range, doppler frequency and polarization [8]. In combination, these parameters can be used to determine a target's position in three dimensions, as well as velocity. Nonetheless, within this study the data will be limited to the scope of the one-dimensional, radial, range and doppler frequency.

The doppler-time domain has been extensively researched for presenting motions of humans, concluding that a *micro-doppler effect* is present for moving targets with vibrations or rotations, thus enabling detailed signature analysis for the domain [19]. However, as the 2D representations of doppler-time, range-time and range-doppler may provide distinct independent information [2], combining the dimensions to a 3D representation of range-doppler-time may provide a more accurate approach for capturing and analyzing human motion signatures. A visualization of this data domain can be seen in Figure 2.11 as reference.

This further entails that the added range dimension of a FMCW radar enables a more precise representation of a motion in comparison to a CW radar.

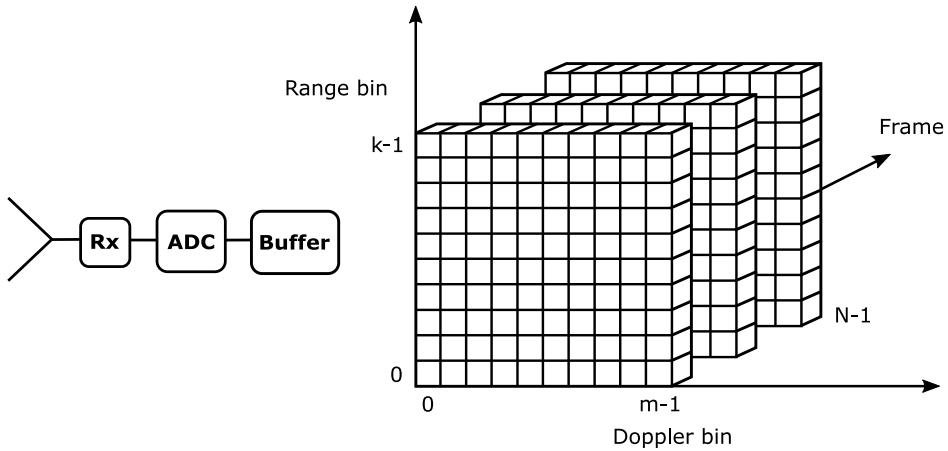


Figure 2.11: Representation of temporal data, range-doppler-time data, that may be captured by a single receiver chain of the FMCW radar.

However, keep in mind that this does not represent the entire, possibly captured, complexity of the motion signature. Many further dimensions of data may be captured by the radar, to the point where real-time processing and storage of the full-spaced data quickly becomes an issue. Hence many target detection methods exist, such as Constant False Alarm Rate [8], that enables higher dimensional processing for a limited space. Nonetheless, this topic will not be delved into.

2.4.2 Recurrent Neural Network

A Neural Network (NN) is a computing system used within machine learning where nodes process and transmit signals. The nodes are often clustered and referred to as layers, where the NN structure often is composed of input layer, a hidden layer and an output layer. The idea of the network is to, given some input experience, become better at a task, evaluated by improving a performance measure [20]. To make this tangible, a simple linear regression task can be taken as example; to map some input experience x to output y , estimated as \hat{y} [21]

$$\hat{y} = \mathbf{w}^T \mathbf{x} + b, \quad (2.15)$$

where \mathbf{w} and b , being weights and bias, are affected by the learning experience and minimizes the error $E = y - \hat{y}$ given further experience \mathbf{x} .

However, Equation (2.15) is too simple to attack most machine learning problems. An extended and yet simple approach is a convolutional network, that may be used for a temporal sequence to obtain some sharing of param-

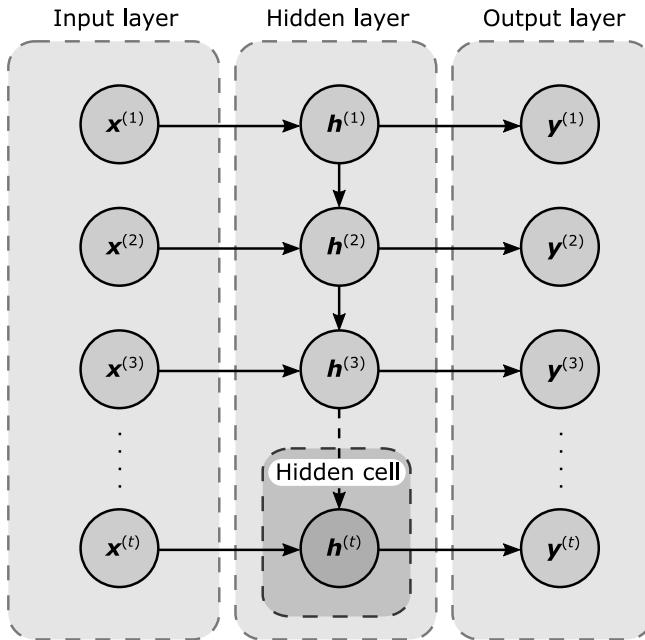


Figure 2.12: Illustration of an unfolded recurrent neural network.

eters across time. This network does thus have memorizing capabilities, but for many applications it is too shallow, as it only shares parameters with few neighbouring inputs [21].

This leads to the Recurrent Neural Network (RNN). This has a similar structure to a traditional Neural Network, but with some form of recurrence that allows it to memorize features across a sequence, e.g. a temporal sequence. Often in the hidden layer, some value is sent forward to the adjacent node

$$\mathbf{h}^{(t)} = f_1(\mathbf{x}^{(t)}, \mathbf{h}^{(t-1)}), \quad (2.16)$$

in which $\mathbf{x}^{(t)}$ is the input at time t and $\mathbf{h}^{(t)}$ is the output of the hidden layer at the mentioned time. The final output $\mathbf{y}^{(t)}$ is further $\mathbf{y}^{(t)} = f_2(\mathbf{h}^{(t)})$, where f_2 , being the final function in the network, often is an activation function of some sort (possibly softmax). An example of a simple RNN with one output per input is shown in Figure 2.12. The input affects the state \mathbf{h} , which is passed forward through time and acts like a memory. Note that the input $\mathbf{x}^{(t)}$ at a specific time τ rarely is a single value. It is often a vector, matrix, or of higher dimensional order referred to as a tensor. For example, this input might be a word (string) or an image, and the total input tensor with time dimension included might be a sentence or a video respectively.

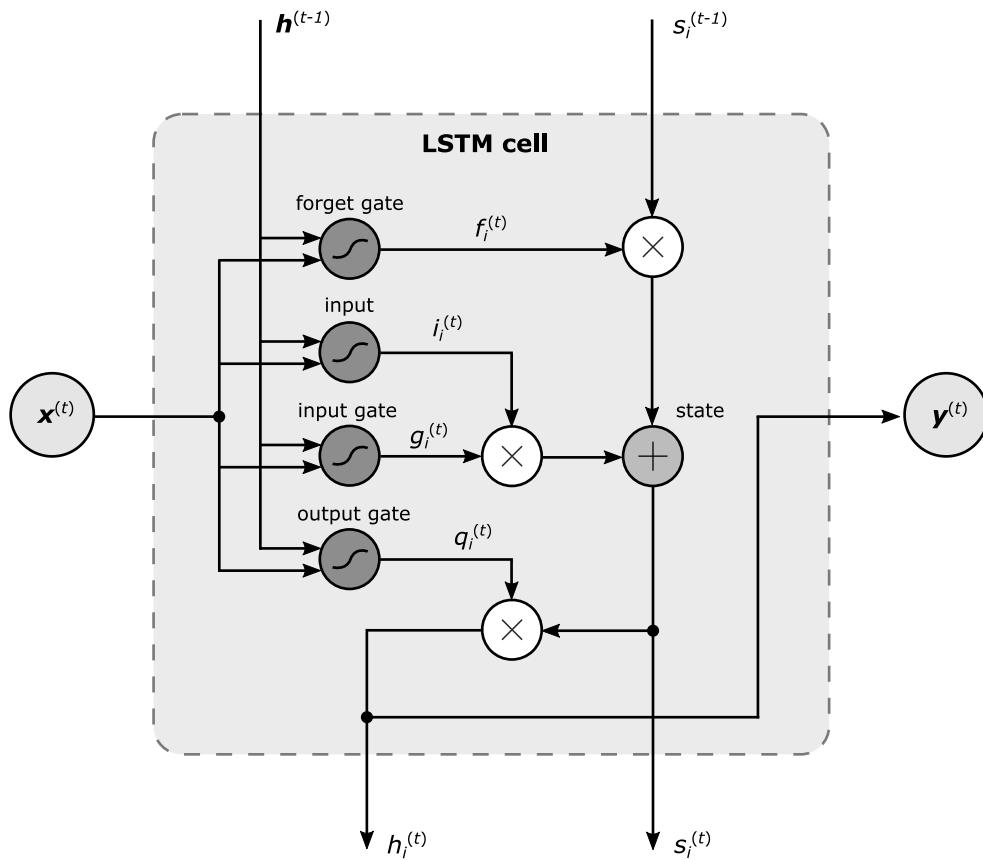


Figure 2.13: Structure of a single LSTM cell.

2.4.3 Long Short-Term Memory

As can be understood from Figure 2.12, it is the hidden layer that is responsible for transforming an input to a certain output. In other words, this is where most of the calculations take place. The units, or cells, of the hidden layer may have many different designs, but one cell that has proven to be increasingly successful [21] is the Long Short-Term Memory (LSTM) cell.

What is special about LSTM, in short, is that it has the ability to both communicate with its neighbouring cells and to discard its memory. In a sense, the cell learns to forget, which entails the possibility of only storing the most relevant data. The structure of a typical LSTM cell is shown in Figure 2.13. In this figure $x^{(t)}$ is the input at time t , which may be a vector, matrix or tensor. For simplicity it is assumed to be a vector here, but the concept of the LSTM cell is the same regardless. $h^{(t-1)}$ is another input that the cell takes, which is the output of all the LSTM cells in the previous time step. $s_i^{(t-1)}$ is a crucial

unit, the state unit, which is responsible for the long memory properties of the cell. This unit communicates across all the LSTM cells and takes part in controlling the forget gate. Moreover, the equations to describe the cell are given below, as formulated by [21], starting with the forget gate

$$f_i^{(t)} = \sigma \left(b_i^f + \sum_j U_{i,j}^f x_j^{(t)} + \sum_i W_{i,j}^f h_i^{(t-1)} \right), \quad (2.17)$$

where σ is an activation function (sigmoid), $\mathbf{h}^{(t-1)}$ is the hidden layer vector for the previous time step and \mathbf{b}^f , \mathbf{U}^f and \mathbf{W}^f are all biases and weights for the forget gate. Keep in mind these are part of the weights that would be determined during the training stage of a neural network model. Furthermore, the input, input gate and output gate are

$$i_i^{(t)} = \sigma \left(b_i + \sum_j U_{i,j} x_j^{(t)} + \sum_i W_{i,j} h_i^{(t-1)} \right), \quad (2.18)$$

$$g_i^{(t)} = \sigma \left(b_i^g + \sum_j U_{i,j}^g x_j^{(t)} + \sum_i W_{i,j}^g h_i^{(t-1)} \right), \quad (2.19)$$

$$q_i^{(t)} = \sigma \left(b_i^o + \sum_j U_{i,j}^o x_j^{(t)} + \sum_i W_{i,j}^o h_i^{(t-1)} \right), \quad (2.20)$$

where \mathbf{b} , \mathbf{b}^g and \mathbf{b}^o are all biases and \mathbf{U} , \mathbf{U}^g , \mathbf{U}^o , \mathbf{W} , \mathbf{W}^g and \mathbf{W}^o are all weights. Analogously to Figure 2.13, the output of the output gate and state are added together to form the final output $h_i^{(t)}$ of a single cell

$$h_i^{(t)} = \tanh(s_i^{(t)}) q_i^{(t)}. \quad (2.21)$$

And in addition, the new state becomes the previous state gated by the forget gate and the current input gated by the input gate

$$s_i^{(t)} = f_i^{(t)} s_i^{(t-1)} + g_i^{(t)} i_i^{(t)}. \quad (2.22)$$

Finally, the outputs of all the LSTM cells $h_i^{(t)}$ are then combined for every cell i and fed as a vector to the cells of the next time step. These outputs are also used to form the final output $\mathbf{y}^{(t)}$, but are often fed to further layers and some sort of activation function first.

Chapter 3

Problem Formulation

In this chapter a brief overview, formulation and modelling of the vital signs problem and fall detection problem will be given. In Section 3.1 a mathematical description of the vital signs detection is given and in Section 3.2 a formulation and model of the fall detection is provided.

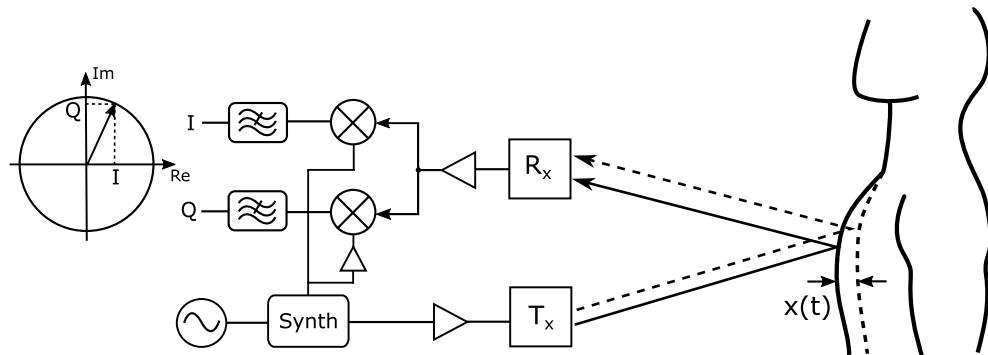


Figure 3.1: Chest displacements $x(t)$ cause phase differences in the EM wave over time that are measured and captured by the IQ demodulator.

3.1 Modelling the Vital Signs Problem

When a human breathes, the chest expands and compresses with every breath. Analogously, this happens for the thorax region and the arteries throughout the body with every heartbeat. From the mm-wave radar's point of view, the thorax region (located at a distance R_0) oscillates with a sub-range resolution motion $x(t)$, causing phase changes measured by the radar. See Figure 3.1. When analyzing this motion, the frequency f_{IF} corresponding to the constant

beat frequency determined by the range to the target, is not of much interest, since the motion is below range resolution. It is primarily the phase difference between consecutive chirps that is useful. Hence focus here will be given to the phase part of the downconverted IF signal. Further assuming the amplitude of this signal to be unity, the I [18] and Q parts can be approximated as

$$I_B(t) = \cos \left[\frac{4\pi R_0}{\lambda} + \frac{4\pi x(t)}{\lambda} + \phi(t) \right], \quad (3.1)$$

$$Q_B(t) = j \cos \left[\frac{4\pi R_0}{\lambda} + \frac{4\pi x(t)}{\lambda} + \phi(t) - \frac{\pi}{2} \right], \quad (3.2)$$

where λ is the center wavelength and $\phi(t)$ is the residual phase. Note that in mid-microwave frequencies and especially mmWave frequencies (≈ 60 GHz) the EM wave primarily scatters off the surface of the thorax region [5], since the waves rapidly attenuate inside tissue at these frequencies. Moreover, it is the motion $x(t)$ that is desired to be detected, as its motion contains both the breathing and heartbeat information. In estimating the heart- and breathing frequency, it is simply the frequencies of two periodic signals that are to be determined

$$x(t) \approx A_B \sin(2\pi f_{B1}t) + A_H \sin(2\pi f_{H1}t), \quad (3.3)$$

in which f_{B1} and f_{H1} are the fundamental breathing- and heart rate frequencies, respectively, while A_B and A_H are their magnitudes. In reality, none of the motions are perfectly sinusoidal; and even if they were they would not be measured as such due to noise in the measurement and superpositioning of the signals. This makes the estimation more troublesome. A first step may be to simulate the motions for a greater understanding. By assuming the breathing and heartbeat movement to originate from the thorax region, at static distance R_0 , it is possible to simulate the superpositioned displacement of both events. For this, a Fourier expansion can be carried out for the baseband signals of Equation (3.1) and Equation (3.2), assuming the signal $x(t)$ to be of the form $x(t) = A \sin \omega t$ yielding [22], [18]

$$\tilde{I}_B(t) = \sum_{n=-\infty}^{\infty} J_n \left(\frac{4\pi A}{\lambda} \right) \cos [n\omega t + \Phi(t)], \quad (3.4)$$

$$\tilde{Q}_B(t) = j \sum_{n=-\infty}^{\infty} J_n \left(\frac{4\pi A}{\lambda} \right) \cos [n\omega t + \Phi(t) - \frac{\pi}{2}], \quad (3.5)$$

where J_n is the n th order Bessel function of the first kind and $\Phi(t) = \frac{4\pi R_0}{\lambda} + \phi(t)$ is a combined residual phase term. Now, these signals are composed only

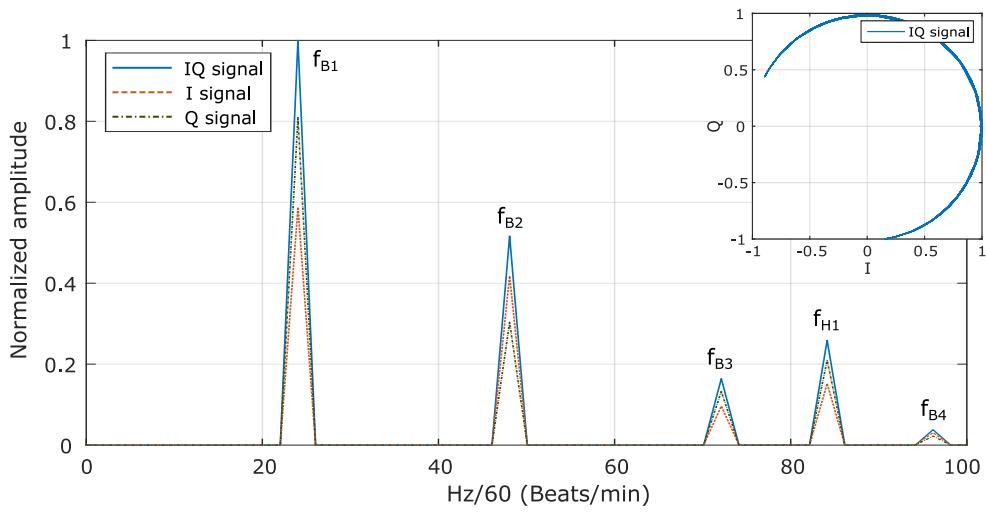


Figure 3.2: Simulated complex IQ signal (top right) and the Fourier transform of its Fourier expansion, illustrating the interference of breathing harmonics.

of multiples of the fundamental frequency, which enables studying their harmonics in frequency domain. As the frequency of interest is $f \approx 60$ GHz, a limitation $A = 4\pi|x(t)|/\lambda \leq \pi$ is held for the simulation as exceeding this value would require dealing with phase unwrapping or similar. Consequently, for the sake of the argument, a modest breathing amplitude of 0.7 mm (1.4 mm peak-to-peak) and frequency of $f_{B1} = 0.4$ Hz (24 breaths/min) is calculated from $n = -20, \dots, 20$ using Equation (3.4) and Equation (3.5). The complex IQ-signal is formed as $S_{IQ} = \tilde{I}_B(t) + \tilde{Q}_B(t)$. Analogously, the signal is superpositioned with a modest heartbeat signal of amplitude 0.12 mm (0.24 mm peak-to-peak) and frequency $f_{H1} = 1.4$ Hz (84 beats/min). For both cases, $\Phi(t) = 0.2\pi$ is assumed. The result is seen in Figure 3.2, showing the complex IQ baseband signal of Equation (3.1) and Equation (3.2) in the top right corner, and the Fourier transform of the Fourier expanded signal of Equation (3.4) and Equation (3.5) in the main window. From the figure, it is evident that detecting the fundamental breathing frequency f_{B1} is straight-forward for simple static cases, but that detecting f_{H1} may be more troublesome due to interference of the breathing harmonics.

3.2 Formulating the Fall Detection Problem

The goal here is to utilize multi-label classification to classify and predict falling motions for human subjects. Multi-label classification presents impor-

tant data about the false alarms and false positives for each class, yielding a greater insight than binary classification would. Moreover, the problem as a whole can be summarized by three parts

1. Acquiring or recording data to be used as both train- and test data.
2. Pre-processing the data to improve arbitrary features of it.
3. The machine learning part; where a neural network is designed to train weights and biases to fit a model to the train data, in order to predict the test data.

Here the recorded information is chosen as range-doppler-time data, which in fact is a set of images where the pixel dimensions are defined by the number of range bins and doppler bins, and the number of frames is determined by the record interval and temporal resolution. As can be understood, this tensor data structure is in fact a video. Furthermore, some digital signal processing to enhance image features aids the machine learning in finding the relevant data. These processes might include normalization, contrast enhancement and noise removal, and is explained in greater detail in Section 5.2.2.

The neural network model will here be split into two parts: A feature extraction network and a time sequence network (refer to Figure 3.3). Firstly, the idea is to design a function f_f and train parameters to it that takes the image input $\mathbf{x}^{(t)}$ to produce a vector of features $\mathbf{E}^{(t)}$

$$\mathbf{E}^{(t)} = f_f(\mathbf{x}^{(t)}). \quad (3.6)$$

Here the function may be a nested function $f_f(\mathbf{x}^{(t)}) = f_1(f_2(\dots(f_N(\mathbf{x}^{(t)}))))$, where each function corresponds to a layer or activation function of the neural network. These features for all times t can further be fed to a time sequence network, containing LSTM cells for instance, in order for the network to learn

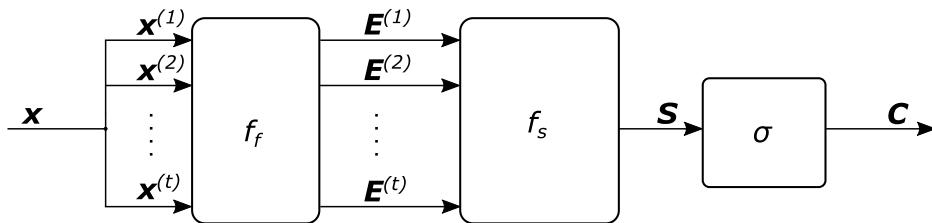


Figure 3.3: Simplified illustration of data flow in the neural networks, to estimate a gesture class from a video input.

the parameters' relation in time. Having f_s denote the function of the time sequence network and \mathbf{E} being the set $\mathbf{E} = \{\mathbf{E}^{(1)}, \dots, \mathbf{E}^{(t)}\}$, this gives

$$\mathbf{S} = f_s(\mathbf{E}), \quad (3.7)$$

where \mathbf{S} is a vector containing time-dependent information of the image features. Yet again, f_s may be a nested function. It is within this block the recurrence takes place, as described in Equation (2.16), or for a LSTM cell by Equation (2.21). Lastly, a softmax activation function σ is used to produce a vector of outputs containing the probabilities of the input data matching each class

$$\mathbf{C} = \sigma(\mathbf{S}), \quad (3.8)$$

in which \mathbf{S} and hence \mathbf{C} are vectors having the same number of elements as the number of classes or labels. An exact mathematical description of the softmax activation function is given in Section 5.2.3. For simplicity the weights and biases are not explicitly stated for the above equations, as there are many thousands of these in the implemented networks. In conclusion, the radar captured tensor video input \mathbf{x} may be fed to a feature extraction network, time sequence network and softmax activation function to estimate the human gesture corresponding to the label $\max(\mathbf{C})$.

Chapter 4

The IWR6843AOPEVM Radar

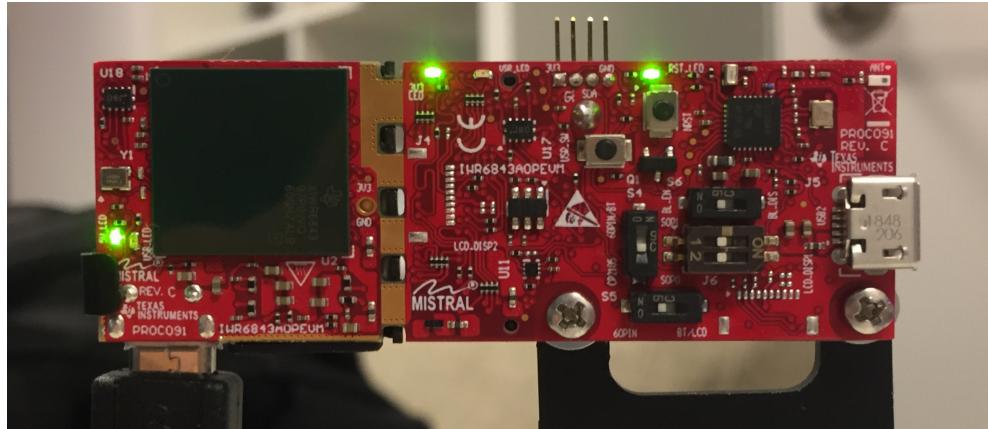


Figure 4.1: Image of the IWR6843AOPEVM radar.

To carry out the project and implement the vital signs and fall detection, the FMCW radar IWR6843AOPEVM from Texas Instruments was used, in stand-alone mode. This is a so called mm-wave radar, whose wavelengths are in the order of millimeters (microwave frequency region). In this chapter it will be briefly described how the radar is designed, how it operates, what configurations it has and how serial communication of the radar is handled.

4.1 Overview

The IWR6843AOPEVM is a (PCB, Antenna on Package) MIMO radar chip with a FMCW transceiver consisting of 4 integrated receivers and 3 transmitters, all being patch antennas with 120° Field of View (FoV). It operates by

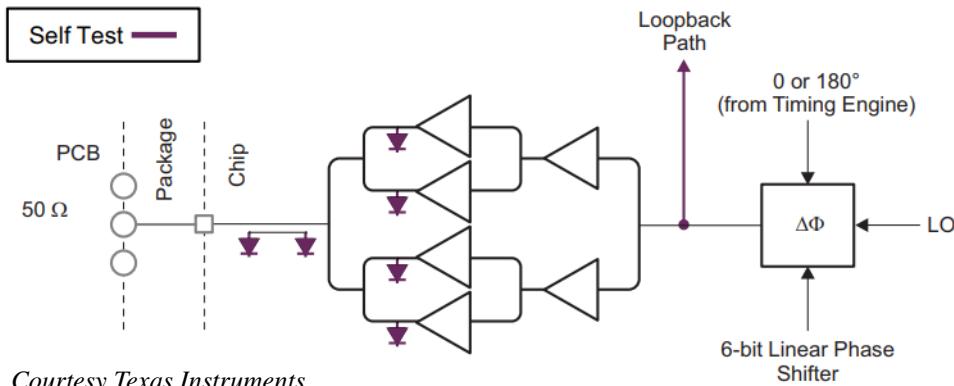


Figure 4.2: Transmit subsystem of the IWR6843AOPEVM board, by [23].

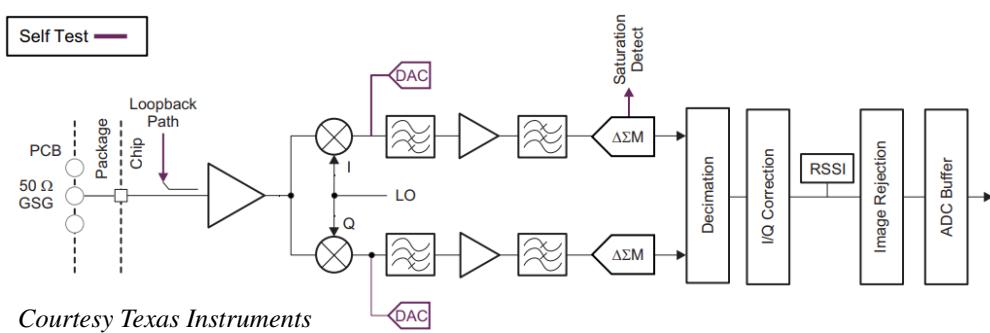


Figure 4.3: Receiver subsystem of the IWR6843AOPEVM board, by [23].

transmitting a sawtooth FM waveform, and it uses Time Division Multiplexing (TDM) or Binary Phase Modulation to obtain orthogonality between transmitted signals. The radar's corresponding virtual antennas are the same as shown in Figure 2.7 of Section 2.2.4, assuming that all Tx and Rx are active. For the LO signal and coherence the chip uses a 40 MHz crystal oscillator, with a phase noise of -92 dBc/Hz at 1 MHz offset. This signal is synthesized to produce a FM chirp that can range from 60 to 64 GHz. For generation of LO signal and structure of clock system, see Figure A.1 in Appendix.

Furthermore, a single transmitter chain of the radar can be seen in Figure 4.2, and the receiver chain is depicted in Figure 4.3. Clearly, the receiver chain uses an LNA to amplify the signal and an IQ demodulator to generate the complex baseband I and Q signals. The signal passes through an ADC and multiple filters, after which the complex digital baseband signal is sent and stored in the ADC buffer on the chip. In addition, the board has a Data Processing Unit (DPU), a hardware accelerator, Micro-Controller Unit (MCU) and a ROM and RAM to process signals, control functions and store data.

In stand-alone mode, the radar connects to a PC through a Universal Asynchronous Receiver Transmitter (UART) to USB interface which has a limited bitrate of 921600 bit/s for transmission and 115200 bit/s for reception. For a complete block diagram of the board, see Figure A.2 in Appendix.

4.2 Flashing Firmware to the Board

Flashing firmware onto the radar board is essential to provide low-level control of its hardware. The flashing contains the basic set of instructions that the radar may be configured with, how the processing is handled on board and what data may be stored and transmitted. Due to its importance, it will be briefly explained how the flashing is carried out.

The radar chip has a mini USB female connection that may connect to a PC through a mini-USB to USB adapter. When plugged in to a PC, two COM-ports should be recognized, where one channel handles data transmission and the other data reception. Whether the COM ports are recognized or not can be tested with a software such as *Tera Term*. When the connection is secured, the chip must be configured for flashing. This is done by toggling the switches on the board to a certain setup. Specifically, the S2 switch should be set ON, and the SOP0 and SOP1 switches set OFF. What this means exactly can be explored in Texas Instrument's user guide of the radar board [24]. Once connected to the COM-ports of the board, with correct switch setup, the flashing of firmware could take place. Texas Instruments has a set of pre-built firmware to the radar in their Software Development Kit (SDK) that can be downloaded along with the evaluation board. To flash arbitrary firmware onto the board, some software such as *Uniflash* is required. In this software, make sure that the COM ports of the radar are found and choose the appropriate .bin file firmware in order to execute the flashing.

When done, the radar should be configured back to functional mode. In stand-alone mode, this is simply achieved by changing the S2 switch to OFF, having the SOP0 and SOP1 switches remain OFF.

4.3 Configuring the Radar Subsystem

With SDK demo firmware flashed onto the board comes an Application Programming Interface (API) that provides a command and response interface which allows configuring and programming the radar subsystem and its parameters. In this configuration numerous properties can be changed, such as

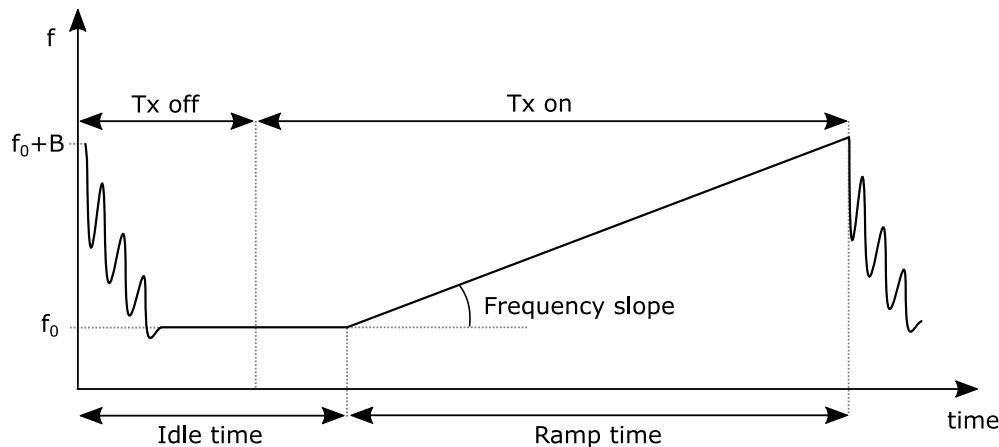


Figure 4.4: Illustration of chirp parameters for a single chirp cycle.

the range resolution, velocity resolution, the number of active transmitters and receivers, the ADC sampling frequency, chirp parameters, the type of data to be processed, and much more. For a full table of configurations the reader is referred to the first table in the SDK User Guide [25].

When configuring the radar, it is important that the chirp parameters, in fact, can be realized. The relation between chirp parameters is most easily illustrated by a graph, which can be seen in Figure 4.4. The ADC sampling rate must always be chosen such that all samples finish within the ramp time, and the total time of a frame T_{PRT} must be chosen larger than the coherent pulse interval with all its contained chirp cycles

$$T_{PRT} > N_{chirp}(T_{idle} + T_{ramp}), \quad (4.1)$$

where N_{chirp} is the total number of chirps in the coherent pulse interval, while T_{idle} and T_{ramp} are the idle time and ramp time. Failing to provide a realizable configuration to the radar will result in an error exception, causing the radar to become unresponsive and requiring reboot. Note that for MIMO-TDM operation the number of chirps increase with the number of Tx, putting a further restraint on the frame time.

Particularly important configuration parameters, in terms of post-processing data, is the type of data that is processed and stored in the radar. With the SDK demo firmware, this may be complex range profiles, point cloud data or range-doppler heat maps to name a few. Furthermore, this is the data that would be transmitted through the UART-USB interface to a PC for post-processing. As will be seen further on, the complex range profile may be used to estimate vital signs, while the range-doppler heat map provides adequate data for the training of neural networks for fall detection.

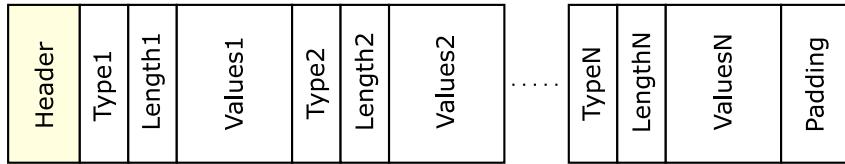


Figure 4.5: Structure of the TLV message.

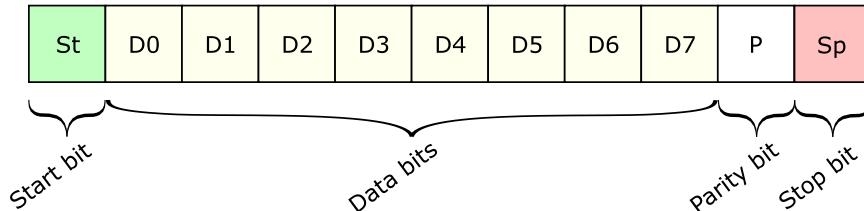


Figure 4.6: The UART frame format.

4.4 Communicating with the Radar

The communication with the radar was carried out between the UART and USB interfaces through serial communication. Firstly, a code was written to send instructions, a configuration file, to the MCU on board. Part of these instructions were the configuration parameters explained in Section 4.3, along with a command ordering the sensor to start. The board expects this message to be sent from the computer, after which data of the radar starts being sent to the computer.

Now, the board sends its stored data to the UART, which is sent as Type Length Value (TLV) format. This format contains a header for the message, and the types and lengths of packages within the message. It is illustrated in Figure 4.5. The type here might be the complex range profile or the range-doppler heat map. Furthermore, once these messages reach the UART for transmission to the PC USB, they obey the UART communication protocol. Each byte (8 bits) of data is sent in conjunction with a start bit, parity bit and stop bit, as presented in Figure 4.6.

When parsing the received 8-bit TLV messages, the header, type and length is parsed to extract the data bits. The header contains a value corresponding to the length of the entire TLV message, and each type is followed by the length of the package, making parsing a simple task. However, the extracted data is 8-bit binary, but the radar supports 16-bit signed integers. Hence pairs of two consecutive bytes must be combined and converted from binary to decimal to reconstruct the value. As this was done for all values, the values of the radar

frame was sent to a buffer. The one-dimensional vector of values in the buffer could then easily be converted to either a range-doppler heatmap matrix or complex range-virtual antenna matrix for further digital processing.

Having an understanding of the communication protocol is essential, as its the limiting factor for transmitting high rates of data. Recall that the UART and thus the radar has a baud rate limitation of 921600 bit/s. It might seem like a lot, but if the radar is to export high resolution range-doppler matrices at a high frame rate, then there is still shortage baud rate. For instance, if the radar is to transmit range-doppler heatmaps with 128 range bins (N_{rb}) and 32 doppler bins (N_{db}) at 9 frames per second (f_{PRF}), this requires a minimum bit rate of

$$R_b = 2f_{PRF}N_{rb}N_{db}N_{bit}, \quad (4.2)$$

where N_{bit} is the number of bits in each UART frame and the factor 2 comes from the 8-bit to 16-bit conversion. Assuming $N_{bit} = 11$ and the values above, this gives $R_b = 811008$ bit/s, which is close to the absolute limit.

Chapter 5

Methods

Here two different methods will be described. Firstly, it will be explained how the IWR6843AOPEVM radar was used to locate a living target and thereafter estimating its breathing- and heart rate in a MIMO approach. Secondly, the fall detection method will be covered. That is, how and what type of data was recorded and what neural networks were used to fit a model to the data.

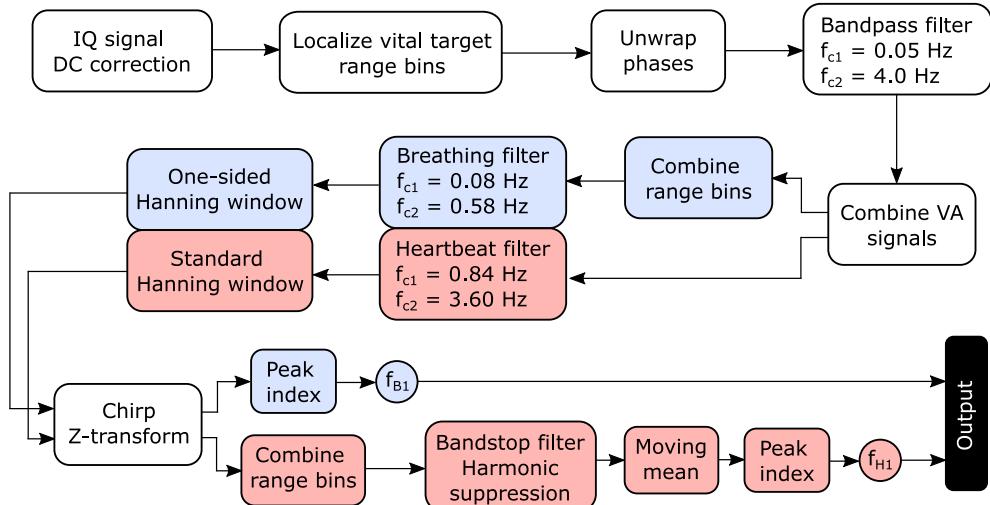


Figure 5.1: Overview of the approach to estimating the vital signs. In this figure f_{B1} and f_{H1} denote the estimated breathing- and heart frequencies, while f_{c1} and f_{c2} denote the cutoff frequencies for the bandpass filters.

5.1 Finding & Estimating the Vital Signs

A flowchart of the procedure to estimating the vital signs can be seen in Figure 5.1. Each one of these steps will be explained in detail. How the IQ signal is corrected is covered in Section 5.1.2, and how the range bins of the vital target is localized is explained in Section 5.1.3. Moreover, the phase unwrapping process is explained in Section 5.1.4. After this, the signals of the virtual antennas are combined to improve SNR, which is explained in Section 5.1.5. Now the signals are separated to a heartbeat signal and breathing signal using bandpass filters, and the signal is further processed. This entire signal processing will be covered in Section 5.1.6. Part of this signal processing is estimating the rates of breaths and heartbeats, which may be achieved with three different methods; the peak finding approach, the FFT approach or the CZT approach. The advantage of the peak finding approach and FFT approach is simplicity, whereas the CZT approach is more accurate. Due to this, the CZT approach will be explained in detail further on, while the other two approaches will only serve as references. Nonetheless, lastly, a simple way of reducing the breathing harmonics will be shown in Section 5.1.7.

5.1.1 Vital Signs Setup

For the vital signs detection, the IWR6843AOPEVM radar was used in standalone mode, flashed with the demo firmware from Texas Instruments. The radar was communicating through a UART serial port to a HP Notebook PC, where all the live signal processing was handled in MATLAB.

Furthermore, due to the limited baud rate of the serial port communication between the PC and radar, and also due to computing power limitations, restrictions had to be put on radar parameters such as range resolution and frame rate. The final configuration is shown in Table 5.1, where it can be seen that 2 Tx and 4 Rx were used, giving 8 virtual antennas. The most important factor here in terms of vital signs detection is the Pulse Repetition Frequency, f_{PRF} , which must be sufficiently high to fulfill two criteria. The first criterion is the Nyquist one, stating that

$$F_s > 2f_x, \quad (5.1)$$

where F_s is the PRF and f_x may be either the heart rate frequency, f_{H1} , or breathing frequency, f_{B1} . However, since $f_{H1} \gg f_{B1}$ it is the heart rate that puts the restriction. Knowing that f_{H1} may be up to 3 Hz [17], it tells that $f_{PRF} > 6$ Hz for the highest heart frequencies. Moreover, since the heartbeat

Parameter	Symbol	Value	Unit
Transmitting antennas	N_{Tx}	2	-
Receiving antennas	N_{Rx}	4	-
Center frequency	f_c	60.4	GHz
Bandwidth	B	834	MHz
Range resolution	R_{RES}	18	cm
Range bins	N_{rb}	64	-
Pulse Repetition Frequency	f_{PRF}	20	1/s

Table 5.1: Radar configuration for the vital signs detection. Note that the PRF here corresponds to the number of frames per second sent from the radar.

waveform is more of an impulse than a pure sine wave, it is safest to be well above this threshold.

The second criterion comes with the unwrapping process due to the fact that any phase $\Phi \in [-\pi, \pi]$, not allowing any consecutive phase jumps greater than π for the process to hold [26]. This is relevant for quick and vigorous breathing motions, putting a restriction $f_{PRF} \gtrsim 13$ Hz for the most extreme cases. This is explained more carefully in Section 5.1.4.

5.1.2 IQ Signal DC Correction

The received data from the radar is a complex range profile, which simply is a Fourier transform across fast time for the I- and Q signals. For a specific range bin r and virtual antenna v the complex value for time sample n can be written as

$$C_{r,v}[n] = I_{r,v}^{AC}[n] + I_{r,v}^{DC} + j(Q_{r,v}^{AC}[n] + Q_{r,v}^{DC}), \quad (5.2)$$

where $I_{r,v}^{DC}$ and $Q_{r,v}^{DC}$ are the static components of the range bin in the Fourier transform, while $I_{r,v}^{AC}[n]$ and $Q_{r,v}^{AC}[n]$ are the oscillating components caused by phase changes due to sub-range resolution displacements. Furthermore, when using arctangent demodulation to find the phase $\Phi[n]$ of the object inside this range bin, it is important that the static part is removed, since the phase computation is highly non-linear [4]

$$\Phi[n] \approx \arctan \left(\frac{\text{Im}\{C_{r,v}[n]\} - Q_{r,v}^{DC,aprx}}{\text{Re}\{C_{r,v}[n]\} - I_{r,v}^{DC,aprx}} \right), \quad (5.3)$$

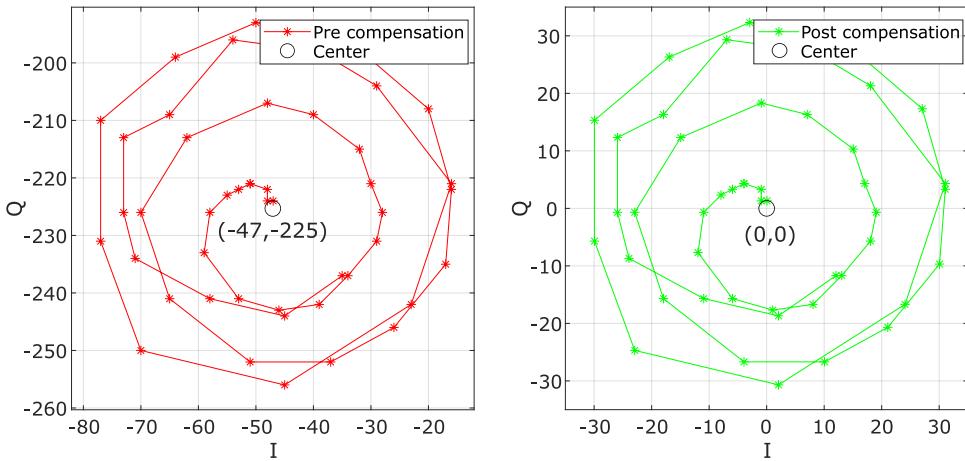


Figure 5.2: IQ-signal measured for 2.5 seconds of a target moving within a range bin. The left graph shows the signal before DC correction and the right shows it after the center has been moved to origo.

in which $Q_{r,v}^{DC,aprx}$ and $I_{r,v}^{DC,aprx}$ are both estimations of the static or DC part of the range bin. One of the fundamental parts of calculating the phase is hence exactly that, to estimate $Q_{r,v}^{DC,aprx}$ and $I_{r,v}^{DC,aprx}$. A straightforward and simple way to do this is to calculate the mean of $C_{r,v}[n]$ over a time that includes multiple periods in the complex plane

$$I_{r,v}^{DC,aprx} = \frac{1}{M} \sum_{n=1}^M \text{Re}\{C_{r,v}[n]\}, \quad (5.4)$$

$$Q_{r,v}^{DC,aprx} = \frac{1}{M} \sum_{n=1}^M \text{Im}\{C_{r,v}[n]\}, \quad (5.5)$$

where M denotes the number of samples to consider. As a window period of 20 seconds and a sampling rate of $f_{PRF} = 20$ were used in this study, giving a total of 400 samples, the most recent half of these were used for the estimation, giving $M = 200$. Equation (5.4) and Equation (5.5) were used to remove the DC part of all the 8 virtual antennas $v \in 1, 2, \dots, 8$ for the 3 located range bins $r \in r_{Loc} - 1, r_{Loc}, r_{Loc} + 1$. The result of the DC compensation for a single range bin and single virtual antenna is illustrated in Figure 5.2. It can be seen that post compensation the data is centered around origo, hence enabling the phase to be easily calculated. In addition, removing the DC part in this sense not only makes the phase calculation possible, but it also helps localizing the vital target. This will be seen in the upcoming section.

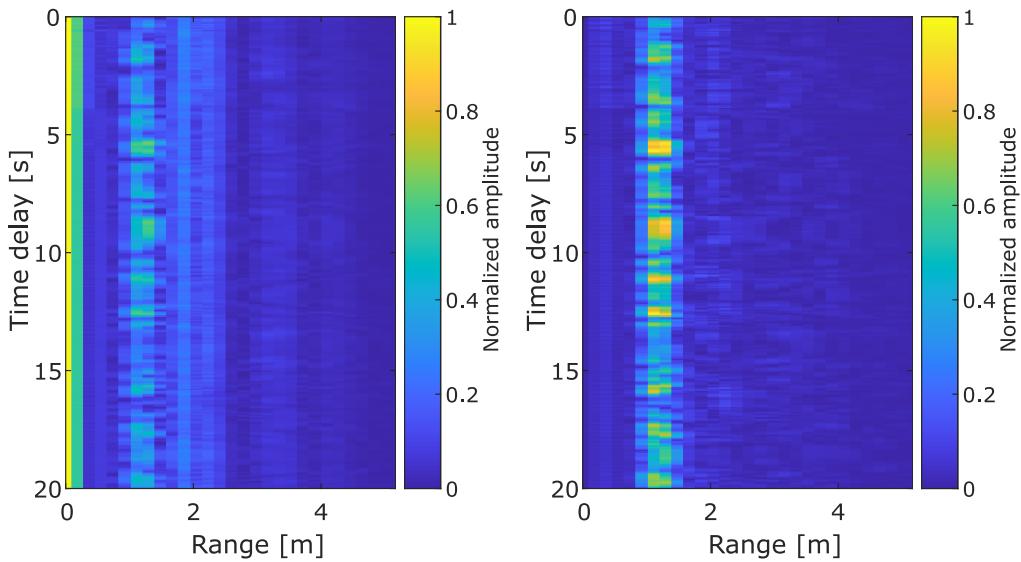


Figure 5.3: Localized range (bins) of a vibrating target, a person, at 1.3 m distance. The left image shows the result before clutter removal and the right image after removal. It is apparent that the dynamic DC removal highlights spatial changes and small displacements.

5.1.3 Vibrating Target Localization

To determine displacements below range resolution for a FMCW radar it is important to keep track of the range bins in which the object is located in. Assuming that the scene is static apart from a single person in the scene, a static clutter removal method is a good start. As a matter of fact, Equation (5.4) and Equation (5.5) described in the previous section may be used to subtract the DC part for all the range bins $r = 1, 2, \dots, N_{rb}$ to dynamically remove static clutter. The result of this procedure can be seen in Figure 5.3, that shows the stored range profiles for a person sitting still and breathing at 1.3 m distance, before and after DC removal. The apparent amplitude changes over time originate primarily from breaths and heartbeats.

With a range resolution of 18 cm, body displacements due to breathing and heartbeats occurred within multiple range bins; approximately 3-4 bins (see Figure 5.3). Because of this, it makes sense to not only find the range bin of maximum amplitude, but instead consider multiple bins. This makes an estimation more robust. Hence one way to estimate the range bins corresponding to the location of the person is to cross-correlate a narrow Gaussian

distribution $G[k]$ with the range profile $r_n[k]$

$$R_n^{xy}[p] = \sum_{k=1}^{N_{rb}} r_n[k]G[k-p], \quad (5.6)$$

where the Gaussian distribution $G[k]$ covers the full extents of a human body; here assumed as 5 range bins. The index of the maximum correlation then gives the estimated range bin of the target. To make the estimation further robust and less sensitive to noise and clutter, the final estimated range bin, r_{Loc} , was determined as the median value of estimated range bins for the 40 most recent frames. The median here is more appropriate than the mean since a mean could offset the estimated range bin if distant clutter is present.

5.1.4 Phase Unwrapping

As stated previously, the phase of the displacements is found using Equation (5.3). However, if the extents of multiple consecutive phases exceed 2π , some procedure is required to keep track of the phase discontinuities to add up the phases. Remember $\Phi \in [-\pi, \pi]$. One procedure that achieves this is called unwrapping, which shifts phase jumps greater than π to their 2π complement. This is most easily explained in pseudo-code:

```
while  $\Phi[n+1] - \Phi[n] \geq \pi$  do
     $\Phi[n] = \Phi[n] + 2\pi$ 
end while
 $\Phi[n+1] = \Phi[n]$ 
```

Note that this phase is the phase at a certain range bin, which only has a single sample per radar frame. Thus this phase must be stored for multiple frames over time to build up the phase history. Furthermore, as seen from the unwrapping process, this puts a restriction on the motion speed that can be tracked. Since any phase jumps greater than 2π would be ambiguous and indistinguishable, the actual consecutive phases from the motion may not exceed this jump. This is analogous to the requirement that any two consecutive displacements must not exceed 0.5λ along Line of Sight (LOS), corresponding to about 2.4 mm for the IWR6843AOPEVM radar (remember the wave travels back and forth). This puts a restriction on the PRF, especially for the breathing, since a breath from full exhale to inhale can expand the chest up to 2 cm along LOS. An approximation of the maximum target LOS motion can be found by [26]

$$d_{max} = \frac{\pi R_{pp} f_x}{f_{PRF}} \quad (5.7)$$

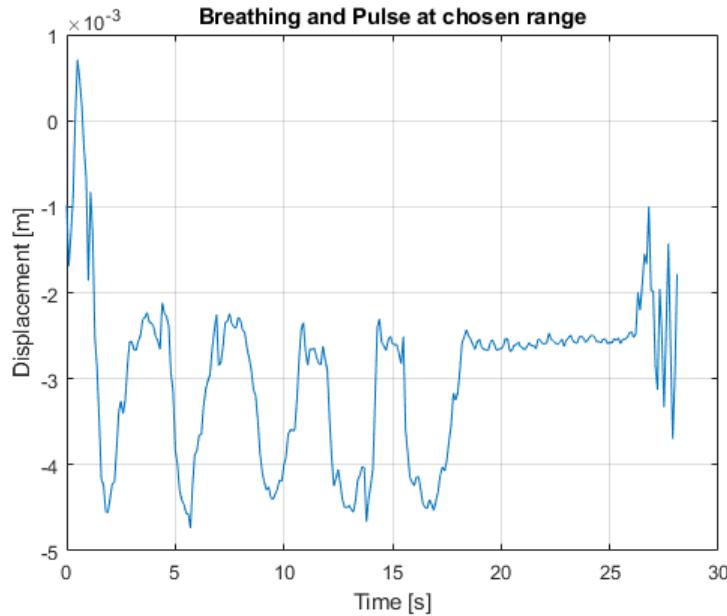


Figure 5.4: Unwrapped phase converted to displacement distance (y-axis) of chest movements for a 28 second interval. The large displacements are caused by breathing and the smaller superimposed displacements by heartbeats. Breath was held at the 18 second mark to isolate the heartbeats.

where R_{pp} is the peak-to-peak sinusoidal motion and f_x is the motion frequency. Assuming $R_{pp} = 2$ cm and $f_x = 0.5$ Hz for an extreme case of breathing, this puts a restriction $f_{PRF} \gtrsim 13$ Hz for the unwrapping process to hold.

It is worth noting that the unwrapped phase is directly proportional to displacement distance as

$$R[n] = \frac{c\Phi[n]}{4\pi f_c}, \quad (5.8)$$

where f_c is the center frequency of the EM wave and c is the light speed in the medium of propagation. Chest displacement of a person, originating from breaths and heartbeats, can be seen in Figure 5.4. This is before any processing or applied filters, but the breaths and heartbeats may still be partially identified.

5.1.5 Improving SNR for MIMO Architecture

As the unwrapped phase (proportional to displacement distance) can be found for one receiver and one range bin, it is also possible to do it for multiple range bins and multiple virtual antennas. Let us denote the time interval unwrapped

phase history for virtual antenna v and located range bin k with time index n as $u_{v,k}[n]$. With goal in mind to increase the SNR by including multiple measurements, it is possible to build a matrix of all the unwrapped phase histories

$$\mathbf{u}_T[n] = \begin{bmatrix} u_{1,1}[n] & u_{1,2}[n] & \cdots & u_{1,K}[n] \\ u_{2,1}[n] & u_{2,2}[n] & \cdots & u_{2,K}[n] \\ \vdots & \vdots & \ddots & \vdots \\ u_{V,1}[n] & u_{V,2}[n] & \cdots & u_{V,K}[n] \end{bmatrix}. \quad (5.9)$$

As the range resolution of the radar was chosen as 18 cm, both breathing displacement and heartbeat displacement covered multiple range bins. Remember that breathing and heartbeat displacements are not located to only the chest, but cause displacements for the shoulders, stomach and arteries all across the body [5]. Furthermore, for the different unwrapped phase histories in Equation (5.9), coherent gain could be achieved by taking the mean of the virtual antenna phase histories and summing up the considered range bins

$$\tilde{u}[n] = \frac{1}{V} \sum_{k=1}^K \sum_{v=1}^V h_F[n] * u_{v,k}[n], \quad (5.10)$$

where V is the number of virtual antennas and K is the number of located and considered range bins. The sign $*$ indicates convolution and $h_F[n]$ is the impulse response of a 4th order Butterworth filter with cutoff frequencies at 0.05 Hz and 4.0 Hz. Filtering prior to summation aids in preserving the AC features of the phase history, reducing unwanted DC components and noise. Ultimately, this results in an increased SNR of $10 \log(V)$ and potentially $10 \log(V \cdot K)$ depending on how many range bins that are affected by the motion. It is worth noting that periodic clutter with same frequency features as the measured breathing- and heart displacement can occur offset from the vibrating target range bin, which further emphasizes improved SNR using Equation (5.10).

It was assumed that breathing could be treated as a bulk motion and hence that the direction of motion would be the same for all range bins covering the breathing. Because of this, Equation (5.10) could be used with $K > 1$ yielding constructive interference. Specifically $K = 3$ was used. For the heartbeats, there is a temporal signal offset in space between chest displacements and artery displacements, and hence a limitation $K = 1$ was set to avoid eventual destructive interference. The signals could have been cross-correlated and time shifted to ensure constructive interference, but it was not deemed as necessary for this study.

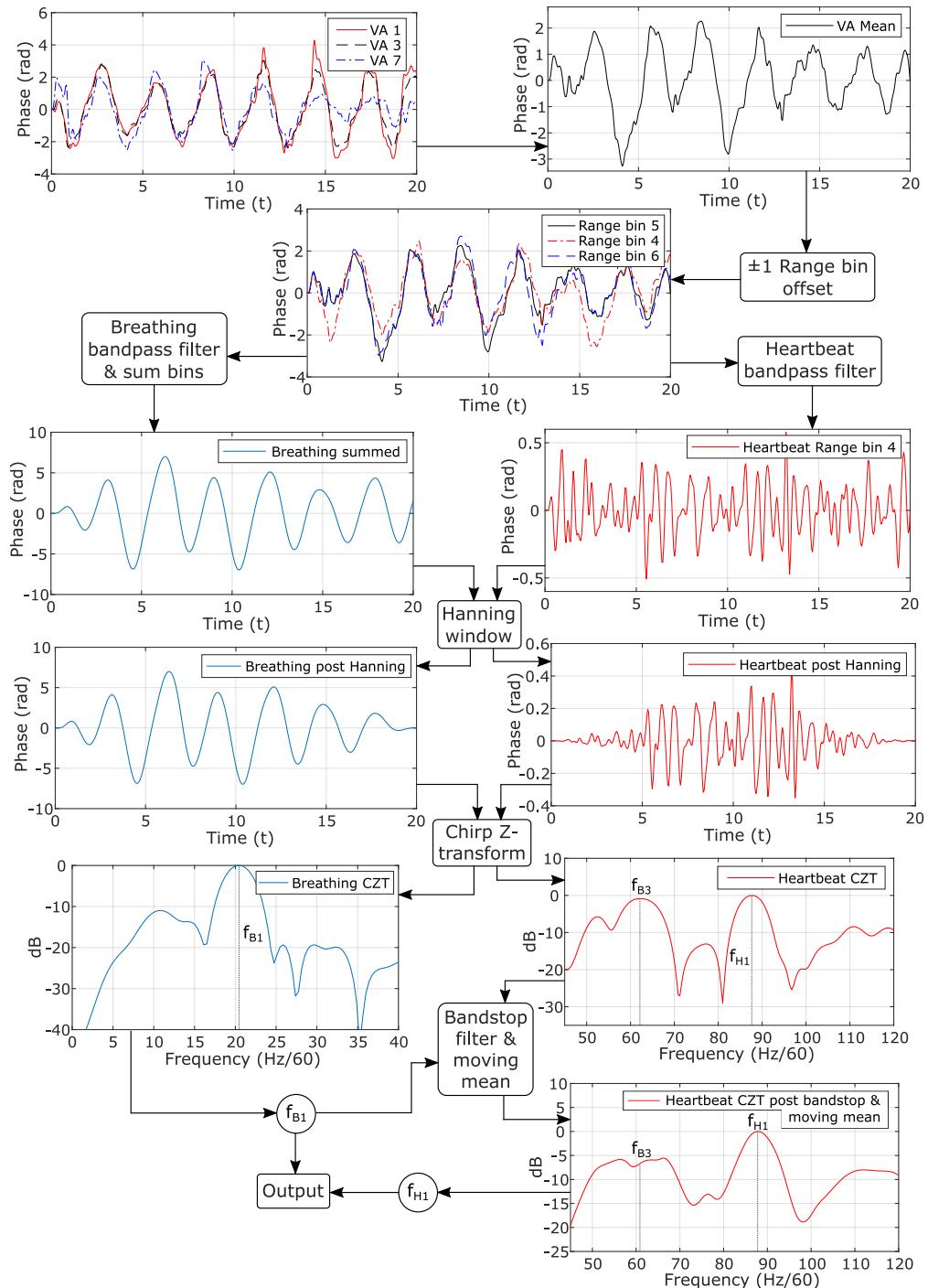


Figure 5.5: Flowchart for the signal processing. Phase histories of 8 virtual antennas and 3 range bins are combined and processed to estimate the fundamental breathing frequency f_{B1} and heartbeat frequency f_{H1} .

5.1.6 Vital Signs Processing Chain

The first step in obtaining frequency estimates for the breaths and heartbeats is to split the superimposed phase history signals based on their different spectral features. This was achieved using two 4th order Butterworth bandpass filters, the breathing bandpass filter having cutoff frequencies at 0.08 Hz and 0.58 Hz, and the heartbeat bandpass filter with cutoff frequencies at 0.84 Hz and 3.60 Hz. Having h_{BR} and h_{HR} denote the impulse responses of the two filters, respectively, also $\tilde{u}[n]$ and $\tilde{u}_k[n]$ denote the phase history means with summation of 3 and 1 range bins, analogously to Equation (5.10), the filtered signals become

$$\begin{aligned} u^{BR}[n] &= h_{BR}[n] * \tilde{u}[n], \\ \mathbf{u}_k^{HR}[n] &= h_{HR}[n] * \tilde{\mathbf{u}}_k[n]. \end{aligned} \quad (5.11)$$

Here $\mathbf{u}_k^{HR}[n]$ is still a vector containing the phase histories of the three located range bins in indices $k = r_{Loc} - 1, r_{Loc}, r_{Loc} + 1$.

Before performing a Fourier transform on the filtered phase histories, it is advisable to utilize a Hanning window or similar to reduce spectral leakage [27]. When applying Fourier transform on a signal that is non-periodic, transients originating from the start and end of the signal cause spectral leakage. This spectral leakage can be greatly reduced simply by ensuring the signal to be periodic, and a simple way to do this is by taking the element-wise product (Hadamard product) of the signal with a Hanning window. The Hanning window is essentially a period of an offset cosine

$$H_1[n] = \frac{1}{2} \left(1 - \cos \frac{2\pi n}{N} \right), \quad (5.12)$$

where $n = 0, 1, \dots, N$ and N is the total number of samples of the signal that the Hanning window should be applied to. Another trick to reduce spectral leakage while preserving more of the signal is that the phase histories $u_{v,k}[n]$ of Equation (5.9) may be phase shifted to put the most recent phase at zero, by subtracting the most recent phase $\Phi[n]$ from $u_{v,k}[n]$. In this sense, only a one-sided Hanning window was required, which was used for the breathing signal to preserve the few breathing periods. The one-sided Hanning window will be denoted $H_2[n] = 1, n < N/2, H_2[n] = H_1[n], n \geq N/2$.

Moreover, there is an issue with the standard FFT approach for estimating breathing rate and heart rate. The frequency resolution of the transform is given as $f_{RES} = F_s/N$, where $F_s = f_{PRF}$ is the sampling frequency and N the total number of samples. But since N is directly proportional to the

time observation window T_{win} and F_s , f_{RES} is only improved with extended observation time $f_{RES} = 1/T_{win}$. This is quickly realized as impractical, as a 60 second observation time would be required for a frequency resolution corresponding to 1 breath/min or 1 beat/min.

This issue can be solved utilizing the *Chirp Z-Transform* (CZT), which is a generalization of the Fourier transform that can be used to increase the spectral resolution while preserving the number of samples [28]. The frequency resolution is further important since breathing harmonics and heartbeat harmonics may be in close proximity in the frequency spectrum [17]. Moreover, since the frequency of breathing may be as low as 0.1 Hz, the frequency resolution of a standard Fourier transform may be insufficient to resolve relevant breathing rates. The transform is formulated as [29]

$$X_m = \sum_{n=0}^{N-1} x_n z_m^{-n}, \quad (5.13)$$

where x_n may be the time-domain discrete phase history corresponding to the body displacements and z_m is a contour in the complex plane. By choosing $z_m = e^{j2\pi m f_2 / (NF_s)}$, with $f_2 = 2.5$ Hz being the upper frequency limit of the new spectrum, this improves the frequency resolution of a 20 second observation window from $f_{RES} = 3$ /min to $f_{RES} = 0.375$ /min, which was deemed sufficient for the application. To conclude, the two signals in frequency domain become

$$\begin{aligned} U^{BR}[m] &= \sum_{n=0}^{N-1} H_2[n] \odot u^{BR}[n] e^{-\frac{j2\pi f_2}{NF_s} mn}, \\ \mathbf{U}_k^{HR}[m] &= \sum_{n=0}^{N-1} H_1[n] \odot \mathbf{u}_k^{HR}[n] e^{-\frac{j2\pi f_2}{NF_s} mn}, \end{aligned} \quad (5.14)$$

where the operator \odot stands for element-wise multiplication. While now in frequency domain, the mean of the heartbeat spectra was taken as $U^{HR}[m] = \frac{1}{K} \sum_{k=r_{Loc}-1}^{r_{Loc}+1} \mathbf{U}_k^{HR}[m]$.

As the frequency spectra are obtained, the fundamental breathing harmonic frequency f_{B1} can be identified in a trivial manner by finding the frequency corresponding to the maximum in the spectrum. This frequency can further be used to reinforce the estimation of the fundamental heartbeat harmonic frequency f_{H1} , which is achieved using a set of bandstop filters to reduce breathing harmonics. This is explained in more depth in Section 5.1.7. Furthermore, after the bandstop filters, a 0.08 Hz (5 /min) width moving mean was performed in the frequency spectrum to even the effect of the bandstop

filter and to reduce sensitivity to occasional sharp peaks originating from different harmonics. This is also realistic for the heartbeat frequency estimation, as the heart rarely holds a constant frequency, but often changes multiple BPM during the course of 20 seconds (T_{win}). A flowchart of the entire processing chain and the corresponding signals can be seen in Figure 5.5. The reader may verify that the processing chain manages to determine both breathing rate and heart rate from the unwrapped phases of 8 virtual antennas and 3 range bins located to the human target. The heart rate was compared to the readings of a pulse oximeter that stated 87 beats/min, which is very similar to the identified heart rate f_{H1} of Figure 5.5. Note that the estimated rates were further averaged over time to present the final estimations.

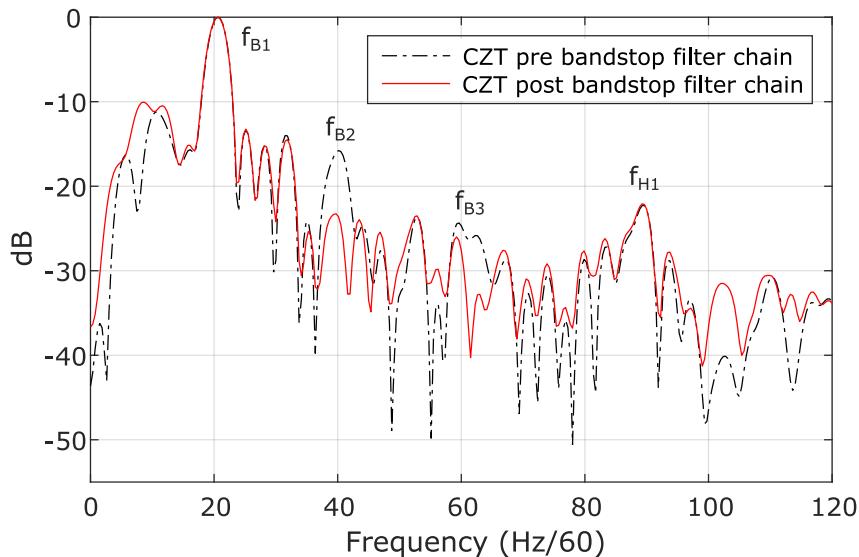


Figure 5.6: Chirp Z-transform for a 20 second recording of unwrapped phases, with and without the bandstop filter chain. It can be seen that both the 2nd breathing harmonic f_{B2} and 3rd breathing harmonic f_{B3} are reduced, making the identification of the heart rate f_{H1} straightforward.

5.1.7 Suppressing Breathing Harmonics

To suppress the breathing harmonics, a simple bandstop filter approach was used. Furthermore, the number of suppressed harmonics were limited to the 2nd and 3rd harmonics, as higher order breathing harmonics tend to be less of a problem due to lower amplitudes [17], and also because having further

bandstop filters in the higher frequency spectrum increases the risk of filtering the actual heartbeat frequency. Nonetheless, the 2nd breathing harmonic was assumed to be located exactly at $2f_{B1}$ frequency and the 3rd harmonic exactly at $3f_{B1}$ frequency. With this in mind, the heartbeat phase history signals $\mathbf{u}_k^{HR}[n]$ were filtered with two bandstop filters for the suppression. Both filters were first order Butterworth bandstop filters, with the first filter having cutoff frequencies at $2f_{B1} - 0.06$ Hz and $2f_{B1} + 0.06$ Hz and the second having cutoff frequencies at $3f_{B1} - 0.04$ Hz and $3f_{B1} + 0.04$ Hz. The frequency spectrum of an unwrapped phase history signal pre and post filtering is shown in Figure 5.6.

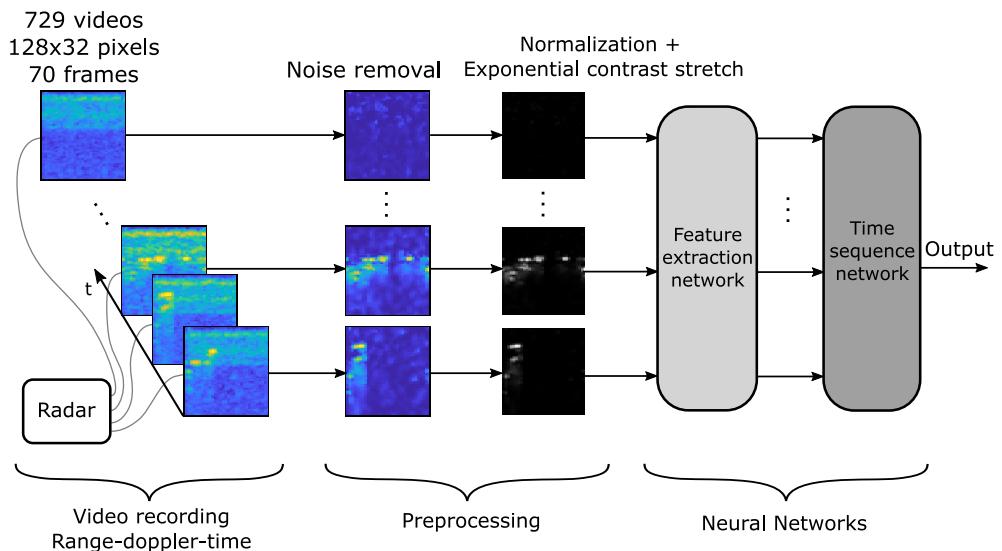


Figure 5.7: Procedure for implementing the fall detection. Range-doppler-time data was recorded with the radar, pre-processed and fed to a set of neural networks to classify gestures.

5.2 Implementing Fall Detection

For the fall detection, the goal was to utilize a machine learning approach with recurrent neural networks to classify different gestures. However, unlike e.g. image facial recognition, there is yet not much publicly available data for radar gesture recognition. Hence data of this sort had to be recorded from scratch. For this study, range-doppler-time data was chosen to be recorded for the algorithm. The radar parameters associated with the recordings can be seen in Table 5.2, from where it is evident that the velocity-, spatial- and

especially temporal resolution are all trade-offs. Due to the limited baud rate of the UART communication protocol, more data could not be transferred per time using the IWR6843AOPEVM radar in standalone mode. Consequently, the resolutions shown in Table 5.2 were assumed to be the best possible trade-off with the available hardware.

An overview of the method in realizing fall detection is illustrated in Figure 5.7. The first step was to record range-doppler-time data, where 729 videos of 70 frames were recorded of different human gestures such as falls. This is explained in detail in Section 5.2.1. Prior to feeding the tensors to the neural networks, image pre-processing was carried out. This included methods such as normalization, noise removal and histogram equalization, where the goal was to provide clear and simple data to the RNNs. This is covered in Section 5.2.2. Furthermore, after processing the tensors were fed to a feature extraction network and time sequence network in order to train, adjust weights of the networks and ultimately classify the different gestures. The design of the neural networks is demonstrated in Section 5.2.3 and the training and classification stage is explained in Section 5.2.4.

Parameter	Symbol	Value	Unit
Center Frequency	f_c	61	GHz
Bandwidth	B	1.9	GHz
Range Resolution	R_{RES}	8	cm
Maximum Range	R_{max}	10.2	m
Velocity Resolution	v_{RES}	0.11	m/s
Maximum Velocity	v_{max}	3.5	m/s
Pulse Repetition Frequency	f_{PRF}	9	1/s

Table 5.2: Radar configuration for the fall detection recordings.

5.2.1 Recording Data

Range-doppler data of a radar is often visualized as an image. Thus the range-doppler-time data can be generalized as a sequence of images; referred to as a video. Furthermore, when capturing the video data of the gestures, static clutter removal was used to minimize the impact of the spatial environment for the data capture. A total of eight different gestures were recorded. This is necessary for the machine learning as both non-fall data and fall data must be

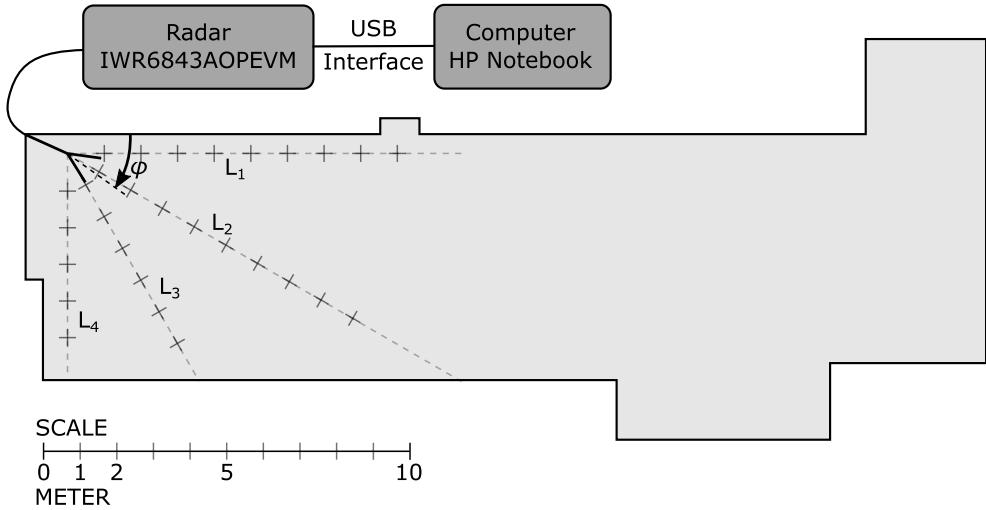


Figure 5.8: Floor plan where the radar was installed and gestures were recorded. The radar was mounted in a ceiling corner of the room with azimuth angle $\varphi \approx 40^\circ$. Along the lines L_1 , L_2 , L_3 and L_4 is where the measurements were taken (at the cross markings), and the angle of the lines were 0° , 30° , 60° and 90° respectively compared to the walls of the room.

represented in order for the algorithm to know what a fall is. The considered gestures were: Falling, bending, jumping, kneeling, lying, sitting, walking and noise. When recording the gestures, all the gestures were performed in four different starting poses; facing the radar, facing the right side to the radar, facing the left side to the radar and facing the back to the radar.

The first recording location is documented by its floor plan in Figure 5.8 and an image of the room seen in Figure 5.9. From Figure 5.9 it can be seen that the room had been stripped from most furniture to ensure people to be within LOS of the radar. Both the radar and the computer (above ceiling) were installed and located within the green circle in the image. When recording this kind of data, it is important to capture a complete and representative data set. Consequently, recordings for the FMCW radar must be performed at different ranges and should be performed at different angles to account for diversity in radiation patterns of the patch antennas of the IWR6843AOPEVM radar. As can be seen from Figure 5.8, this was taken into account as the recordings were carried out at the cross markings of the figure.

A further measure to improve the representation of the data set is to use multiple people for the gesture recordings. This was also done, as a total of three different people were recorded performing the gestures. However, all were male in the age 20-30 years old with a height of 170-185 cm. Capturing

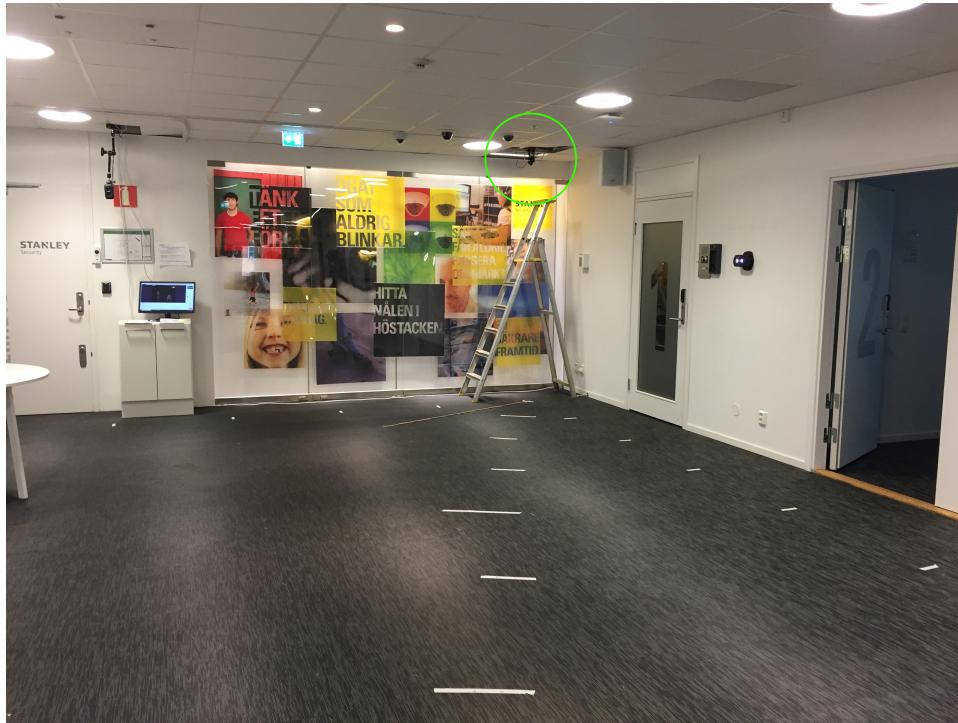


Figure 5.9: First room where the radar was set up and recordings captured. The radar is located within the green circle.

Motion	Nr recordings	Forward/Backward/Sides	Horizontal Range [m]
Fall	141	Yes	1.0-9.0, 1.0-7.0
Bend	77	Yes	1.0-9.0, 1.0-7.0
Jump	70	Yes	1.0-9.0, 1.0-7.0
Kneel	81	Yes	1.0-9.0, 1.0-7.0
Lie	71	Yes	1.0-9.0, 1.0-7.0
Sit	88	Yes	1.0-9.0, 1.0-7.0
Walk	87	Yes	1.0-9.0, 1.0-7.0
Noise	114	-	-

Table 5.3: All radar recorded data. An even distribution of the motions were recorded for 1) facing the radar 2) facing back to radar 3) facing left side to radar and 4) facing right side to the radar. In the first recording location, all motions were recorded in the horizontal ranges 1.0-9.0 m and in the second location the ranges were 1.0-7.0 m.

the data at different locations improves the representation too, as despite utilizing static clutter removal, there may still be dynamic clutter originating from scatterings of the static scene. Hence a second location was used to record further data, which resembles a home care environment. See Figure A.3 of Appendix.

A table of all the 729 recorded gestures can be seen in Table 5.3, where half of the recordings were captured from the room of Figure 5.9, and rest of the half from the smaller 3.1x7.5 meter home environment room of Figure A.3.

5.2.2 Preprocessing

Preprocessing to extract the most relevant data of the videos was carried out in multiple steps; noise and static radar return removal, image normalization and exponential contrast stretching.

The noise and static return removal was achieved by ensuring the first five frames, $N = 5$, of the video to be used as noise reference. The average signal was taken from these frames and then subtracted from all images of the video

$$T_{k,l,t}^a = T_{k,l,t} - \frac{1}{N} \sum_{t=1}^N T_{k,l,t}, \quad (5.15)$$

where $T_{k,l,t}$ are indices of the recorded video tensor with range index k , doppler index l and time index t . The operation was performed for all indices in the

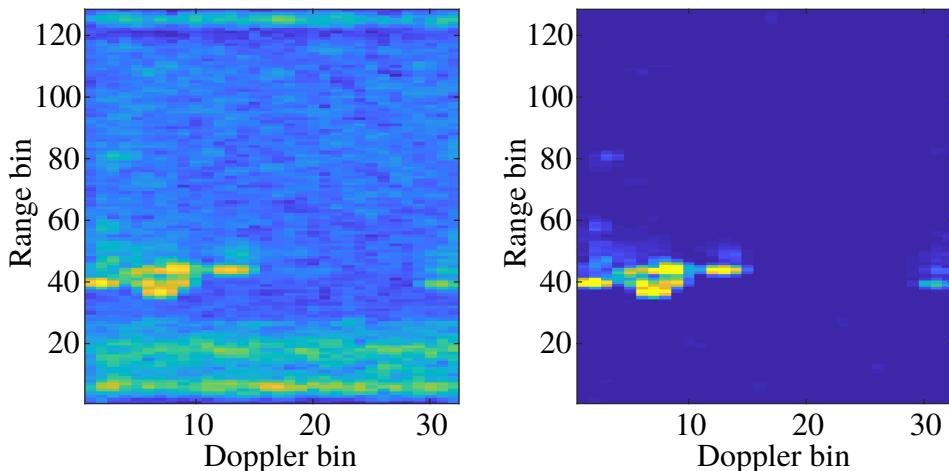


Figure 5.10: Range-doppler plot for a single frame of a fall recording, prior to processing (left) and post processing (right).

video, for $t = 1, 2, \dots, 70$, $k = 1, 2, \dots, 128$ and $l = 1, 2, \dots, 32$. Noise was further reduced by filtering the images with an adaptive low-pass wiener filter, with a kernel size for the convolution of 5×2 , with 5 being the pixel filter width for the range bins and 2 being the pixel filter width for the doppler bins.

Note that the tensor $T_{k,l,t}$ is a set of images, hence every value of the tensor is a pixel value, here ranging from 0 to the max value of a 16 bit signed integer, $2^{15} - 1$ (32767). Furthermore, to improve the clarity of the images a contrast stretch can be performed. However, in this type of data the lowest pixel values are of little interest, as this does not correspond to a strong radar return and is often remainders of noise. Consequently, an exponential contrast stretch was carried out to highlight the strongest radar returns. Normalization and exponential contrast stretch to rebuild new images was carried out as

$$T_{k,l,t}^{out} = \left(\frac{T_{k,l,t}^a - T_{k,l,t}^{min}}{T_{k,l,t}^{max} - T_{k,l,t}^{min}} \right)^\alpha. \quad (5.16)$$

Here $T_{k,l,t}^{min}$ and $T_{k,l,t}^{max}$ are the lowest and highest pixel values of the old image, chosen as 0 and 800, and α is a scaling factor which was chosen as $\alpha = 2$. Note that Equation (5.16) is *not* limited for $T_{k,l,t}^{min} \leq T_{k,l,t}^a \leq T_{k,l,t}^{max}$, because it

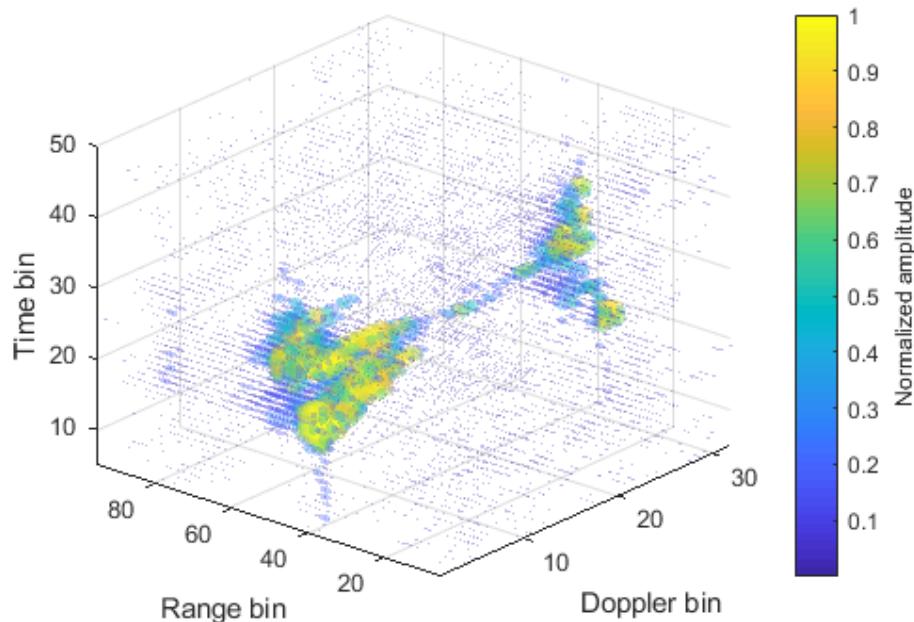


Figure 5.11: Range-doppler-time scatter plot after processing of one of the recorded falls.

is not desirable to remove values $T_{k,l,t}^a > T_{k,l,t}^{max}$ as these correspond to strong radar returns. These values were simply floored at the pixel value $T_{k,l,t}^{max}$, just as the pixel values $T_{k,l,t}^a < T_{k,l,t}^{min}$ were floored at value $T_{k,l,t}^{min}$.

A single range-doppler image pre and post processing can be seen in Figure 5.10, from where it is evident that noise has been reduced and the relevant data has been preserved. Figure 5.11 shows the entire processed range-doppler-time data of a fall recording represented as a scatter plot. This is the tensor that further will be fed into the neural networks. It can be seen that a sudden event takes place around the 20th frame and 40th range bin, which corresponds to the radar return of a person falling. Figure 5.12 shows the same data, but with one dimension stripped to present the traditional doppler-time image and range-time image.

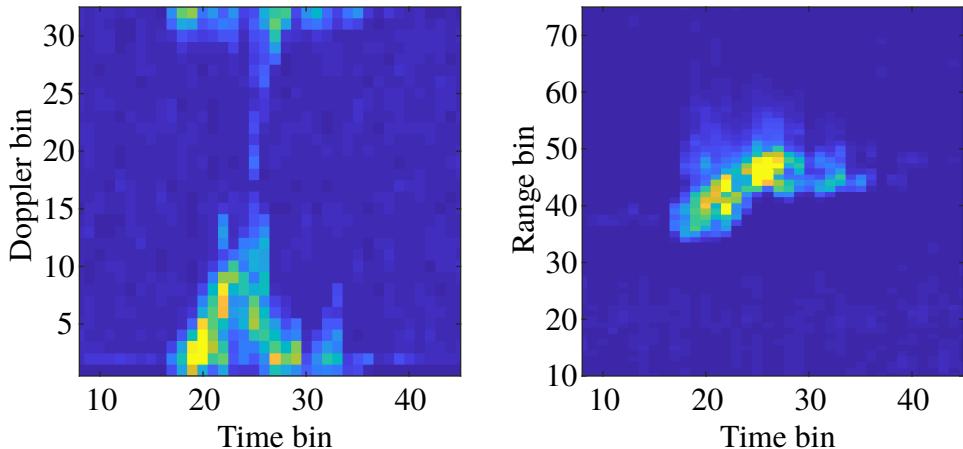


Figure 5.12: Doppler-time plot (left) and range-time plot (right) of the same recorded fall as in Figure 5.11. Note that static clutter removal was used.

5.2.3 Video Classification RNN Model

The Recurrent Neural Network (RNN) model was designed in Python, with Keras and Tensorflow 2.3.1. Furthermore, the model used two different networks with two distinct purposes. The first neural network had the purpose to extract features of the recorded doppler-range images, and the second neural network was designed to classify a gesture to the time sequence of the image features.

The final design of the image feature extraction network can be seen in Figure 5.13. The input x_t corresponds to a single frame; a range-doppler image.

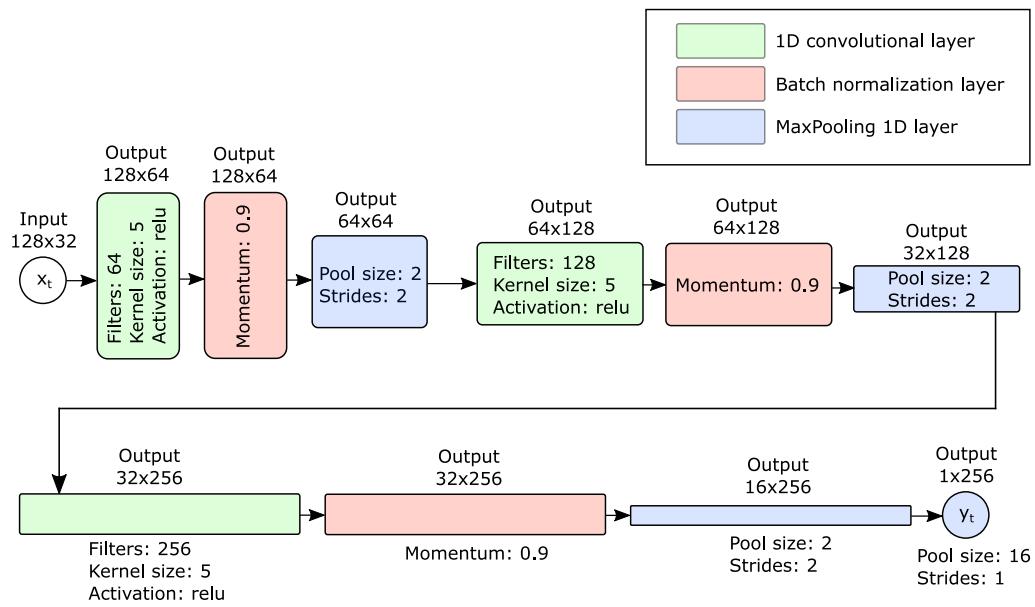


Figure 5.13: Feature extraction neural network for the range-doppler images. A combination of convolutional layers, batch normalization layers and max-pooling layers were utilized to reduce the dimension of the input and produce 256 features per image.

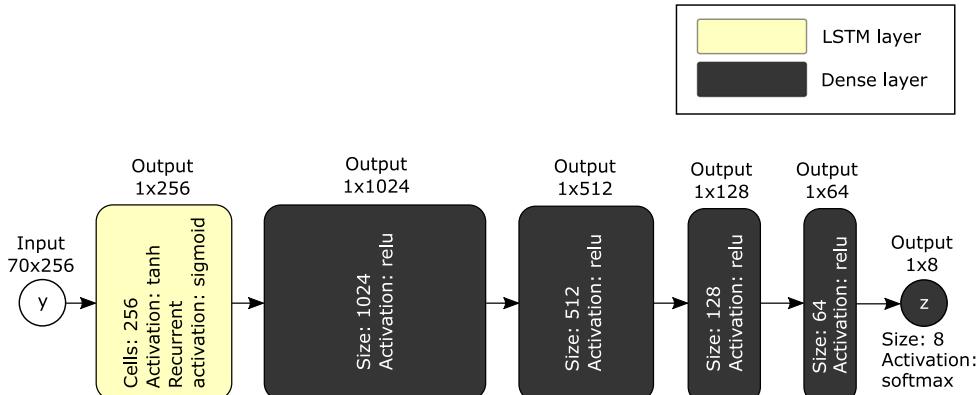


Figure 5.14: Time sequence neural network to produce output classes, corresponding to gestures, from the image features of all the video frames. Note that the output has shape 1×8 , containing the probabilities of the gesture class relating to the input data.

The input is analogous to the result of Equation (5.16) for $k = 1, 2, \dots, 128$ and $l = 1, 2, \dots, 32$. As can be seen, convolutional layers, batch normalization layers and max-pooling layers were used to produce the image fea-

tures. The purpose of the convolutional layers were to identify simple patterns within the range-doppler images. Meanwhile, the relu activation function $f_{relu}(x) = \max(0, x)$ ensured the output of the layer to be positive. Note that padding was used for the convolution layers to preserve the dimension of the input. Moreover, the batch normalization layer was used to normalize the data in each batch of videos to minimize the internal covariate shift problem, making the network train and converge faster [30]. Furthermore, the max-pooling layers were used to down-sample the input and extract the maximum input values, here corresponding to the maxima of the normalized convolutions.

Now as features exist for the video's images, it is time to feed the features to a time sequence network. The designed time sequence network is illustrated in Figure 5.14. Here, the input y corresponds to the output of the previous neural network of Figure 5.13, but for all time frames of the video $t = 1, 2, \dots, 70$. As seen from the image, a LSTM layer and multiple dense layers are used to obtain the final gesture class output. The LSTM layer has the objective to learn the relevant image features across time, and the dense layer has the function to reduce the dimension of its input based on an activation function, here f_{relu} , for all layers except the last. The last dense layer is where probabilities are assigned to each gesture class based on the input, using a softmax activation function

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{n=1}^N e^{z_n}}, \quad i = 1, 2, \dots, N, \quad (5.17)$$

where N is the number of gesture classes, here taking the value $N = 8$ for the classes: Fall, bend, jump, kneel, lay, sit, walk and noise.

5.2.4 Training and Classification

To train the model, Tensorflow 2.3.1 Graphics Processing Unit (GPU) environment was used along with NVIDIA's CUDA 10.1 and cuDNN 7.6 packages. Computations were performed on the NVIDIA graphics card GeForce GTX 1060.

The recorded data was split into two parts; a training set and a test set. The training set was chosen as a random 80% of the recorded and processed gesture videos, and the test (evaluation) files were the remaining 20%. Furthermore, a model was fit through regression, in order to train and optimize weights of the neural networks, to the training set using the "Adam" optimizer of Keras with a learning rate of 0.0018. The batch size was chosen as 32 and the number of epochs was chosen to be 25. Once the model was fit to the data, it was used to predict the classes of the test data for results.

Chapter 6

Results

Various results will be presented here. Firstly, in Section 6.1, the vital signs results will be shown. It will be shown how well the IWR6843AOPEVM radar performs with the explained methods to estimate both breathing rate and heart rate, with different estimation approaches and at different ranges. In Section 6.2, the multi-label fall detection results will be presented in the form of a confusion matrix, a ROC plot, F1-scores and accuracy and loss plots.

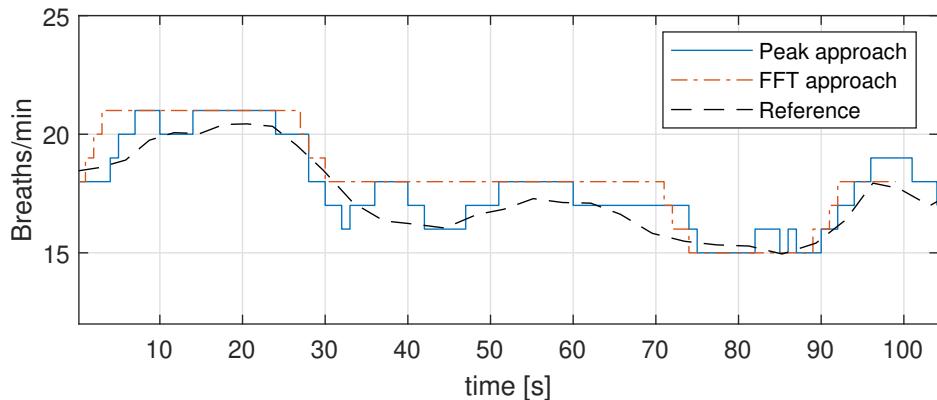


Figure 6.1: Human breathing rate measured by the radar at 4.5 m distance. The black dashed graph is the breathing reference and the two other graphs are different estimations of breathing rate.

6.1 Vital Signs Results

As no chest strap could be used as reference for breathing, a simple clocking approach was used instead. The breaths were clocked at the instant of every

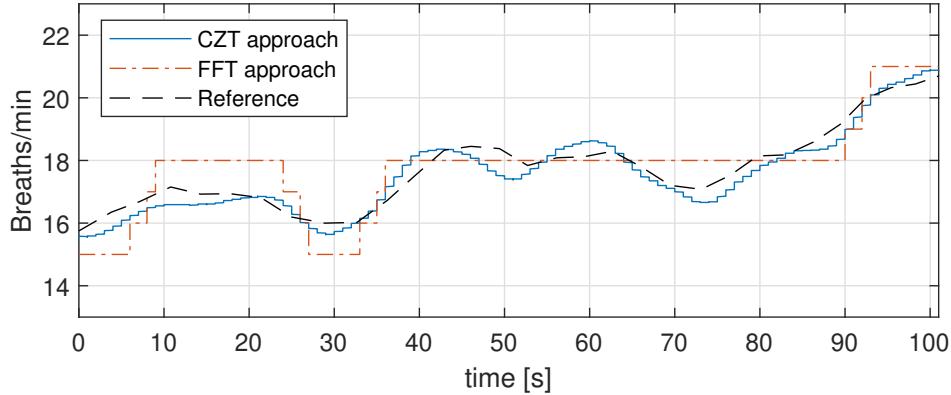


Figure 6.2: Human breathing rate measured by the radar at 2.3 m distance. The reference is illustrated as the black dashed graph, while the blue and red graphs correspond to FFT- and CZT breathing rate estimations.

exhale, and the momentaneous breathing rate reference was determined as the inverse of the average period of the three closest breaths to the current time.

A breathing measurement performed at 4.5 m distance can be seen in Figure 6.1. Here the FFT approach was carried out analogously to the CZT approach in Section 5.1.6, with the chirp Z-transform replaced by the Fourier transform. In contrast, the peak approach was carried out by taking the autocorrelation of either the breathing- or heartbeat signal, and finding the time between its first two peaks. The inverse of this time yields the estimated frequency. Note that in Figure 6.1, the FFT- and peak finding approach estimates have been shifted in time by 10 s and 6 s, respectively, to account for the estimation delay. It is evident that both estimations follow the reference, albeit with a low resolution. As previously mentioned, this is one of the flaws with the standard FFT approach for estimating the breathing rate.

To tackle the resolution problem, the chirp z-transform (CZT) approach was from here on utilized for the estimations. A breathing estimation of this sort is presented in Figure 6.2, while still presenting the FFT approach as a comparison. It is clear that the CZT approach outperforms FFT in this estimation, which can be understood analyzing the Root Mean Square Error (RMSE)

$$E = \sqrt{\frac{1}{N} \sum_{i=1}^N (Y_i - \hat{Y}_i)^2}, \quad (6.1)$$

where \hat{Y}_i is the estimated value, Y_i the true value and N the number of samples. Performing the calculation on both approaches of Figure 6.2 gives $E = 0.73$ breaths/min for the FFT approach and $E = 0.35$ breaths/min for the CZT.

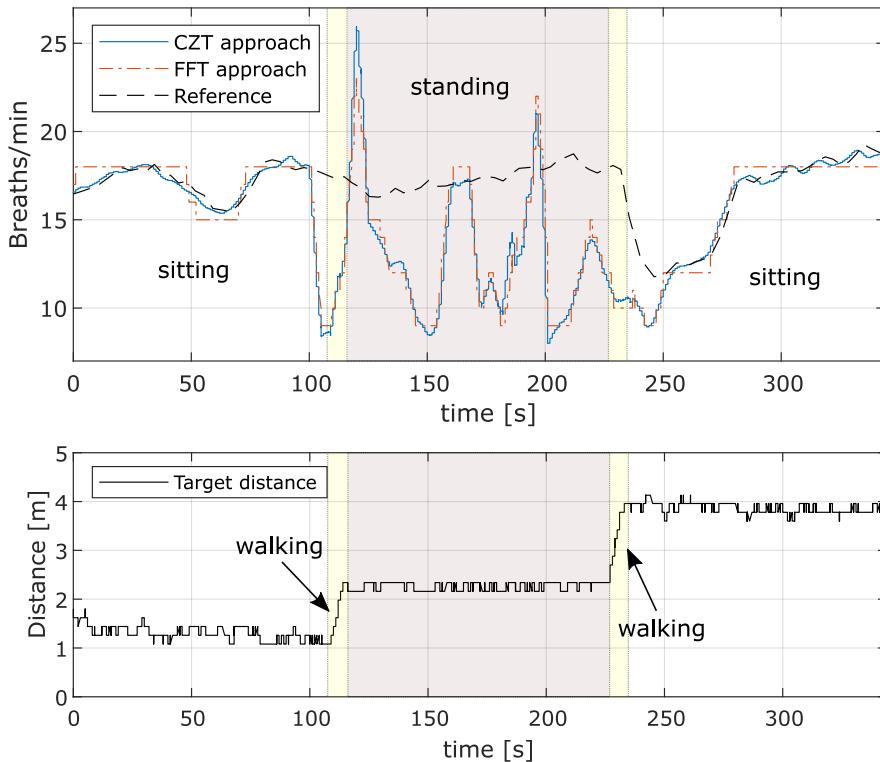


Figure 6.3: Breathing rate- and distance measurement for a semi-moving human target. It can be seen that the estimation is accurate for a person sitting down, but less so when a person is standing.

Another interesting measurement is presented in Figure 6.3, that shows the breathing rate estimation for a semi-moving target in a home care environment. This measurement was performed for almost six minutes of a person sitting down close to the radar, then walking away and standing in front of the radar, and lastly walking further away to sit down at a chair further down the room. The previously mentioned vibrating target localization was here used to track the breathing human target, while the phase unwrapping was performed across range bins. It is clear from the measurement that both localizing the human target and performing an accurate breathing rate estimation is possible in a home care environment for a person sitting down, or for a person relocating between sitting positions at different ranges. However, while walking and while standing the estimation performance is greatly reduced. For the standing case, the RMSE increases from 0.27 breaths/min to 5.74 breaths/min with the CZT approach. This issue comes from the fact that a standing human constantly oscillates back and forth trying to keep balance, which creates a radar return strong enough to mask the breathing rate in frequency domain.

Measurement nr.	Distance [m]	Pose	Approach	RMSE
1	4.5	Sitting	Peak	0.80 /min
1	4.5	Sitting	FFT	1.21 /min
2	2.3	Sitting	CZT	0.35 /min
2	2.3	Sitting	FFT	0.73 /min
3	1-4	Overall	CZT	3.75 /min
3	1-4	Overall	FFT	3.69 /min
3	1-4	Sitting	CZT	0.27 /min
3	1-4	Sitting	FFT	0.66 /min
3	1-4	Standing	CZT	5.74 /min
3	1-4	Standing	FFT	5.59 /min

Table 6.1: Performance of breathing rate estimations for all measurements.

Nonetheless, Table 6.1 concludes the calculated RMSE for all the measured breathing estimations. From the RMSEs it is clear that the CZT approach performs best for the breathing estimation; at least for a sitting person.

The first heart rate measurement is illustrated in Figure 6.4. This measurement was performed of a person sitting down at 0.9 m distance from the radar, while carrying a pulse oximeter on the index finger for reference. Here the RMSE was determined as $E = 5.73$ beats/min. One more measurement,

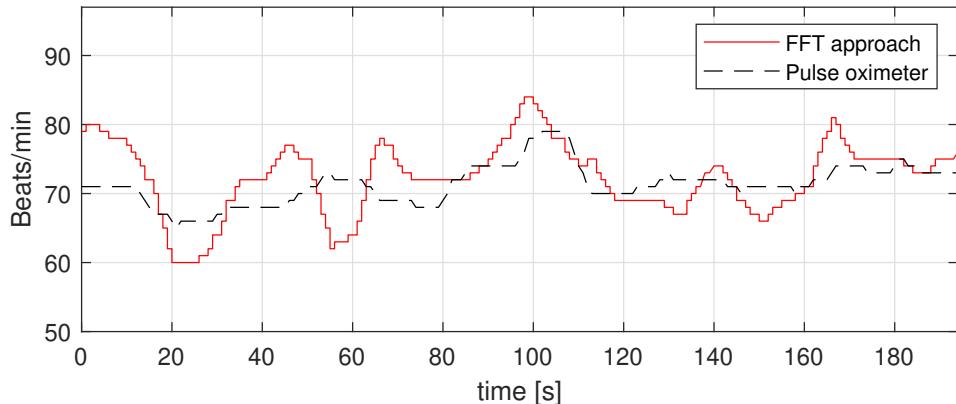


Figure 6.4: Heart rate measurement by the radar at 0.9 m distance. The reference signal was measured with pulse oximeter CMS50D+. Note that the signal has been time shifted to account for the estimation delay.

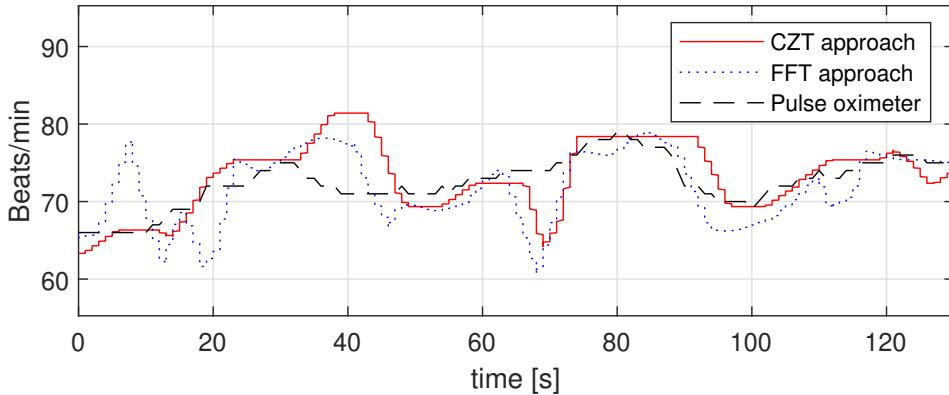


Figure 6.5: Heart rate measurement by the radar at 1.0 m distance. The reference signal was measured with pulse oximeter CMS50D+. Note that the signal has been time shifted to account for the estimation delay.

illustrating the CZT approach performance, can be seen in Figure 6.5. Here the error in comparison to the reference was lower, with a RMSE of $E = 3.85$ beats/min for the FFT approach and $E = 3.49$ beats/min for the CZT approach. The CZT method appears to outperform the FFT, but discrepancy between reference and estimation still exist due to interference of harmonics, phase noise, and difference of compared methods (the pulse oximeter uses a time domain peak counting method, whereas the proposed method is in frequency domain, yielding the most occurring heart rate, which may be different to the mean heart rate). The greatest discrepancy can be seen at the 40 second mark and 70 second mark of Figure 6.5, where both estimations deviate from the reference.

A compilation of the RMSEs of the heart measurements is presented in Table 6.2, where it can be seen that the CZT approach was the least erroneous. Measurements were only performed for sitting poses, although any similar static pose would produce similar results, e.g. for a person lying down.

Measurement nr.	Distance [m]	Pose	Approach	RMSE
1	0.9	Sitting	FFT	5.73 /min
2	1.0	Sitting	CZT	3.49 /min
2	1.0	Sitting	FFT	3.85 /min

Table 6.2: Performance of heart rate estimations for all measurements.

6.2 Fall Detection Results

Multi-label classification for the fall detection was used to preserve scientific transparency and not skew the results in any favor. If binary classification of non-falls and falls is used, it is easy to skew the results by including a great number of non-fall-like gestures. Furthermore, a confusion matrix of the multi-label classification for the fall detection is presented in Figure 6.6, where label 0 denotes the fall class. The notation of the indexes and their corresponding classes, along with their corresponding precision, recall and F1-score is presented in Table 6.3. From the confusion matrix, it is evident that the walking gesture (label 6) and noise (label 7) was easily classified. Never did the model falsely predict these classes as falls, and never were falls predicted as these classes. This is a good sign, as walking and especially noise will be part of the most frequent signals in a home care environment; meaning false alarms for these signals are highly undesirable. Moreover, as seen from the confusion matrix, 85% of the fall test files were correctly predicted

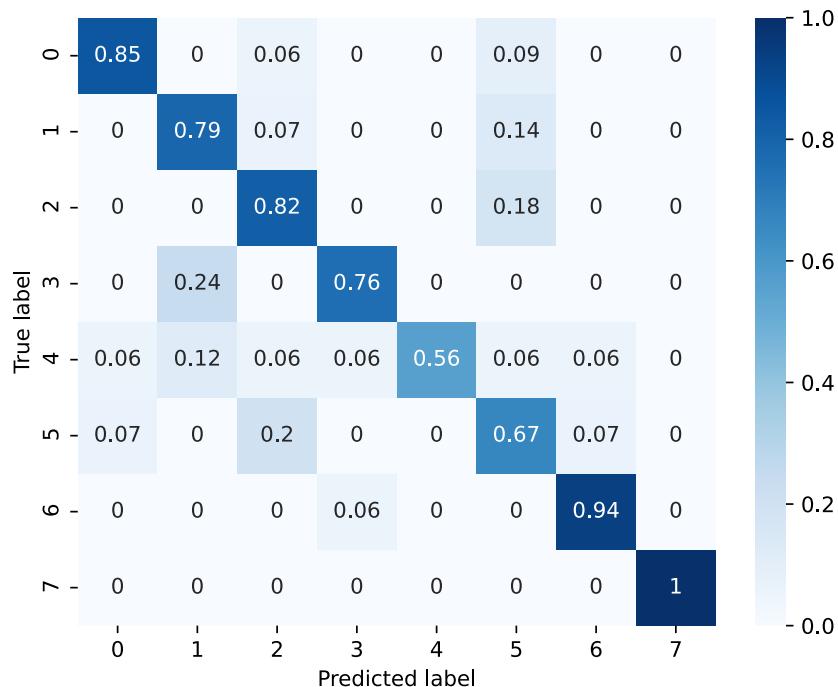


Figure 6.6: Confusion matrix of the multi-label classification. Y-axis are the true labels and X-axis the predicted ones. Label 0 denotes falling, 1 bending, 2 jumping, 3 kneeling, 4 lying down, 5 sitting down, 6 walking and 7 is noise (no action).

as falls. The false negative rate is hence 15%, with 6% jumps (label 2) and 9% sitting motions (label 5) being the false negatives. For the false positives (false alarms), this included 6% of the lying down motions (label 4) and 7% of the sitting motions (label 5).

To elaborate on evaluation measures, Table 6.3 also presents the precision,

Class	Index	Precision	Recall	F1-score	Support
Fall	0	0.94	0.85	0.89	34
Bend	1	0.65	0.79	0.71	14
Jump	2	0.56	0.82	0.67	11
Kneel	3	0.87	0.76	0.81	17
Lie	4	1.00	0.56	0.72	16
Sit	5	0.56	0.67	0.61	15
Walk	6	0.88	0.94	0.91	16
Noise	7	1.00	1.00	1.00	23

Table 6.3: Evaluation measures of the classified test data.

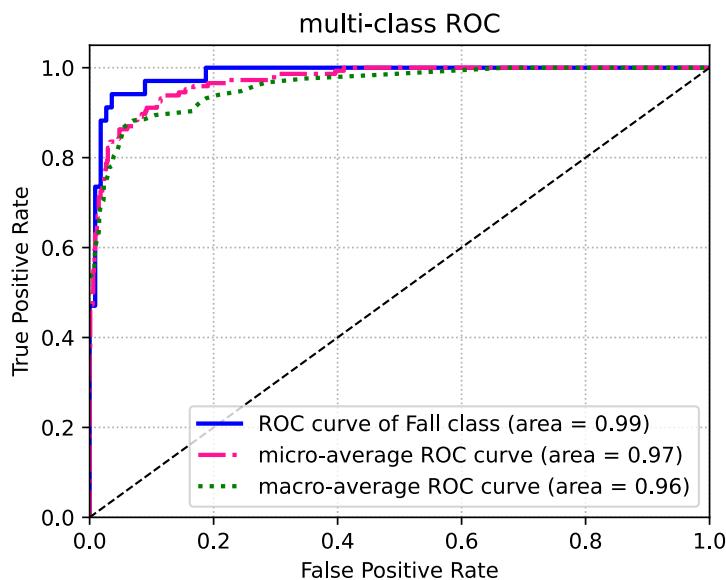


Figure 6.7: ROC plot, illustrating the relation between true positives and false positives for different thresholds of accepting a gesture as a fall (blue graph). The two other graphs are averages of all gestures; showing the overall performance of the classifier.

recall and F1-score of all the predicted classes. The F1-score is defined as

$$F_1 = 2 \frac{P_{pr} P_{re}}{P_{pr} + P_{re}}, \quad (6.2)$$

where P_{pr} is the precision of the predicted class and P_{re} is its recall. It is clear from the table that the classifier performed best at predicting noise, walking and falls.

The trade-off between false negatives and false positives, or the relation between true positives and false positives is presented in a Receiver Operating Characteristic (ROC) plot in Figure 6.7. Here the threshold for accepting a gesture as a fall is varied, resulting in either few false alarms or few false negatives. A sweet spot in the graph may be identified at the 3.5% false positive rate and 94% true positive rate (6% false negative rate).

To reinforce the results, an accuracy plot and loss plot of the classifier's training stage is also provided. The accuracy plot is presented to the right in Figure 6.8 and the loss plot can be seen to the left. As seen, the graphs seem to converge over the epochs, albeit the model may be slightly overfit judging the deviation between loss of train data and test data after the 18th epoch.

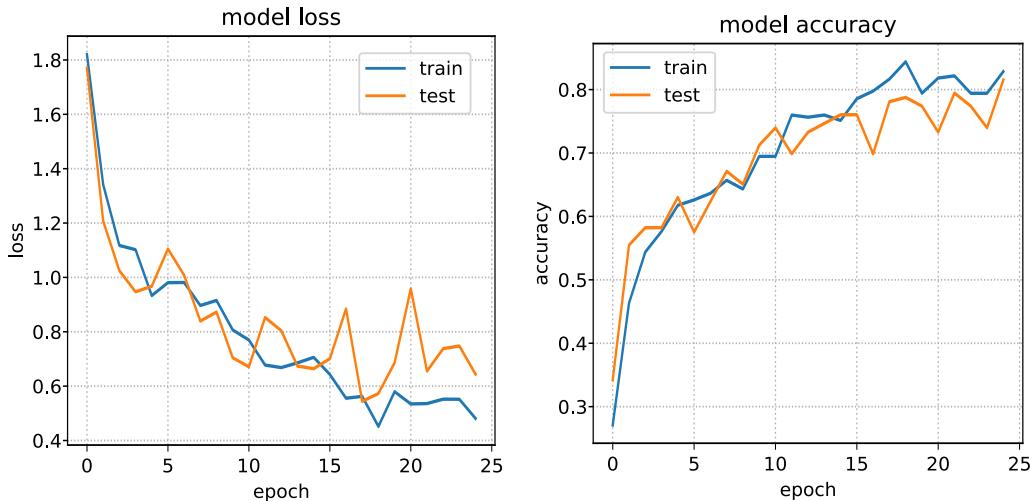


Figure 6.8: Loss (left) and accuracy (right) for all epochs during training stage for train data and test data.

Chapter 7

Discussion

7.1 Evaluation of Vital Signs Results

From the results of the breathing estimations and heart rate estimations, it is evident that there is a fairly low error between reference signals and the results of the radar. However, an estimation as simple as taking the mean of the signal may also yield an error of similar magnitude. Hence it is of interest to study a correlation that disregards the mean, to ensure that the measurement follows the reference. Introducing the Pearson correlation

$$r_{XY} = \frac{\sum_{n=1}^N (x_n - \bar{x})(y_n - \bar{y})}{\sqrt{\sum_{n=1}^N (x_n - \bar{x})^2 \sum_{n=1}^N (y_n - \bar{y})^2}}, \quad (7.1)$$

where N is the number of samples, x_n is the estimated frequency, y_n the reference frequency, while \bar{x} and \bar{y} are their corresponding means. For values of r_{XY} , 1 and -1 both correspond to full positive or full negative correlation, while $r_{XY} = 0$ indicates no correlation. Carrying out the calculation for the heart rate CZT result of Figure 6.5 results in $r_{XY} = 0.655$, and for the breathing CZT estimation of Figure 6.2 a correlation of $r_{XY} = 0.968$ is obtained. This can be compared to other works [4], [5], [6], achieving heart rate estimations with r_{XY} in the range $0.552 \leq r_{XY} \leq 0.872$ and breathing rate estimations in the range $0.911 \leq r_{XY} \leq 0.94$, all utilizing FMCW radars with frequencies ranging from 9.6 GHz to 80 GHz. Judging from this data, it appears that the heart rate estimation with the adopted methods and IWR6843AOPEVM radar is mediocre. However, the breathing rate estimation seems to excel the other works'.

In terms of real life scenarios, both heart rate- and breathing rate estimations appear to be limited to at least semi-static subjects (with the utilized

methods), considering the results of Figure 6.3. Here semi-static refers to a subject that remains still (e.g. sitting or lying) for approximately the window processing period in time (20 s in these results), but is allowed to relocate position once the time has passed. Furthermore, there are many possible obstacles in front of the body or chest region to worsen the detection, such as jackets and blankets. Even chairs, tables and walls can block the line of sight. With this in mind, the estimations of a radar implemented in a home care environment, with non-ideal conditions, is deemed to produce worse results than achieved here.

7.2 Evaluation of Fall Detection Results

To understand the performance of the implemented multi-label classification, a comparison can be carried out of the ROC curve result of Figure 6.7 with the works' of others. In [3] binary classification was used with two classifiers on two separate data sets, yielding ROC curve areas ranging from $0.918 \leq A_{ROC} \leq 0.974$. Furthermore, [1] provided similar results with the best classifier having $A_{ROC} \approx 0.97$. In comparison, the proposed feature extraction and time sequence LSTM network provided $A_{ROC} = 0.99$ for the falls of the recorded data, which is higher than the mentioned sources. However, a higher number does not necessarily entail a better classifier, especially not when evaluating the performance based on merely 34 test falls. It is possible that the recorded data is biased due to the recording process. To elaborate, in machine learning, it is important to include a complete and representative set of data for the training and classification. In this instance, all the motions were performed by males aged 20-30. The falling event of an elderly person would likely be slower than that of a young male, hence there is certainly a bias in the trained model. Consequently, fall detection accuracy would likely decrease when including a more versatile population for the classification. However, the model could of course train on this data as well, but it requires far more ambitious data capturing, with many thousands of video recordings instead of 729.

If the motion of the fall of an elderly is too slow to show clear and significant doppler features, it may be more relevant to use elevation Angle of Arrival (AoA) to include data capture of the person's height. This would give more false alarms for the lying down class, but has greater potential in discriminating the fall class from many other gesture classes. The issue with the IWR6843AOPEVM radar in this regard is that it only has 4 independent virtual antennas for measuring elevation angle. As explained later, this yields an angular resolution of 28° , which is very coarse for distances over 4-5 me-

ters. The angular resolution could be improved by cascading radars, which is further discussed in Section 7.4.2. Nonetheless, having such a coarse angular resolution is the reason as to why a range-doppler-time approach was explored within this work. Keep in mind that the (rather good) results were obtained with recordings having only a frame rate of 9/s, range resolution of 8 cm and velocity resolution of 0.11 m/s, where all the resolutions had to be restricted due to the UART baud rate bottleneck. This is promising results for the explored approach, as obtaining finer resolutions is a rather trivial task that could further improve the fall detection.

7.3 Fields of Application

It is clear that a fall detection radar could be implemented within home care to aid elderly people. Nonetheless, as with any detection algorithm, there are risks of not detecting the fall, and for false alarms to occur. Hence it is most likely best to implement the radar in conjunction with some other solution, such as personal alarm. Especially in today's smart homes, with present Bluetooth and WiFi, the multiple solutions could be used in conjunction and communicate with each other to produce far more accurate results than one solution could alone. Moreover, the actual implementation of the fall-detection radar could be quite simple, as the processing and real-time data comparison to machine learning model could be handled on a Raspberry Pi connected to the IWR6843AOPEVM radar, protected by some form of radome. This package could be installed in the top corner of a room, ready to detect.

For the vital signs, the estimations require the subject to remain mostly still. It may be used within home care too, but limited to certain positions. For instance, the radar could monitor breathing rate and heart rate during sleep. It could also determine the rates for someone sitting down in a chair, in a sofa, in front of a desk, in a car seat, etc. It is worth noting that the breathing- and heart rate estimation is not limited to human subjects. It may be used for animals as well, simply by altering the cut-off frequencies of the different filters in the method to adapt it to their standard rates. For instance, it could be used to track the vital signs of a horse in its stable box, or cattle in their cowshed. These are both limited and mostly static environments which suits the purpose well. A FMCW radar further has the possibility of detecting vital signs of multiple subjects simultaneously, which could make it a feasible option.

7.4 Future Work

Multiple measures can be taken to improve or extend the utilized methods. Below, a few of such possibilities are explored.

7.4.1 Linking Harmonics for Improved Estimation Performance

In estimating breathing rate and heart rate, with the methods' of this work, only the fundamental harmonic frequencies were utilized to obtain the actual estimations. However, multiple harmonics of breaths and heartbeats are present in the frequency spectrum of the received and processed IQ-signal. With this in mind, it is wasteful to simply low-pass filter the signal to suppress the harmonics. It would be possible to extend the processing to detect and isolate a number of harmonics through band-pass filters, and through cross-correlations or similar, increase the estimation accuracy of the heart rate and breathing rate.

7.4.2 Cascading of Radars

Multiple radars could be cascaded to capture more data and to dramatically increase the measured angular resolutions corresponding to the angle of arrival. Even with the range-doppler-time approach, cascaded radars would improve the accuracy of capturing human motion signatures, due to greater total

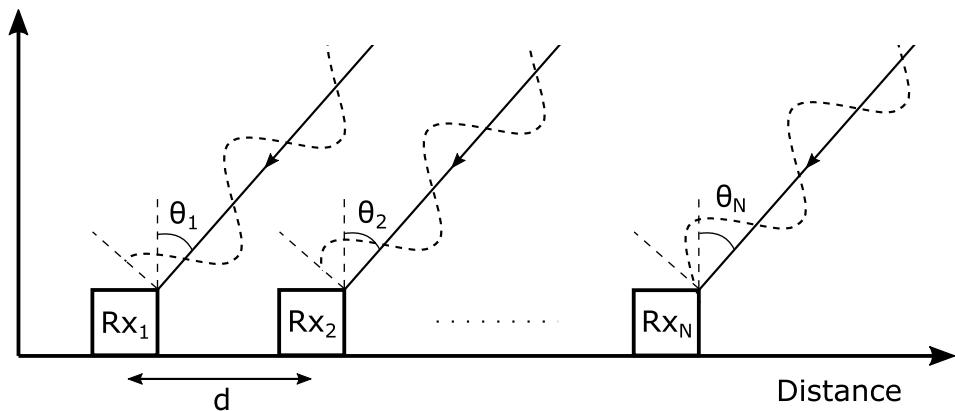


Figure 7.1: The incident electromagnetic waves reach the antennas in different phases, with the angle of arrival shown as θ .

transmission rates, allowing improved temporal-, spatial- and velocity resolutions. Moreover, with an elevation angle resolution greater than $\theta_{RES} \approx 28^\circ$, it would be possible to accurately include height measurements for the fall detection algorithm, which could improve the fall detection accuracy. Furthermore, by only cascading one more IWR6843AOPEVM radar, twice the transmitters and receivers are obtained. With correct setup of timings and chirps, a MIMO-TDM approach could result in four times the number of virtual antennas along elevation axis, and thus an angular resolution 4 times smaller ($\theta_{RES} \approx 7^\circ$)

$$\theta_{RES} = \frac{\lambda}{Nd \cos(\phi)}. \quad (7.2)$$

Here the spacing between antennas in one dimension is denoted $d \cos \phi$, λ is the wavelength, N is the number of unique virtual antennas along one axis and θ_{RES} is the angular resolution (in radians) for each independent axis. See Figure 7.1 for reference.

Most important in terms of fall detection is to increase the number of virtual antennas along the elevation axis, to resolve height measurements. An approach of this sort could be used with CFAR detection to find the three-dimensional position of a number of scattered points. The position of these points could further be recorded for different gestures and used to train a neural network model. This would likely yield better fall detection results in comparison to the range-doppler-time approach for slow, real life, falling motions as the height data could outshine the radial velocity data in the training stage of the machine learning model.

7.4.3 Machine Learning for Estimating Vital Signs

Another interesting way to estimate the vital signs is to do it with machine learning. A simple approach of achieving this would be to record a couple of hundred files, including complex range data, of a person breathing in front of the radar. If the breathing rate and heart rate would be simultaneously measured and recorded by e.g. a chest strap and a pulse oximeter, the radar data could be processed and fit through machine learning to match their results, utilizing some kind of time sequence neural network.

A challenge with this approach would be to record a complete representative set of samples; including low to high heart frequencies and breathing frequencies, at different ranges and different angles between the radar and the human subject. However, there is a chance that a machine learning approach could further reduce the issues of breathing harmonics. This could in turn yield more robust heart rate estimations.

Chapter 8

Conclusions

By this study it is deemed that both vital signs detection and fall detection is possible within a home care environment. The fall detection shows great possibilities, as promising results were reached despite a clear bottleneck in UART transmission rates for capturing the gesture data. A similar range-doppler-time procedure could be adopted with better results by simply using a different communication protocol to allow higher data transmission rates and thereby better range-, velocity- and time resolution. Other methods to use point-cloud data to capture the elevation angle with cascaded radars for sufficient angle resolution may provide more accurate results; albeit at the cost of requiring further antenna elements. The range-doppler-time approach may hence be a cost-effective and yet well-performing solution.

For the vital signs, it is clear that monitoring breathing rate and heart rate requires the monitored person to remain sufficiently still, as excessive motions cause too much spectral noise for the signs to be accurately estimated. However, in a home care environment this is probably the most important time to study the signs; as it includes sleeping, sitting and lying down. Sleep apnea and cardiac issues could hence be tracked remotely while sleeping. Another idea is to combine the fall detection and vital signs solutions in one and specifically activate the vital signs detection after a fall incident has occurred. This could then, through the alarm service provider, provide critical information for a paramedics team ready to support the injured person who fell.

Acknowledgements

I would like to thank numerous people for aiding and guiding me through this project and making all of this possible. Firstly, I would like to thank Gustav Norvik, being my supervisor at the company STANLEY Security where I carried out the thesis, for counseling and providing materials such as the IWR6843AOPEVM radar. Moreover, obtaining the machine learning results from the fall detection would not have been possible without Mirza Okanovic, who dedicated many hours to discuss the machine learning and support coding the algorithms.

Umer Abbas Shah, being my supervisor at Royal Institute of Technology (KTH), assisted with many fruitful discussions to ease understanding of the radar hardware. He also reviewed the report multiple times to give feedback, which I am very grateful for.

I am grateful for both Gurpreet Singh and Gustav Norvik in their efforts to assist in recording data for the fall detection. Being so many recordings, this included many hours in time and multiple bruises from all the falls, but ultimately resulted in a better representation of data.

Lastly, I would like to thank my father. Not only did he endure my constant talk of mathematical equations and radars, but he also provided a second location for recording of fall detection data, which further improved the results. In the times of COVID when many studying locations were closed, he also gave me the opportunity to utilize a workplace in his company, helping me uphold my routines and finishing the thesis.

Bibliography

- [1] Moeness G Amin et al. “Radar signal processing for elderly fall detection: The future for in-home monitoring”. In: *IEEE Signal Processing Magazine* 33.2 (2016), pp. 71–80.
- [2] Baris Erol and Moeness G Amin. “Radar data cube analysis for fall detection”. In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2018, pp. 2446–2450.
- [3] Liang Liu et al. “Automatic fall detection based on Doppler radar motion signature”. In: *2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth) and Workshops*. IEEE. 2011, pp. 222–225.
- [4] Mostafa Alizadeh et al. “Remote monitoring of human vital signs using mm-Wave FMCW radar”. In: *IEEE Access* 7 (2019), pp. 54958–54968.
- [5] Laura Anitori, Ardjan de Jong, and Frans Nennie. “FMCW radar for life-sign detection”. In: *2009 IEEE Radar Conference*. IEEE. 2009, pp. 1–6.
- [6] Siying Wang et al. “A novel ultra-wideband 80 GHz FMCW radar system for contactless monitoring of vital signs”. In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2015, pp. 4978–4981.
- [7] Merrill Ivan Skolnik. “Radar handbook”. In: (1970).
- [8] Mark A Richards et al. *Principles of modern radar*. Citeseer, 2010.
- [9] Reinhard Feger et al. “A 77-GHz FMCW MIMO radar based on an SiGe single-chip transceiver”. In: *IEEE Transactions on Microwave theory and Techniques* 57.5 (2009), pp. 1020–1035.
- [10] Merrill Ivan Skolnik et al. *Introduction to radar systems*. Vol. 3. McGraw-hill New York, 1980.

- [11] Cesar Iovescu and Sandeep Rao. “The fundamentals of millimeter wave sensors”. In: *Texas Instruments, SPYY005* (2017).
- [12] Andrew G Stove. “Linear FMCW radar techniques”. In: *IEE Proceedings F (Radar and Signal Processing)*. Vol. 139. 5. IET. 1992, pp. 343–350.
- [13] Mark A Richards. *Fundamentals of radar signal processing*. Tata McGraw-Hill Education, 2005.
- [14] Alex Zwanetski and Hermann Rohling. “Continuous wave MIMO radar based on time division multiplexing”. In: *2012 13th International Radar Symposium*. IEEE. 2012, pp. 119–121.
- [15] Frank C Robey et al. “MIMO radar theory and experimental results”. In: *Conference Record of the Thirty-Eighth Asilomar Conference on Signals, Systems and Computers, 2004*. Vol. 1. IEEE. 2004, pp. 300–304.
- [16] Christoph Will et al. “Radar-based heart sound detection”. In: *Scientific reports* 8.1 (2018), pp. 1–14.
- [17] Antonio Lazaro, David Girbau, and Ramon Villarino. “Analysis of vital signs monitoring using an IR-UWB radar”. In: *Progress In Electromagnetics Research* 100 (2010), pp. 265–284.
- [18] Changzhi Li, Yanming Xiao, and Jenshan Lin. “Experiment and spectral analysis of a low-power *Ka*-band heartbeat detector measuring from four sides of a human body”. In: *IEEE Transactions on Microwave Theory and Techniques* 54.12 (2006), pp. 4464–4471.
- [19] Victor C Chen et al. “Micro-Doppler effect in radar: phenomenon, model, and simulation study”. In: *IEEE Transactions on Aerospace and electronic systems* 42.1 (2006), pp. 2–21.
- [20] Tom M Mitchell et al. “Machine learning”. In: (1997).
- [21] Ian Goodfellow et al. *Deep learning*. Vol. 1. MIT press Cambridge, 2016.
- [22] David C Champeney and Raymond M Fuoss. “Fourier transforms and their physical applications”. In: *PhT* 26.10 (1973), p. 57.
- [23] *IWR6843AOP Single-Chip 60- to 64-GHz mmWave Sensor Antennas-On-Package (AOP)*. Texas Instruments. Apr. 2020. 74 pp.
- [24] Chethan Kumar Y.B. *60GHz mmWave Sensor EVMs*. Version Rev. D. Texas Instruments. Dec. 2020. 81 pp.

- [25] *MMWAVE SDK User Guide*. Version 1.0. Texas Instruments. Jan. 18, 2019. 70 pp.
- [26] Guochao Wang et al. “Application of linear-frequency-modulated continuous-wave (LFMCW) radars for tracking of vital signs”. In: *IEEE Transactions on Microwave Theory and Techniques* 62.6 (2014), pp. 1387–1399.
- [27] Bram Lohman et al. “A digital signal processor for Doppler radar sensing of vital signs”. In: *IEEE Engineering in Medicine and Biology Magazine* 21.5 (2002), pp. 161–164.
- [28] Lawrence R Rabiner, Ronald W Schafer, and Charles M Rader. “The chirp z-transform algorithm and its application”. In: *Bell System Technical Journal* 48.5 (1969), pp. 1249–1292.
- [29] L Rabiner, RW Schafer, and C Rader. “The chirp z-transform algorithm”. In: *IEEE transactions on audio and electroacoustics* 17.2 (1969), pp. 86–92.
- [30] Sergey Ioffe and Christian Szegedy. “Batch normalization: Accelerating deep network training by reducing internal covariate shift”. In: *arXiv preprint arXiv:1502.03167* (2015).

Appendix

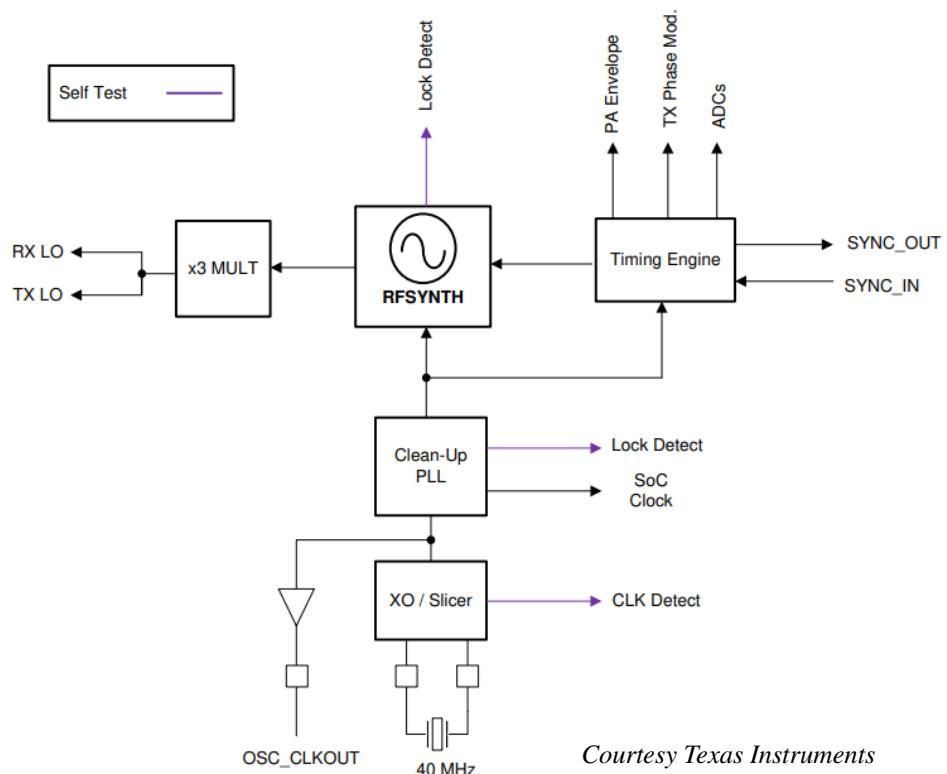


Figure A.1: Clock system of the IWR6843AOPEVM board, by [23].

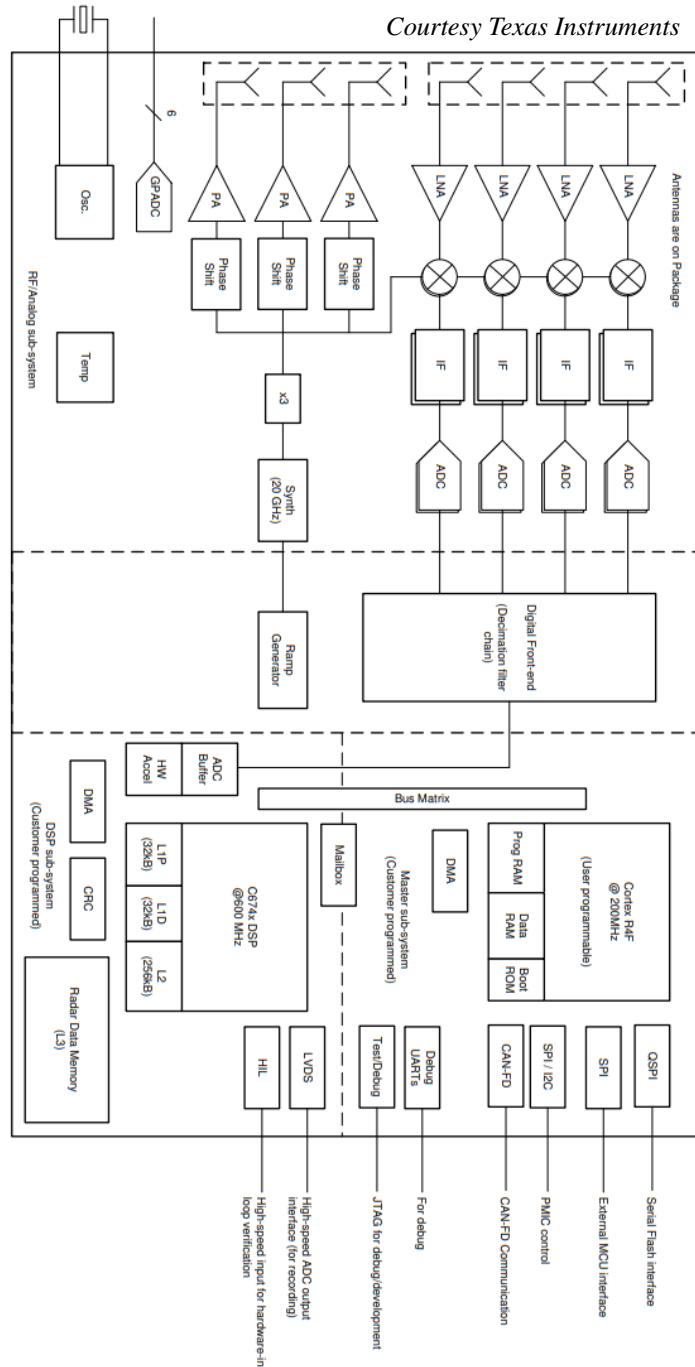


Figure A.2: Block diagram of the IWR6843AOPEVM board, by [23].



Figure A.3: Second location where the radar was installed (green circle) and gestures were recorded. This location was intended to be more similar to a home care environment.

TRITA -EECS-EX-2021:76