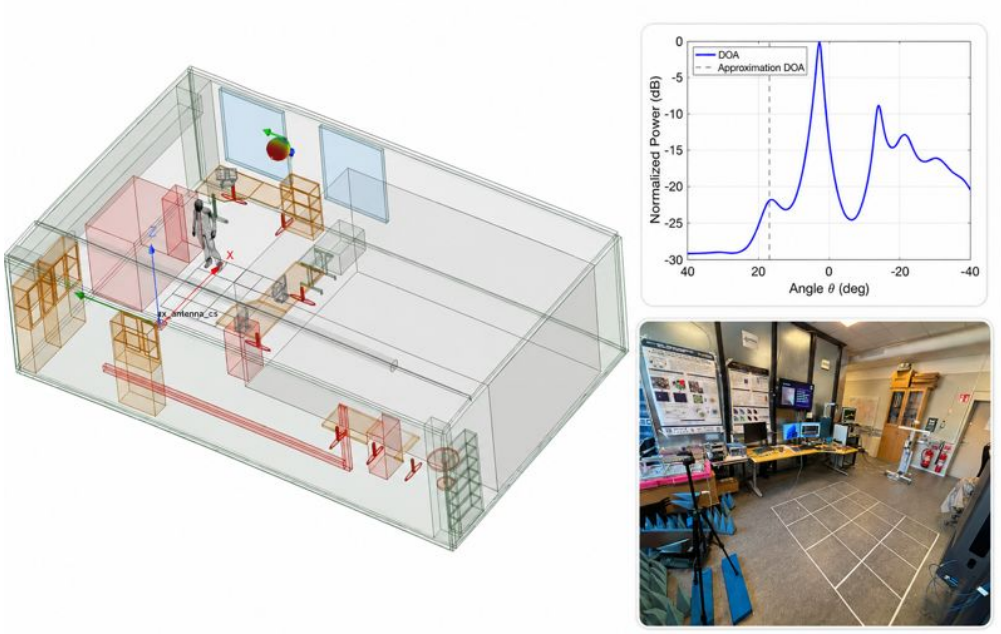
CHALMERS




# 6G Sensing Systems: From Radio Channels to Machine Learning (wECHO2)

Bachelor thesis in Electrical Engineering

DEPARTMENT OF ELECTRICAL ENGINEERING

CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2026
www.chalmers.se

Bachelor's thesis 2026

# 6G Sensing Systems: From Radio Channels to Machine Learning (wECHO2)

HANNA ALTSTEN
OSSIAN BREDMAR
NATALIE JOHANSSON
HANNA JÖNSSON
ASAL MAHMOODI
ERIK ROSLUND


CHALMERS

Department of Electrical Engineering
*CAOS division*
Antenna group
Chalmers University of Technology
Gothenburg, Sweden 2026

6G Sensing Systems: From Radio Channels to Machine Learning (wECHO2)
H. ALTSTEN, O. BREDMAR, N. JOHANSSON, H. JÖNSSON, A. MAHMOODI, E. ROSLUND



Supervisor: Pavlo Krasov, Department of Electrical Engineering
Examiner: Erik Ström, Department of Electrical Engineering

Bachelor's Thesis 2026
Department of Electrical Engineering
CAOS division
Antenna group
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: A picture of the lab environment on the bottom right, a plot from the MUSIC algorithm and the digital twin of the same environment as the background.

Typeset in LaTeX
Gothenburg, Sweden 2026

# Abstract

In the context of future 6th generation (6G) and Integrated Sensing and Communication (ISAC) applications, this thesis explores the possibility of using data from the $S_{21}$ parameter for indoor detection of human presence in the 4 GHz to 16 GHz frequency range. Physical measurements in Chalmers' antenna lab were combined with simulations in a digital twin of the lab environment to analyze how a stationary human affects the wireless channel. The Multiple Signal Classification (MUSIC) algorithm was applied to both physical and simulated data in order to analyze directions of arrival (DoA) and reflections caused by human presence. The results show that the MUSIC algorithm was able to detect changes caused by a human in the wireless channel, although the accuracy depended on the environment, the placement of the antennas as well as the assumed number of sources. Additionally, a Convolutional Neural Network (CNN) was trained on simulated and measured data in several dataset combinations. As a result it could be observed that combining simulated and real data in training improved the models ability to adapt to realistic environments, while training on only simulated data resulted in reduced transferability to the real measurements.

# Contents

# List of Figures

# List of Tables

# 1

# Introduction

This chapter presents the background of the thesis and explains why it is relevant for current sixth generation (6G) research. The purpose of the thesis is stated, as well as the scope which includes important limitations and assumptions.

## 1.1 Background

The sixth generation of mobile networks is envisioned to not only extend current network capabilities, but also to support future technology applications. Building on fifth generation (5G) networks, 6G is expected to drastically increase throughput and reduce latencies, as well as introduce new opportunities such as the internet of things [1]. These opportunities may utilize integrated sensing and communication, in which radio signals can be used for data transmission as well as for obtaining information about the surrounding environment. 6G could therefore be used to largely improve entire systems that require awareness of the physical environment around them.

Wireless channels play a central role in understanding such systems, as they describe the transmission path of the wave between the transmitter and receiver [2]. This thesis will simulate these wireless channels in an indoor environment with a deterministic model. The simulations will be compared with real-world measurements done with a Vector Network Analyzer (VNA) to design efficient signal processing algorithms.

This thesis focuses on human detection, which can be realized with integrated sensing and communication (ISAC) [3][4]. Currently, infra-red sensors are widely used for this purpose but are limited in their precision. 6G has the potential to replicate this function with the help of software. The accuracy need further testing, which this thesis will account for.

## 1.2 Purpose

The purpose of this thesis is to detect human presence in an indoor physical environment using a VNA and its $S_{21}$ parameter as a wireless channel between a transmitting and receiving antenna. To support the physical measurements, human presence is to be detected in a simulated digital twin of the laboratory that suffi-

ciently matches the real environment's parameters. The simulations are done with the software ANSYS HFSS. The measurement data collected in both environments are then inserted into a MUSIC algorithm and used to train a machine learning algorithm, aiming to assist the cause of human detection.

## 1.3 Scope

- Only fixed and low power (<1 mW) measurements will be done, where the position of the antennas will be fixed and the measured persons position will be changed.
- The attenuation, such as multi-path fading, will be neglected because the MUSIC algorithm primarily focus on the direction of the wave.
- The frequency span for the measurements will be limited to 4-16 GHz due to the VNA's range and the physical constraints from the antennas and the coaxial cables.
- The scope will focus on trying to detect one human.
- Only compare measurements with the digital twin of the room and not with other indoor environments. The simulations will also use the built-in component library.
- Only one configuration will be used for the antennas position.

# 2

# Theory

The following section presents the necessary theoretical background for this thesis. First, the fundamentals of 6G, Integrated Sensing and Communication, and the multipath effect are introduced. Furthermore, the chapter covers S-parameters, Vector Network Analyzer, the principle of Fourier Transform, the fundamentals of the MUSIC algorithm, machine learning, and windowing.

## 2.1 Fundamentals of 6G

Understanding the evolution towards the 6th generation (6G) of wireless systems requires evaluating both the achievements and limitations of 5th generation (5G) systems. 5G was introduced as a key enabler of Internet of Everything (IoE), with the goal of enabling large scale connectivity between users, devices, and machines [5]. It was intended to operate at high-frequency millimeter wave, while supporting heterogeneous IoE services. Certain capabilities, such as basic IoE functionality and Ultra-Reliable Low-Latency Communication (URLLC) for applications including industrial automation, were not fully implemented at the large scale that was initially anticipated. Future applications will require communication systems capable of providing URLLC for heterogeneous devices across both up-link and down-link communication, while also integrating computing, control and communication through end-to-end co-design. These requirements exceed the capabilities of 5G, making it insufficient for supporting emerging IoE applications such as extended reality, including mixed reality, augmented reality, and virtual reality.

Therefore, one expected cornerstone of 6G is that its design will inherently support a true IoE, enabling an immersive and fully connected cyber-physical world [5][6]. This is expected to be achieved through capabilities such as extreme reliability and determinism, digital twin environments, massive-scale IoE ecosystems, and Integrated Sensing and Communication (ISAC). In addition to the IoE capabilities, 6G is expected to introduce the use of new spectrum bands extending into the 1 terahertz (THz) band. Although communication at these frequencies is associated with challenges such as molecular absorption and severe path loss, 6G will provide substantial improvements in capacity, precision, and sensing integration. Furthermore, terabit per second peak rates, sub-millisecond latency, and microsecond synchronization are expected to enable advanced application such as holographic telepresence, Tactile Internet 2.0, and brain-to-computer interfaces.

In all wireless communication systems, the wireless channels are important [7]. They can be defined as the proportion medium between a transmitter and a receiver, where the transmitted signal is affected by both large-scale and small-scale propagation effects. These effects are commonly modeled using geometry-based stochastic channel models. In 5G millimeter wave systems, the small-scale fading is typically represented by clusters of multiple propagation rays.

## 2.2 Integrated Sensing and Communication (ISAC)

ISAC is a key technology for the development of 6G networks, and it refers to wireless systems that can perform data communication and environmental sensing simultaneously [8]. Traditionally, wireless systems are designed where communication systems focuses on maximizing the data rate, reliability, and efficiency. Additionally, radar systems aim to detect objects, estimate distance, velocity, and direction. ISAC seeks to combine these functionalities into one unified system in order to improve spectral efficiency, reduce hardware complexity, and enable new applications such as smart environments, localization, and human presence detection.

The fundamental principle behind ISAC is that wireless signals interact with the surrounding environment as they propagate between a transmitter and a receiver [9]. During propagation, signals undergo reflection, scattering, diffraction and attenuation by objects such as walls, furniture, and humans. These interactions create a multipath propagation environment where multiple delayed versions of the signal reach the receiver. As a result, the received signal contains embedded information about the environment through the wireless channel response.

The wireless channel in conventional communication systems is often treated as a distraction that must be estimated and compensated for in order to achieve reliable data transmission [9]. However, in ISAC systems the channel is considered a valuable source of sensing-information where the changes in the environment, such as the presence or movement of a human, lead to measurable variations in channel characteristics including amplitude, phase, delay, and multipath structure. By analyzing these variations, it becomes possible to detect objects and human presence without the need for dedicated sensing hardware. By extracting parameters such as the angle of arrival and time delay, ISAC systems reuse communication signals and channel estimation data for sensing, which enables efficient use of spectrum and hardware resources.

This integration of sensing and communication is expected to support a wide range of future applications, such as indoor localization, security monitoring, smart homes, and industrial automation [8]. Despite its potential, ISAC also introduces challenges including balancing communication performance with sensing accuracy and addressing privacy and ethical considerations related to environmental monitoring or human detection.

## 2.3 Multi-path effect and human presence

In a wireless channel, the signal will travel from the transmitting antenna to the receiving antenna through multiple paths rather than a direct trajectory [10]. This occurs due to objects that act like reflectors and scatterers in the surrounding environment. Each propagation path may differ in parameters such as amplitude, angle of arrival, and phase, due to constructive and destructive interference caused by these objects.

According to Friis transmission formula, the power received by a receiving antenna as a function of distance can be expressed according to equation (2.1),

$$P_r(d) = \frac{P_t G_t G_r \lambda^2}{(4\pi d)^2} \tag{2.1}$$

where $P_t$ is the transmitted power, $G_r$ is the receiver antenna gain, $G_t$ is the transmitter antenna gain, $\lambda$ represents wavelength in meters, and $d$ represents distance from the transmitter to receiver in meters [11].

When a human is introduced, the characteristics of the wireless channel are altered, and the received signal will consequently include information about the humans presence [11]. Additional paths are introduced by the human body and the total power at the receiver can be represented by equation (2.2).

$$P_r(d) = \frac{P_t G_t G_r \lambda^2}{(4\pi)^2 \left(d^2 + 4h^2 + \Delta^2\right)} \tag{2.2}$$

The parameter $\Delta$ has been introduced, representing human body path length. $h$ is a parameter that accounts for the distance from the line of sight path to a reflection point on the floor or the ceiling. As can be seen from equation 2.2, the received signal will include changes in the propagation path when a human is present. A human alters parameters such as phase and amplitude and causes doppler shift as well as multipath fading.

When analyzing wireless signals for human detection, the wireless channel can be divided into a target channel and a background channel [12]. The background channel consists of multi-path components arising from environmental objects such as walls, ceilings, and furniture. The target channel, in the scenario of sensing a human, instead consists of all the propagation paths that are influenced by the human body. The target channel itself can be interpreted as a concatenation of two segments, a transmitter to human link, and a human to receiver link. Depending on the environment, each segment may contain line-of-sight or non line-of-sight components. The resulting propagation paths arriving at the receiver will subsequently include direct human reflected components as well as indirect paths involving additional reflections from other objects.

## 2.4 S-parameters

Scattering parameters, which also are called S-parameters, are used to observe how radio frequency signals behave in a system [13]. At radio and microwave frequencies, direct measurement of voltages and currents becomes inefficient because of the behavior of the wave, impedance mismatches, and distributed effects. Instead, network behavior is described using S-parameters, which characterize how much of a signal that is transmitted, reflected, and received between ports of a system. For a two-port network, the S-parameters are described in matrix seen in equation (2.3),

$$\begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} \tag{2.3}$$

where $a_i$ is the incident wave at port i, and $b_i$ is the reflected wave at port i.

In wireless channel measurements, the most important parameter is $S_{21}$. It describes how much of the transmitted signal the transmitter that reaches the receiver. The $S_{21}$ parameter is mathematically calculated with equation (2.4).

$$S_{21} = \frac{\text{Received signal at port 2}}{\text{Transmitted signal from port 1}} = \frac{b_2}{a_1} \tag{2.4}$$

When antennas are connected to the ports, $S_{21}(f)$ represents the wireless channel transfer function between the transmitting and receiving antennas [13]. It includes propagation effects such as reflection from walls and objects, scattering, diffraction, attenuation, and multipath propagation. When a person enters the area of measurement, these effects change, which causes measurable variations in the $S_{21}$ parameter.

## 2.5 Vector Network Analyzer

The Vector Network Analyzer (VNA) is a frequency domain channel sounder, used to measure complex valued S-parameters in both magnitude and phase across a specified frequency range [14]. Since the VNA measures both amplitude and phase it allows reconstruction of the complete complex function of the system. A known signal is first generated at port 1 by the VNA, and the signal propagates through the environment via the transmitting antenna. Later on, the received signal at port 2 is measured and the ratio between transmitted and received waves also meaning the $S_{21}$ parameter is computed. Finally, the procedure is repeated over many frequency points.

Although it is a popular instrument, it has several limitations when it comes to developing 6G [14]. VNA have flexible frequency and bandwidth settings within their specifications, which can be extended to high radio frequencies near and above 1 THz, but they still function within hardware limits. The VNA's measurement range is limited because there is a need to use radio frequency (RF) cables to connect

remote antennas, which introduces high losses, specially when it comes to millimeter-wave and sub-THz frequency bands. Most VNAs have only two ports, which restricts their ability to support multi-antenna and multi-link 6G studies. In addition, VNA-based channel sounding also requires long measurement times because frequency sweeping is slow and becomes even slower when many antenna directions or spatial positions are measured.

## 2.6 Fourier Transform and Time-Domain Conversion

The Fourier Transform (FT) is used to convert signals from the time domain to the frequency domain [15]. In the time domain, a signal is described as a function of time and after applying the FT, the same signal is represented in terms of its frequency components. This means that it decomposes a time dependent signal into a sum of sinusoidal functions with different frequencies. Many signal processing operations are easier to perform in the frequency domain. The original signal can be reconstructed using the inverse Fourier Transform (IFT). When converting between the domains, no information is lost.

For digitally sampled signals, the Discrete Fourier Transform (DFT) is used [16]. The DFT operates on a finite number of samples into its discrete frequency representation. The DFT is given by equation (2.5),

$$F(\nu) = N^{-1} \sum_{\tau=0}^{N-1} f(\tau)\, e^{-i2\pi \frac{\nu}{N}\tau} \tag{2.5}$$

where $\nu = 0, 1, \ldots, N-1$ [17]. $F(\nu)$ describes the amplitude and phase of the frequency component with frequency $\nu/N$ and the factor $N^{-1}$ make sure that $\nu = 0$ corresponds to the mean value of the signal. In many applications, the DFT is used to compute the power spectrum $|F(\nu)|^2$, or to manipulate the frequency spectrum by removing low frequencies, enhancing high frequencies, or filtering noise. After such processing, the original signal can be reconstructed using the inverse DFT, given by equation (2.6).

$$f(\tau) = \sum_{\nu=0}^{N-1} F(\nu)\, e^{i2\pi \frac{\nu}{N}\tau}. \tag{2.6}$$

Often, the Fast Fourier Transform (FFT) is used instead of the DFT, as it is a faster algorithm [17]. Rather than computing the DFT directly, which can be expensive for large data sets, the FFT divides the data into smaller parts. Once a signal is transformed to the frequency domain with the FFT, it can be reconstructed with the inverse FFT (IFFT), analogous to the inverse DFT.

## 2.7 Windowing

Windowing is a technique commonly used in digital signal processing [18]. It uses a function called a window function to truncate a signal. Given a long sequence $x(n)$, and if $w(n)$ is a window function of length $N$, multiplying these together gives a truncated, N-point sequence $x_n(n)$, which can be seen in equation (2.7).

$$x_n[n] = x[n]w[n] \tag{2.7}$$

This is equivalent to convolution in the frequency domain. As the sampling length is not increased during truncation, the bandwidth of the signal is expanded. This, in turn, results in spectral leakage. This is remedied by choosing a window function with desired properties. Both the window function and the signal has to be considered when choosing window function, as different window functions have different effects on different signals. A commonly used window function is the Kaiser window. The Kaiser window is defined by the equation (2.8),

$$w(n) = \frac{I_0(\beta\sqrt{1 - (\frac{n-(N-1)/2}{(N-1)/2})^2}}{I_0(\beta)}, 0 \leq n \leq N-1 \tag{2.8}$$

where $I_0$ is the zeroth-order modified Bessel function of the first kind, $N$ is the window length, and $\beta$ is a parameter defining the shape of the window [19]. Increasing $\beta$ decreases the width of the main lobe, as seen in Figure 2.1.

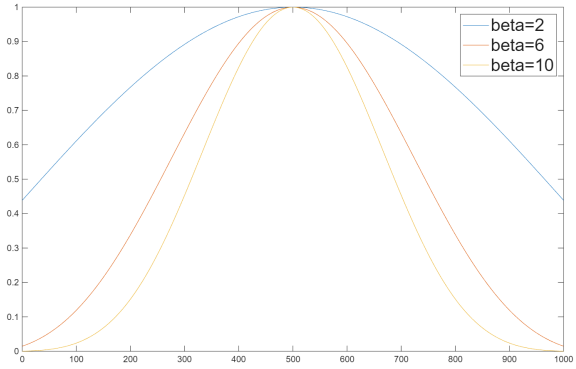



**Figure 2.1:** Three Kaiser window functions with different $\beta$; blue for $\beta = 2$, orange for $\beta = 6$ and yellow for $\beta = 10$.

## 2.8 MUSIC algorithm

To amplify the signal data of the $S_{21}$ parameter data, the algorithm Multiple Signal Classification (MUSIC) can be used [20][21]. The algorithm uses the $S_{21}$ parameter to derive a covariance matrix and linear algebra to find the eigenvalues of the matrix. The eigenvalues then get sorted from highest to lowest and are separated into a signal subspace and a noise subspace, where the amount of eigenvalues in the signal subspace is decided by the number of sources set. From there, the directions of arrival (DoA) are estimated and plotted as a MUSIC pseudospectrum.

### 2.8.1 Direction of Arrival and MUSIC resolution

The DoA is dependent on certain parameters. Firstly, the receiving antenna needs to be a uniform linear antenna array and isotropic [21]. Being a uniform linear antenna array means having a row of antennas with the same characteristics and being isotropic means receiving waves equally from all directions.

Other parameters are the number of sources set, the number of antenna elements M in the uniform linear antenna array, as well as having the optimal spacing of $d \leq \frac{\lambda}{2}$ between them [21]. The reason for having a spacing of less than a half of a wavelength of the highest frequency is to avoid grating lobes, which are defined as having a power value similar to the main lobe [22]. This decreases accuracy.

There are some parameters that increase the resolution of the algorithm [21]. The number of array elements provides a more accurate beam but with higher computational effort. The number of snapshots also improves the resolution but needs more computational effort. A greater Signal-to-Noise Ratio (SNR) and a coherent signal also improves the resolution.

### 2.8.2 The math behind the MUSIC algorithm

For MUSIC, the input data can be represented in equation (2.9),

$$x_m(t) = \sum_{k=1}^{D} s_k(t) a_m(\theta_k) + n_m(t) \tag{2.9}$$

where $s_k(t)$ = signal strength and $a_m = e^{-j(m-1)\frac{2\pi d sin\theta_k}{\lambda}}$ is the response function [21]. It can be represented as a matrix seen in equation (2.10),

$$\mathbf{X} = \mathbf{A}\mathbf{s}(\mathbf{t}) + \mathbf{n}(\mathbf{t}) \tag{2.10}$$

where $\mathbf{A} = [a(\theta_1)...a(\theta_D)]^T$ and $\mathbf{s}(\mathbf{t}) = [s_1(t), ...s_D(t)]^T$

The covariance matrix $\mathbf{R_X} = E[\mathbf{X}\mathbf{X}^H]$ can then be created with a mathematical model, where the assumption is that the signal and noise is uncorrelated with a zero noise mean, which gives equation (2.11),

$$\mathbf{R_x} = \mathbf{A}\mathbf{R}_s\mathbf{A}^H + \mathbf{RN} \tag{2.11}$$

where $\mathbf{R_s}$ is the signal correlation matrix and $\mathbf{RN} = \sigma^2\mathbf{I}$.

The eigenvalues can be retrieved from $\mathbf{R_x}$ by separating the weight attached to the eigenvalue from equation (2.12),

$$\mathbf{R_x}v_i = \lambda_i v_i \tag{2.12}$$

and then sorted from highest to lowest [21]. The number of sources specifies the number of signals in the signal subspace and sets the rest into noise subspace [20].

If the amount is unknown it needs to be estimated, which can be done by using methods such as minimum description length (MDL) [23]. However, for low SNR environments it starts to break down in finding the minimal similar sized eigenvalues, due to a shorter spacing between the eigenvalues. Instead, an algorithm can be used that orders the eigenvalues and then observes the difference between the eigenvalues and finds the maximum difference between them. Then it can subtract the index position of where the largest difference exists between two eigenvalues from the total number of antenna elements to get an estimation of the number of sources.

To find the peaks of where an object is, the MUSIC pseudospectrum from equation (2.13) is used,

$$\mathbf{P_{mu}}(\theta) = \frac{1}{\mathbf{a}^H(\theta)\mathbf{U_n}\mathbf{U_n}^H\mathbf{a}(\theta)} \tag{2.13}$$

where $\mathbf{U_n} = [V_{D+1}, ...V_M]$ is represented by the basis for the noise eigenvectors when the vectors contain low eigenvalues [21].

## 2.9 Machine Learning theory

Machine learning is a subset of artificial intelligence that enables algorithms to learn from data and make predictions and decisions without being explicitly programmed [24]. To achieve this, a model needs to undergo model training. The goal of training is generalization, meaning that a model performs well not only on the data it was trained on but also on new, unseen data.

Neural network models are used in machine learning to learn and identify patterns from data by designing it with different layers [25]. One commonly used type is the convolutional neural network (CNN), which is well suited for structured, grid-like data [26].

Several types of layers can be used in a CNN [26]. The most relevant include convolutional layers, which apply filters, also called kernels, to the input that can detect patterns. Each filter produces a feature map, marking where a specific pattern occurs in the data. In addition, activation functions, such as Rectified Linear Unit (ReLU), are used to add non-linearity, allowing the model to learn complex relationships better [27]. Batch normalization is another commonly used technique that stabilizes the learning process by normalizing the inputs of each layer so they have a mean of zero and standard deviation of one [28].

To reduce dimensionality, pooling layers can be applied [26]. It helps to reduce computational costs and make the model more robust to small variations. Before passing data to the fully connected layers, flattening is used to convert multi-dimensional feature maps into a one-dimensional vector. Then, dropout is often used for regularization, which helps to reduce overfitting [29].

# 3

# Methods

The methodology was divided into tasks which were carried out in parallel. Physical measurements were conducted in the laboratory both with and without human presence, while simulations were used to support and complement the measurement data. Using the data, signal processing and machine learning algorithms were used to detect human presence based on the obtained data.

## 3.1 Physical measurements

Before conducting the measurements, a suitable method for collecting the data needed to be defined. Firstly, the positions of the transmitting and receiving antenna needed to remain fixed to ensure consistency in the measurements. To achieve this, tape was used to mark their assigned locations on the floor of the lab. Additionally, a $4 \times 4$ grid was marked on the floor using tape, where each of the 16 squares in the grid had the dimensions of $50 \times 50$ cm.

The next step was to determine the parameters of the antenna and the settings of the VNA. The transmitting and receiving antenna were both connected to the VNA using coaxial cables with N type connectors. Due to the properties of the receiving antenna array, the spacing between the elements was chosen to be approximately half a wavelength in order to avoid grating lobes. The operating frequency range for the measurements was set as 4 GHz to 16 GHz, which corresponded to an element spacing of approximately 1 cm. The $S_{11}$ parameter was measured using the VNA to verify the frequency range, as can be seen in Figure 3.1.

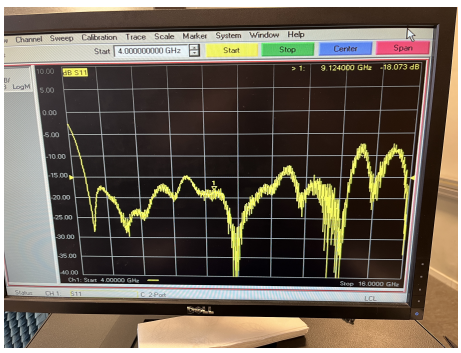



**Figure 3.1:** $S_{11}$ response of the antenna within the selected frequency range.

The receiving antenna's aperture was 200 mm, which resulted in a total of 21 elements in the virtual antenna array. The number of measurement points was then determined by considering the length of the room, the antenna aperture, and the number of array elements, all of which were constrained by predefined system constraints. This resulted in a total of 1601 measurement points. Furthermore, the intermediate frequency bandwidth was set. A value of 7 kHz was initially tested, but was later reduced to 3 kHz in order to decrease measurement noise. The power was set to 0 dBm, which provided sufficient signal strength for reliable measurements. As a first test, the transmitting antenna was positioned on the floor in order to maintain line-of-sight (LoS) between the antennas. This can be seen in Figure 3.2.

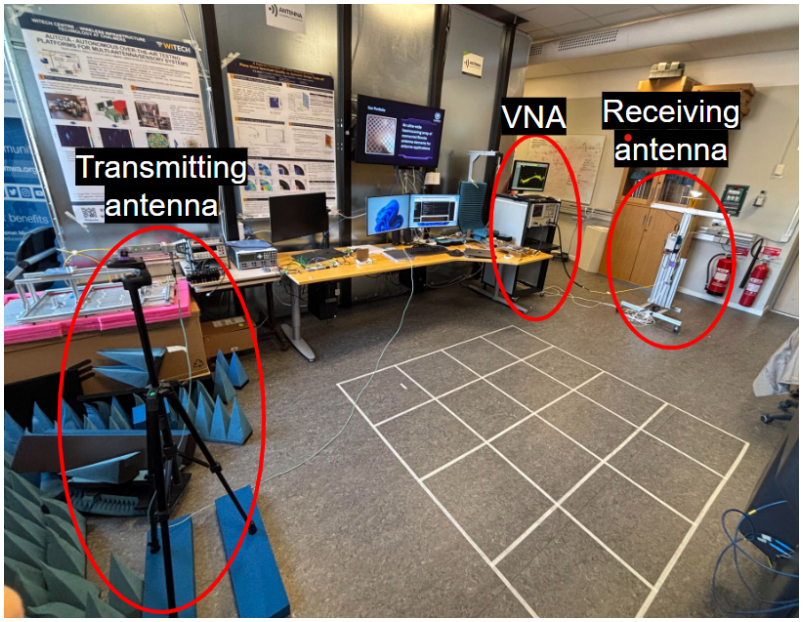



**Figure 3.2:** First set of measurements with antennas on the floor.

In Figure 3.3 it can be observed that the receiving antenna was placed more to the left of the center of the grid. This was also in order to maintain LoS, and its specific position was kept throughout all measurements. The first set of measurements were then conducted, including three measurements per square on a stationary human.

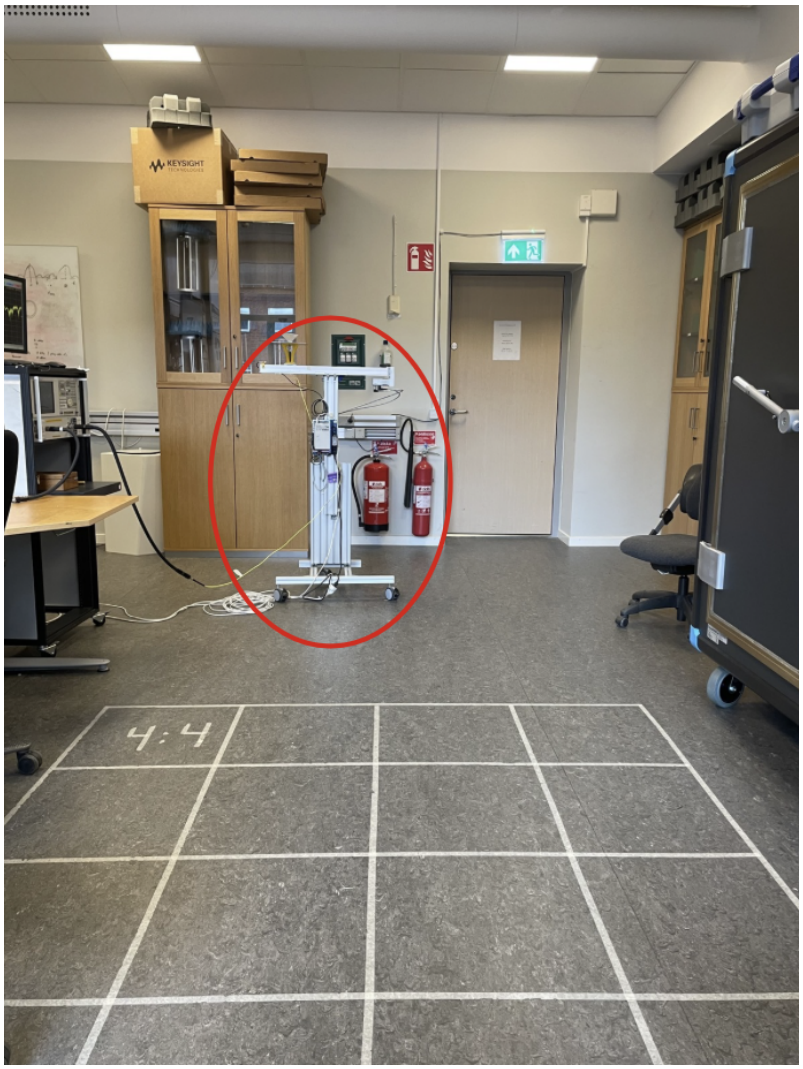



**Figure 3.3:** The position of the recieving antenna.

The same measurement procedure, including three measurements per square, was then repeated with the difference that the transmitting antenna was placed on a shelving unit. This was done to represent a more realistic human sensing scenario, where a device may be placed on a wall or the ceiling. This setup can be seen in Figure 3.4.

**Figure 3.4:** The transmitting antenna standing on a shelving unit.

In addition, measurements were conducted on an empty room with no human present. This was done to observe differences in the results, comparing the case of an empty room with that of human presence.

The distances from the receiving antenna to the center of each square were measured and are presented in Table 3.1. The distances were measured at approximately the same height as the receiving antenna, and were to be used for comparison with the results of the MUSIC algorithm. The rows and columns of the grid were defined as illustrated in Figure 3.5 from the perspective of the receiving antenna. In order to refer to these as coordinate positions on the form (x,y), the convention of writing Row as the x coordinate and Column as y coordinate was decided.

**Table 3.1:** Distance from the receiving antenna to the center of each square in cm.

| | Column 1 | Column 2 | Column 3 | Column 4 |
|---|---|---|---|---|
| Row 1 | 380 | 365 | 358 | 359 |
| Row 2 | 317 | 320 | 311 | 313 |
| Row 3 | 293 | 274 | 263 | 261 |
| Row 4 | 251 | 231 | 214 | 213 |

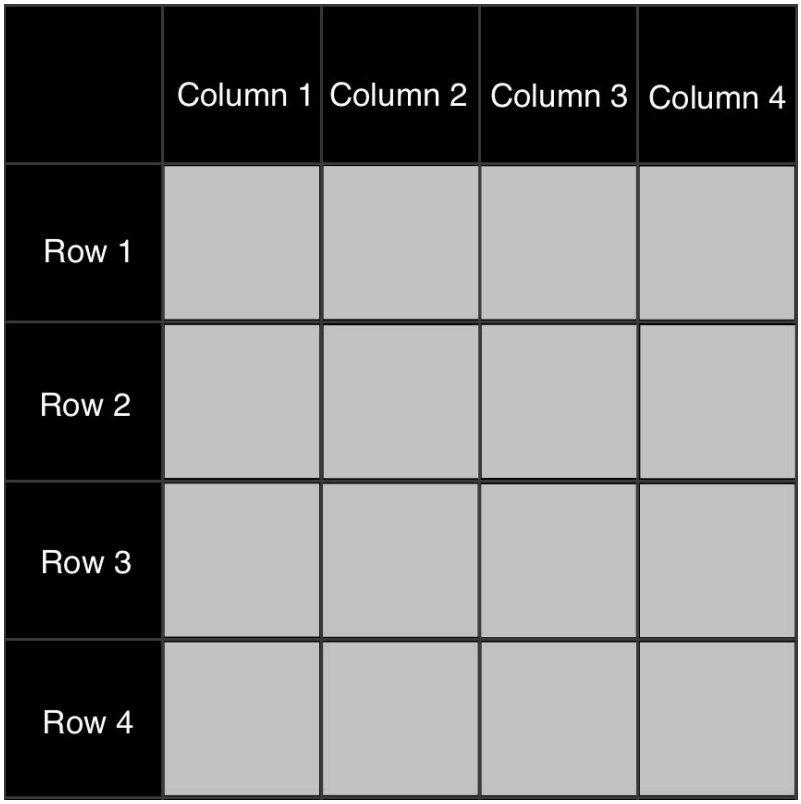



**Figure 3.5:** The grid seen from the perspective of the receiving antenna.

The transmitting antenna was kept on the shelving unit, and several more sets of measurements were conducted throughout the course of the study, all following the described setup with a stationary human in each square of the grid. When building the datasets for machine learning, thirty additional measurements were performed in each square as well as on an empty room.

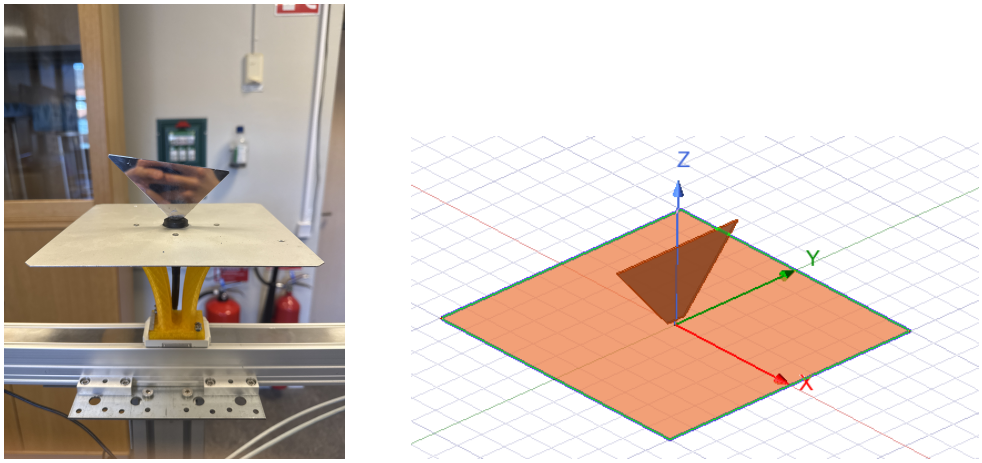
## 3.2 Simulations

Along with the physical measurements, simulated data was gathered for testing with the MUSIC algorithm and for comparison with measured data. This was done in the software ANSYS HFSS, specifically using the SBR+ solver. ANSYS HFSS is an electromagnetic simulation software made for simulating high-frequency electronic products [30]. The SBR+ solver uses shooting and bouncing ray (SBR) technology to compute how the simulated waves should bounce and scatter. Using this, the

software analyzes the radiation patterns of antennas placed in a 3D-environment and how other placed objects reflect and affect said patterns. Objects can either be created directly in HFSS using elementary shapes, or created in another software such as AutoCAD. The later option is more appropriate for complex shapes, such as furniture, while simpler objects, such as walls, can be created directly in the HFSS. As simulations were carried out to mirror the real world measurements, parameters such as frequency range and step size were set to be equal in simulation and physical data.

### 3.2.1 Antenna design

To improve the accuracy of the simulation data, a monopole trapezoidal antenna was created and modified to match its return loss with the VNA's $S_{22}$ parameter. Figure 3.6 shows the physical receiving antenna and the modeled one.




**(a)** The real antenna. **(b)** The modeled antenna.

**Figure 3.6:** Real and modeled antenna.

Because of limitations in modifying the dimensions of the model antenna, some options such as having a different port, or making the antenna thin and tilted, were not changed. The return loss was compared with the VNA $S_{22}$ parameter to make sure the model gave a similar result.

### 3.2.2 The open environment

After designing the receiving antenna (Rx), the model was imported to an open environment, where 5 bounces were used with a ray density of 2 per Min lambda, and "Fast Frequency Looping" was enabled. The initial validation was performed in a simplified open environment by introducing a single spherical object with the material perfect electrical conductor (PEC) to study basic propagation behavior. Rx antenna was placed in front of the human model and the transmitting antenna (Tx) was located behind the human model. The distance between the antennas was set to 4.25 m, where the z-coordinate for the Tx antenna was 1.25 m and the z-coordinate for the Rx was 1.41 m. Later, a human model with the property of

sea water was imported to the model instead of the metallic sphere, and the $4 \times 4$ spatial grid was defined with dimensions identical to those used in the laboratory measurements, as seen in Figure 3.7.

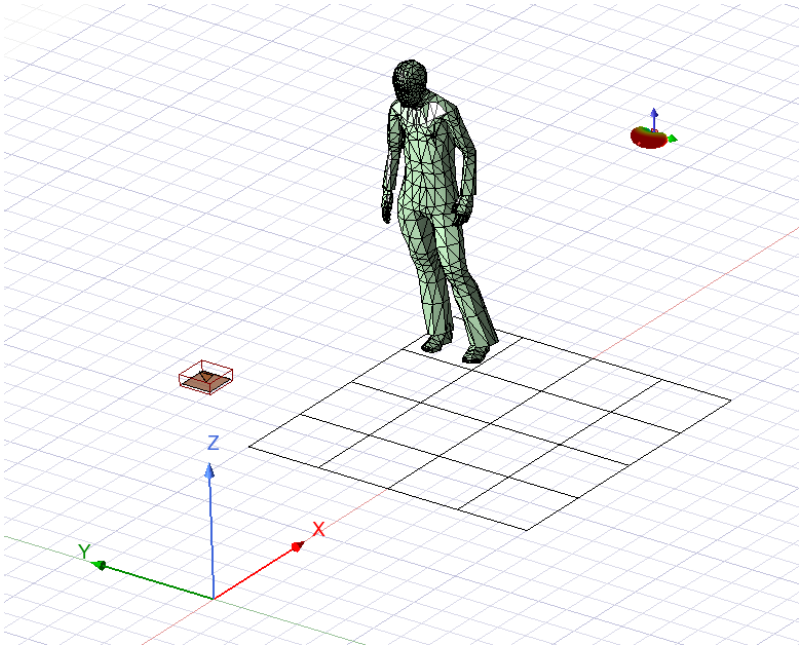
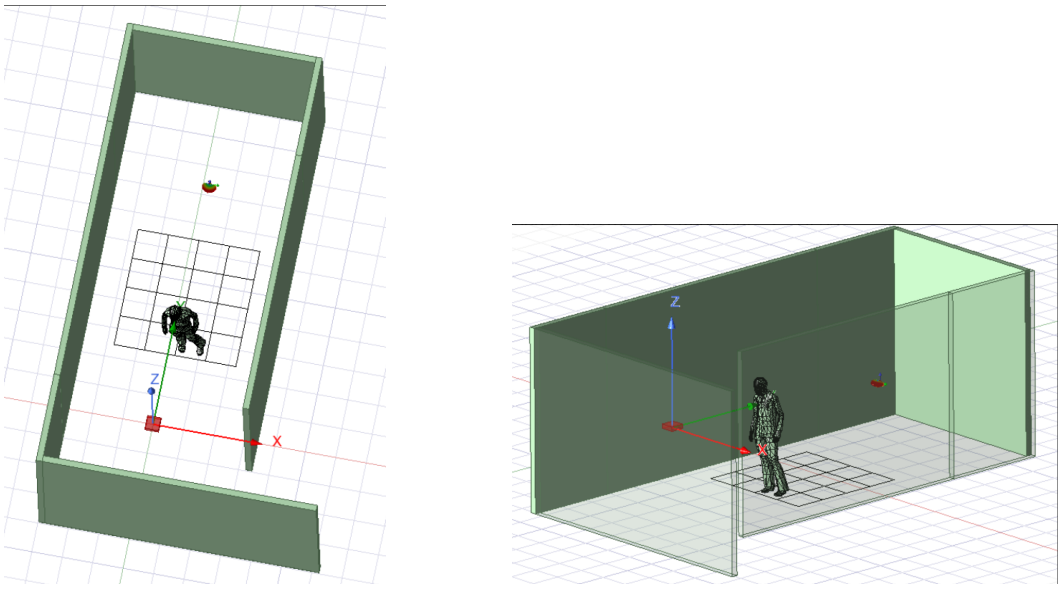



**Figure 3.7:** The open environment with the grid and the human with sea water as material.

Parametric sweeps were conducted using the Optimetrics module. The position of the human model was controlled by the variables x_human and y_human, while the Rx antenna position was defined by the variable y_pos. The Rx antenna was translated along the y-axis in 1 cm increments over 20 positions, consistent with the experimental procedure. The human model was sequentially placed at all 16 grid locations. For each configuration, simulation data were exported as magnitude and phase in radians for further analysis.

Additional scenarios were evaluated to assess environmental effects. A cylindrical object was introduced, and simulations were performed with the cylinder positioned both near and at a distance from the human model. Furthermore, measurements were conducted with only the cylinder present, positioned along the grid diagonal at (1,1), (2,2), (3,3), and (4,4).

To improve realism, surrounding walls were incorporated with dimensions matching the physical laboratory environment, as seen in Figure 3.8. The side walls were set 50 cm from the start of the grid. The wall close to the Rx antenna was located at a distance of 1.4 m, while the wall next to the Tx antenna was positioned 1.75 m away. The same measurement procedure was repeated under these conditions to enable direct comparison with experimental data.




**(a)** **(b)**

**Figure 3.8:** The environment after adding walls with asphalt as material, seen from different angles.

The center coordinates for each position in the grid in the open environment were measured based on the start position of the Rx antenna, which can be noted in Table 3.2.

**Table 3.2:** Distance from the receiving antennas first element to the center of each square with coordinate (x,y) in m.

| | Column 1 | Column 2 | Column 3 | Column 4 |
|---|---|---|---|---|
| Row 1 | (3.45; 0.75) | (3.45: 0.25) | (3.45; -0.25) | (3.45; -0.75) |
| Row 2 | (2.95; 0.75) | (2.95; 0.25) | (2.95; -0.25) | (2.95; -0.75) |
| Row 3 | (2.45; 0.75) | (2.45; 0.25) | (2.45; -0.25) | (2.45; -0.75) |
| Row 4 | (1.95; 0.75) | (1.95; 0.25) | (1.95; -0.25) | (1.95; -0.75) |

### 3.2.3 The digital twin

To help evaluate the DoA of the real environment, a digital twin was used to compare and support the data. The twin used was a modified version from a previous bachelor's thesis [31]. The Tx antenna was placed 2.63 m above ground, when standing on a shelf. The Rx antenna was placed 1.38 cm from the nearest wall and 1.41 m from the ground, in front of the human model, which gave an angle of around 12.5 degrees. The distance between the antennas was about 5.50 m LoS. The digital twin can be observed in Figure 3.9.

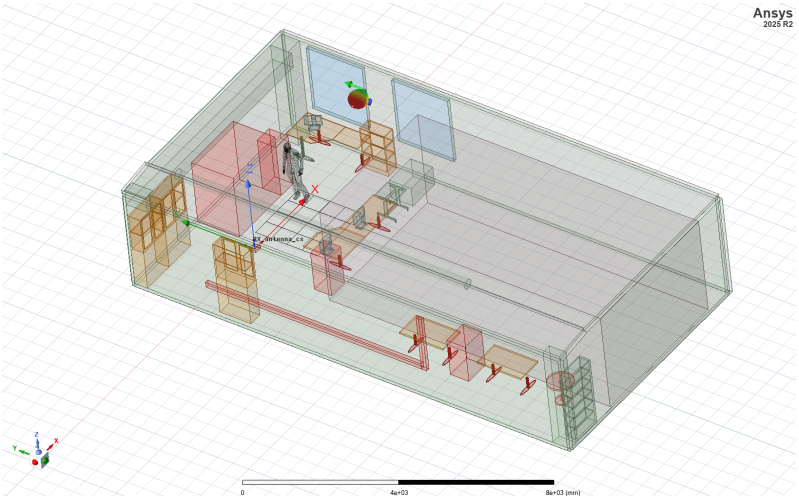



**Figure 3.9:** The digital twin, where the red and purple color objects are in material pec, the green objects are asphalt, the yellow materials are wood, and the blue objects are glass.

The antenna model was fully integrated into the digital twin, and the same Optimetrics based parametric sweeps were applied for both the human and Rx positions. Simulations were performed for all grid configurations, ensuring consistency between simulated and measured datasets.

During the HFSS simulation, Rx positions was swept in the intended spatial direction, from higher to lower positions. However, it was observed that the exported CSV data did not preserve this ordering, and instead listed the samples starting from the lowest Rx position. To ensure consistency with the physical measurement setup and the simulation sweep direction, the data was reordered in post-processing by sorting the samples according to the Rx position in descending order.

The center coordinates for each position in the grid was measured based on the start position of Rx antenna, which can be noted in Table 3.3.

**Table 3.3:** Distance from the receiving antennas first element to the center of each square with coordinate (x,y) in m.

| | Column 1 | Column 2 | Column 3 | Column 4 |
|---|---|---|---|---|
| Row 1 | (3.65; 1.30) | (3.65; 0.80) | (3.65; 0.30) | (3.65; -0.20) |
| Row 2 | (3.15; 1.30) | (3.15; 0.80) | (3.15; 0.30) | (3.15; -0.20) |
| Row 3 | (2.65; 1.30) | (2.65; 0.80) | (2.65; 0.30) | (2.65; -0.20) |
| Row 4 | (2.15; 1.30) | (2.15; 0.80) | (2.15; 0.30) | (2.15; -0.20) |

Since the model was deterministic, only one sweep was necessary. This means that the only thing that would change the data for one specific position was when other obstacles were added or moved.

### 3.2.4 Noise augmentation of digital twin data

To enable more data for machine learning and improve comparability with measured data, stochastic noise was added to the simulated data obtained from the digital twin model by writing a Matlab code. The original dataset, obtained from the digital twin, consisted of magnitude and phase values that were first manually selected, but later automatically identified based on their filenames. Both files were imported as tables while preserving the original column naming and structure. A consistency check was performed to ensure that the two datasets contained the same number of rows and a minimum number of columns, guaranteeing that each magnitude sample corresponded to a phase sample. These datasets were combined into a complex representation of the electromagnetic field.

Additive white Gaussian noise (AWGN) was then applied to the complex signal using a SNR value of 15. The value of SNR was chosen based on the SNR from the physical measurements. This approach ensures that both amplitude and phase are affected simultaneously, ensuring a physically realistic perturbation. The noise level was controlled with the help of the SNR parameter, allowing systematic variation of the measurement uncertainty. To further emulate realistic measurement imperfections, small random variations in amplitude scaling and phase offset were introduced before adding noise.

To ensure physical consistency with the measurement setup, the data were reordered such that the receiver positions followed a descending spatial sequence. This was achieved by sorting the dataset based on the Rx positions and applying the same ordering to both magnitude and phase data.

After adding noise, the noisy magnitude and phase data were written back to CSV files while preserving the original data format. The output filenames were constructed to clearly indicate the addition of noise and the realization index, while maintaining the original distinction between magnitude and phase data. The processed datasets were also saved in MATLAB format for further analysis and visualization. The Matlab code can be viewed in the drive link provided in appendix A.3.

## 3.3 Signal processing with MUSIC

To determine which data was useful, the MUSIC algorithm was used together with IFFT.

Since MUSIC requires choosing the number of sources, the algorithm for estimating the number of sources mentioned in Chapter 2.8 was used. After the eigenvalues were retrieved from the covariance matrix they were sorted in descending order. The eigenvalues were then compared to find the largest difference between two eigenvalues. Then the estimated number of sources were determined by subtracting the 21 antenna elements from the index in the ordered eigenvalues. This index was determined from where the two eigenvalues had the largest difference.

For the open environment, the number of sources was known only when the human, with or without a cylinder, was present. This was used to test the algorithm. When testing the algorithm, it was concluded that the simulation takes one more source into account. For this reason, if there was only one human in the open environment, it provided two number of sources instead of three. The same pattern was also present in some of the post-processing when the cylinder was added into the environment. Therefore, the algorithm was modified so that it produced one more than the calculated value, which also guaranteed that all of the signal was accounted for instead of risking losing one in the noise subspace.

When implementing it into the code, some if-statements were also incorporated, along with the three largest differences to obtain more estimations to compare with. The if-statements were used to remove the bug when the algorithm gave two number of sources on the more complex environments due to the big differences the largest eigenvalue got from LoS. Hence, the storage of the second and third largest differences was also used to confirm that the number of sources was greater than four on the digital twin and the physical environment.

Sub-bands were also implemented, where the code provided the possibility to choose the length of the sub-band and the number of sources. For an effective comparison between different sub-bands, 1 GHz was chosen. To determine how a different number of sources affected the resolution, testing of the number of sources between 7 and 19 was conducted for position (1,1) on both the digital twins data and the physical environments data. The reason for this was the assumption that its position had less reflective materials and would therefore improve the likelihood of detecting a human. When the lowest possible number of sources was found to get some detection that aligned with the approximated distance, a larger number of sources was used to compare the number of sources affect. Post-processing on position (4,4) was also done with the same procedure of finding the lowest number of sources with some resolution before using a larger number for comparison. The reason for only choosing positions (1,1) and (4,4) was because they were furthest apart from each other.

To get a sense of the distance, the data were first windowed with Kaiser using a $\beta$ before going through IFFT and then multiplied by the speed of light to get the distance. The plot was also normalized to the actual distance by shifting the LoS peak to match up around 5.6 m. The plot was decided to be between 5-7 m due to the fact that the longest path between the antennas and the human was under 7 m. The choice of using Kaiser was mainly due to optimization possibilities. Different $\beta$ values were tested before setting the simulated digital twin data to $\beta = 5$ and the physical data to $\beta = 6$. The code and the data used can be found in appendix A.3, where the MUSIC algorithm is inspired by the code from a previous paper [21].

## 3.4 Implementation of machine learning

To create suitable conditions for training the model, both real and simulated data were discussed and analyzed. The real measurement data reflected true real-world

conditions, and therefore contained irregular and unpredictable noise patterns. Due to this, simulated data was expected to provide more controlled conditions and potentially improve the accuracy of human detection.

The primary software tools were Visual Studio Code (VSCode) together with PyTorch's library, directly implemented in VSCode as an extension. ChatGPT was consulted for guidance on data labeling, splitting, and class organization. For both of the real and simulated datasets, a DataSet object was used to separate the complex $S_{21}$ parameter into a real and imaginary part. The full dataset was divided into 70% training data, 15% validation data, and 15% test data using a stratified split to ensure even split of labels across each dataset. A normalizer function was only applied on the training data, since it would cause data leakage applying it to the other two sets. The split and classified data was then used as input to the neural network.

The neural network was designed as a convolutional neural network and can be seen in Figure 3.10 since this type of architecture is well suited for learning patterns and spatial relationships in structured, grid-like data, which the input consisted of. The input consisted of two channels representing the real and imaginary parts of the measured signal.

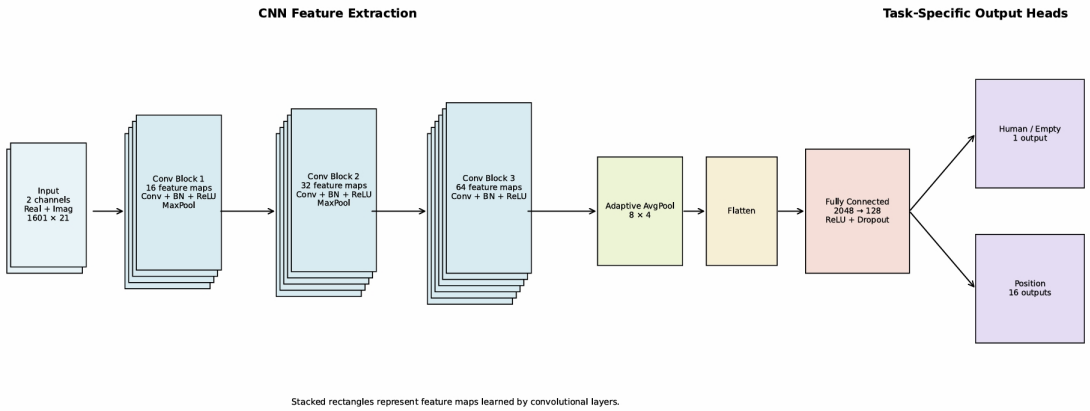



**Figure 3.10:** The convolutional neural network.

The first part of the network consisted of three convolutional blocks, where each block included a convolutional layer followed by batch normalization and a ReLU activation function. The three layers were chosen to provide sufficient depth for learning both simple and more complex spatial features, while still keeping the network relatively lightweight. Since the amount of real data measurement data was limited, using a much deeper architecture could increase the risk of overfitting, where the model memorizes the training data instead of learning patterns that generalize to unseen data.

The first convolutional layer used a larger kernel size to capture broader spatial structures and global relationships in the input data, while the following convolutional layers used smaller kernels to focus on more local and detiled features. Throughout

the network, the number of feature maps increased, allowing the network to gradually learn more complex spatial features from the input data. Batch normalization was applied after each convolutional layer to stabilize the learning process, while the ReLU activation function introduced non-linearity for better computational performance.

The first two convolutional blocks also included max pooling to reduce the spatial dimensions, which helped the model find prominent features. However, to preserve spatial information, max pooling was not applied after the third convolutional layer, since excessive downsampling could remove important information needed for the localization of the human.

Later on, an adaptive average pooling layer reduced the feature maps to a fixed size. This ensured that the fully connected layer always received the same number of input features. Since fully connected layers required a fixed input size, adaptive pooling made the network robust to variations in earlier layers and simplified the overall architecture.

Then the pooled feature maps were flattened and passed through a fully connected layer with 128 neurons, which combined the learned spatial features into a global representation. A ReLU activation function was again applied together with a dropout to reduce overfitting and improve generalization.

After the fully connected layers, the network split into two output branches. The first branch performed binary classification to determine whether a human was present or not, which was stored as a metric called human/empty. The second branch performed multi-class classification to estimate which position the human was in if a human was present, and was shown as the position accuracy metric. In addition, a third evaluation metric called the average distance metric was used, which was the average euclidean distance between the predicted and true position.

During training, loss functions were also calculated to evaluate the performance of the metrics human/empty, position accuracy, and average distance. This was later combined into a metric expressing the total loss. All of the metrics were evaluated for different combinations of training epochs when training on simulated data followed by fine-tuning on real measurement data, as well as for training using only real measured data and for training on a mixed dataset consisting of both simulated and real data.

When training, different number of epochs were also evaluated together with different learning rates. The number of epochs determined the amount of training iterations, while the learning rate controlled how much the model weights could change per iteration.

## 3.5 Use of AI in the thesis

AI has mainly been used as a tool to get an introduction into the subject as well for brainstorming in different stages of the project. The AI tools ChatGPT from

OpenAI and Claude from Anthropic have also been used to create a groundwork when writing the code for the MUSIC algorithm and the machine learning which was then built on by the members of the thesis.

# 4

# Results

Based on physical measurements and simulations, this section presents results from the MUSIC algorithm and the machine learning algorithm. The results are organized to illustrate the impact of different environmental conditions and parameter choices. This section will be used to draw further conclusions about human detection. .

## 4.1 Human detection using the MUSIC algorithm

Using the digital twin of the physical environment together with the physical environment, forms a foundation to determine the possibility if a human can be detected in the MUSIC spectrum. To complement the detection method, the distance will also be plotted.

### 4.1.1 Simulation results with MUSIC

In this section the results from different environments with the MUSIC algorithm will be presented. Each plot have a x-axis that is the angle of the DoA in degrees and a y-axis that is the magnitude of DoA in dB. The frequency spectrum is between 4 to 16 GHz. The expected direction of the human is indicated by the approximated DoA as the pointed gray line. The figure of the simulated environment will also be presented again in this section.

Figure 4.1 shows the open environment used to simulate the following results, with a modeled human placed at position (1,1) in the grid. The Rx antenna is placed in front of the human and Tx antenna is located behind the human.

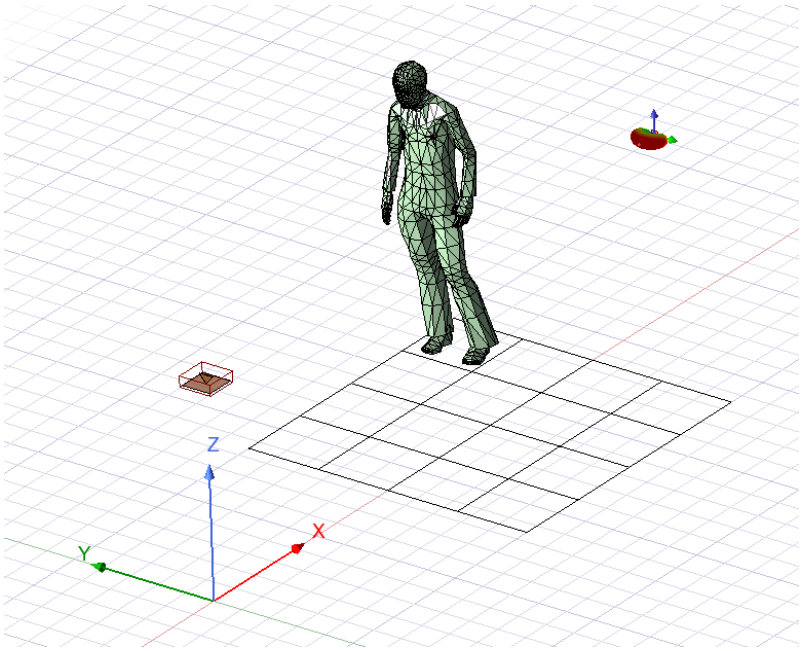



**Figure 4.1:** The open environment with the grid and the human with material sea water.

Figure 4.2 shows four plots corresponding to different positions of the human in the diagonal of the grid. The human is located at (1,1) in 4.2a, at (2,2) in 4.2b, at (3,3) in 4.2c, and at (4,4) in 4.2d. There are two spikes in each plot, where the highest one represents the LoS and the other one represents the human. The approximated DoA does not align completely with the humans position, except for 4.2d, but it is located around the same area.

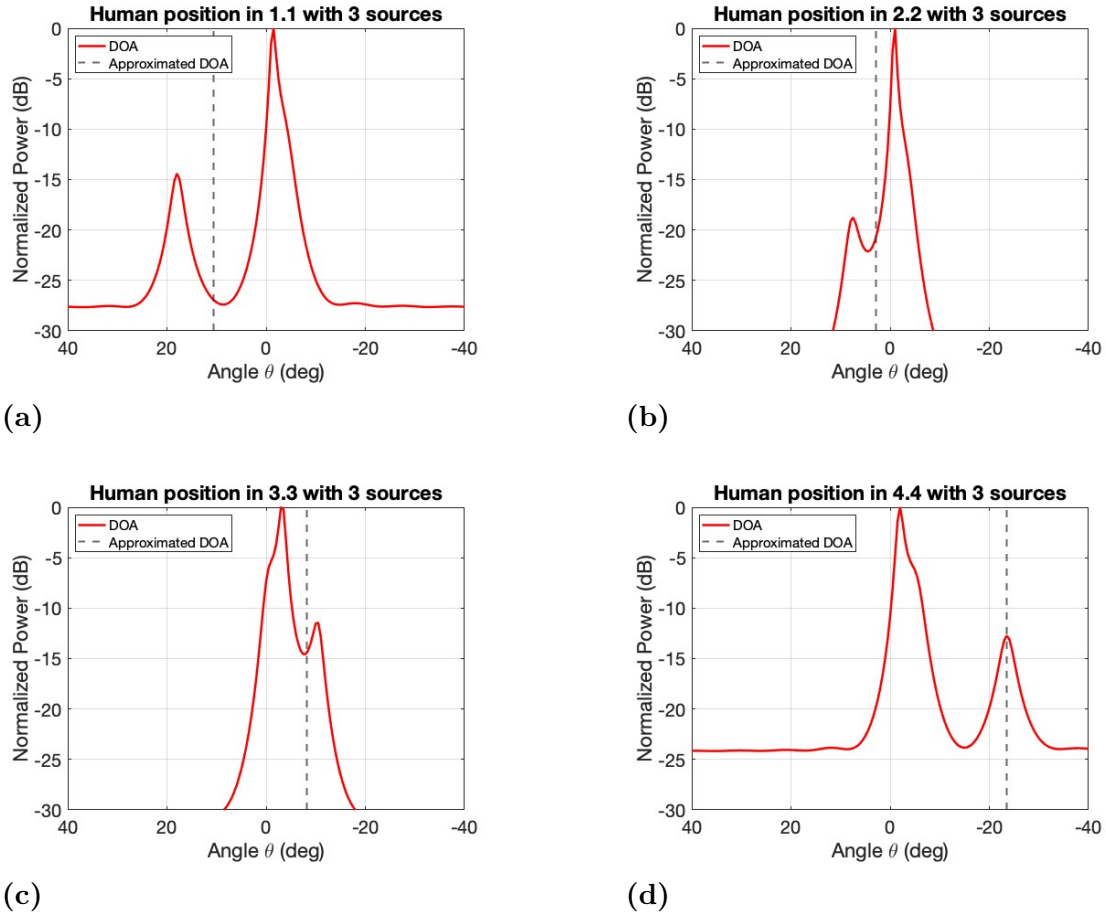



**Figure 4.2:** MUSIC plots over four different positions of the human in an open environment.

The plots in Figure 4.3 show the MUSIC spectra for two configurations in an open environment with both a human and a metallic cylinder. For each configuration, results are shown using three number of sources and four number of sources. For 4.3a the human is located at (1,1), and the cylinder at (1,4), and the spectrum has three sources. In that plot one can clearly see the LoS and a secondary peak that is perfectly aligned with the approximated DoA of the cylinder, but the peak that is close to the approximated DoA of human is extremely weak. The positions for the human and the cylinder in 4.3b are the same as in plot 1 but the number of sources are 4. It can be noted that the peak that is close to the approximated DoA of human is bigger when the number of sources increased. In 4.3c and 4.3d the human is positioned at (4.4) and the cylinder at (4,1). The number of sources are three for plot 4.3c and it can be noted that there is almost no peak at the area close to the approximated DoA of human, but there is a clear peak near the approximated DoA of the metallic cylinder. For plot 4.3d, with 4 sources, additional peaks appear in the spectrum, with responses observed near both the approximated human and cylinder directions.

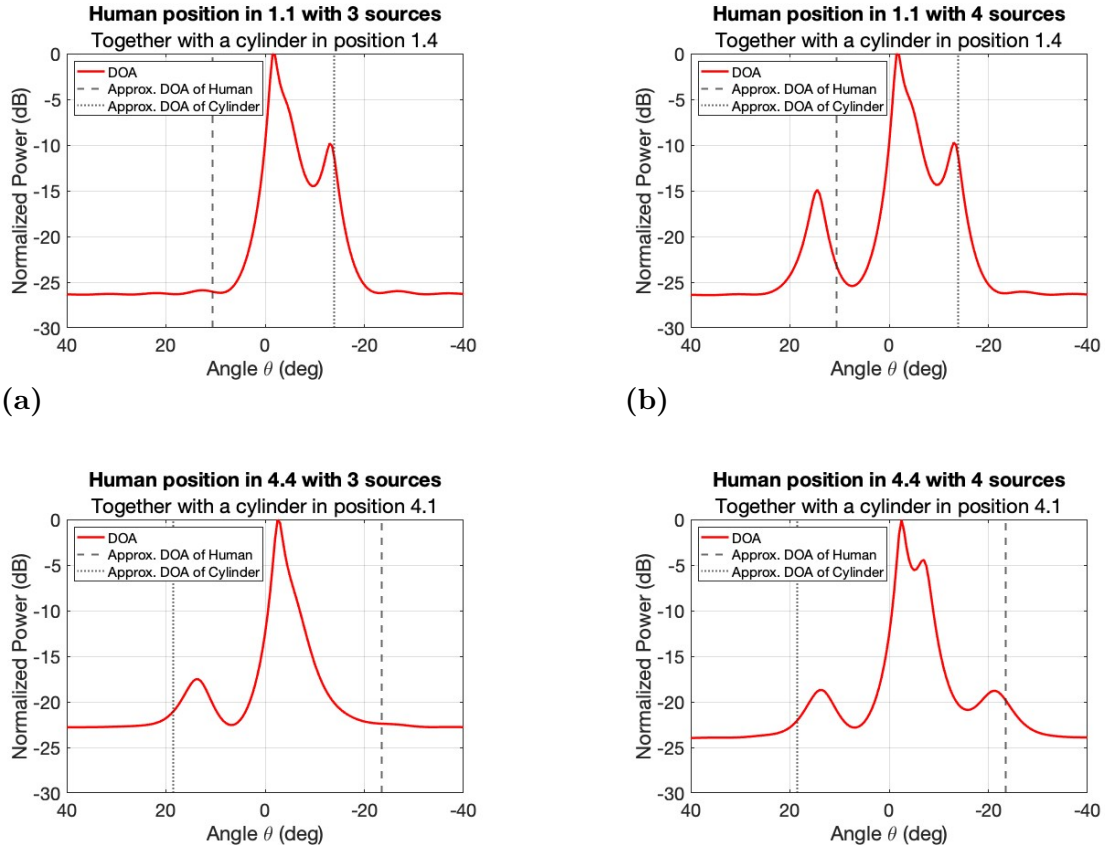



**Figure 4.3:** MUSIC plots over two different positions of the human and a metallic cylinder in an open environment, one with three sources and one with four sources.

Figure 4.4 shows the environment used to simulate the following result after adding walls, with the human placed at position (4,3) in the grid.

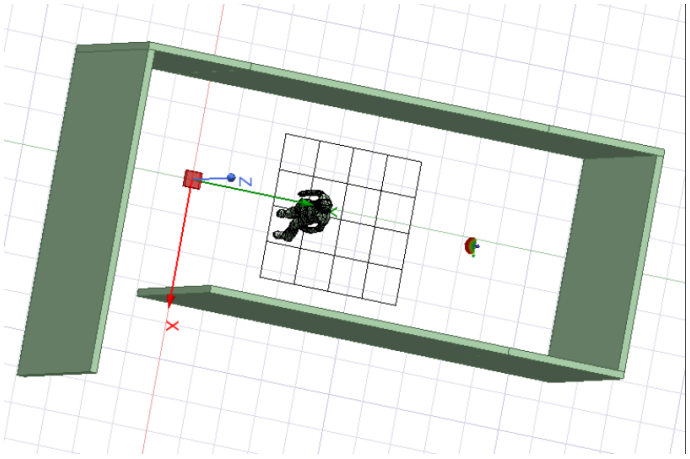


**Figure 4.4:** Picture of the environment after adding walls, with material asphalt.

The Figure 4.5 shows the MUSIC spectra for the same human positions in the diagonal of the grid, in an environment with walls. The red curve shows the DoA with the highest peak being the LoS. A secondary peak would be expected due to scattering from the human. However, across all four positions, no strong peak is observed at the approximated DoA compared to the results from the open environment without walls. For the positions in 4.5b and 4.5c, where the human is located close to the LoS direction, the main peak appears broadened and slightly shifted relative to the reference DoA. For the positions in 4.5a and 4.5d, the human is located further from the LoS, smaller secondary responses can be observed in the same area as the approximated DoA. These responses are not perfectly aligned with the reference but there are some peaks near the reference, which might indicate the human's presence.

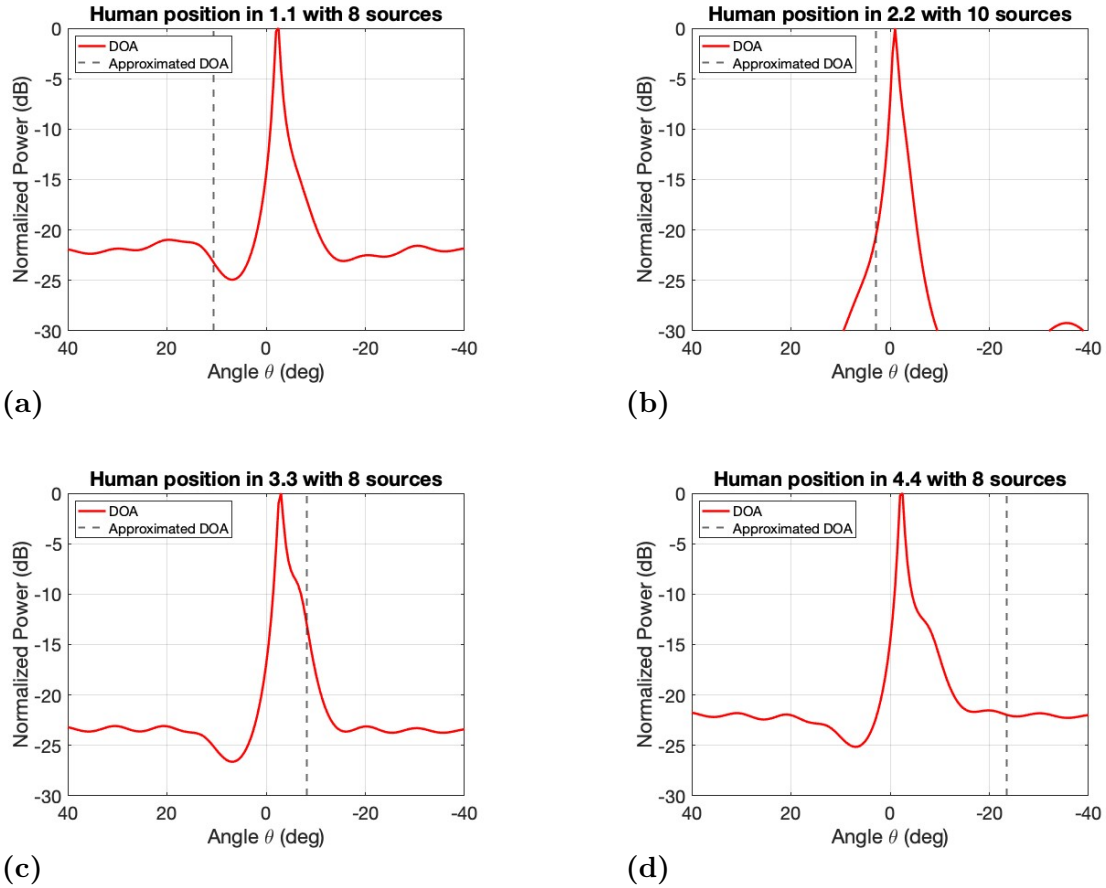
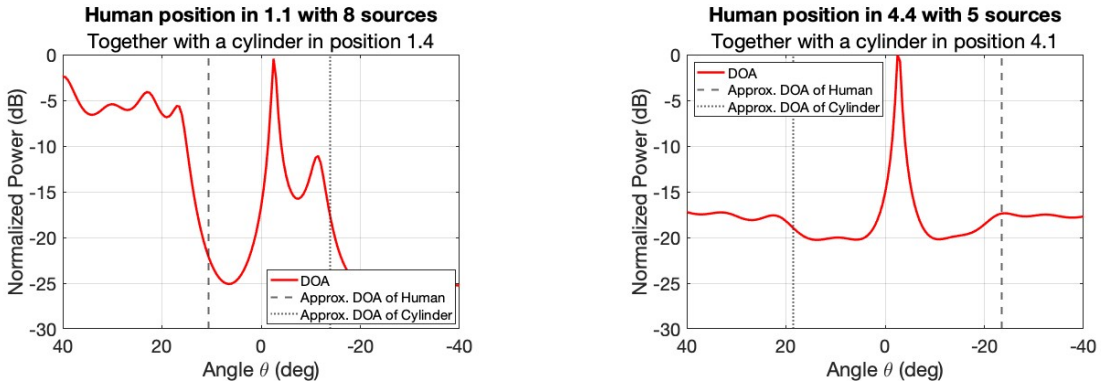



**Figure 4.5:** MUSIC plots over four different positions of the human in an environment with walls.

The plots in Figure 4.6 show the MUSIC spectra for two configurations, where both a human and a metallic cylinder are present in an environment with walls. In both plots, the dominant peak corresponds to the LoS. In 4.6a, the human is located at position (1,1) and the cylinder at (1,4). Some additional secondary responses appear that are not perfectly aligned with the reference angles. They appear in the angular regions corresponding to the approximated direction of the human and the cylinder, and are represented as gray lines. In 4.6b, the human is positioned at (4.4) and the cylinder at (4.1). The secondary responses, near the approximated DoA of the human and the cylinder, are weaker compared to 4.6a, but the lines are more aligned with the peaks, where the approximated DoA for the human is perfectly aligned with a secondary peak.




**(a)** **(b)**

**Figure 4.6:** MUSIC plots over two different positions of the human and a metallic cylinder in an environment with walls.

Figure 4.7 shows the digital twin used to simulate the following result, with the human placed at position (1,1) in the grid.

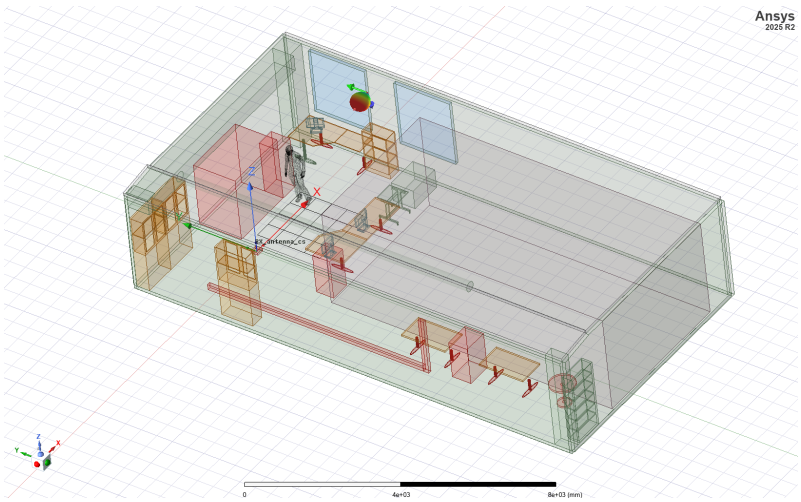
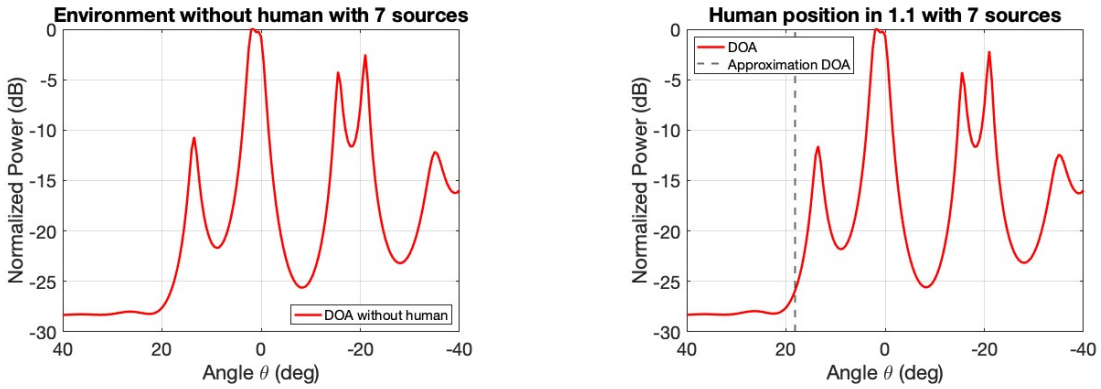



**Figure 4.7:** The digital twin, where the red and purple color objects are in material pec, the green objects are asphalt, the yellow materials are wood and the blue objects are glass.

The plots in Figure 4.8 present the MUSIC spectra obtained from the digital twin of

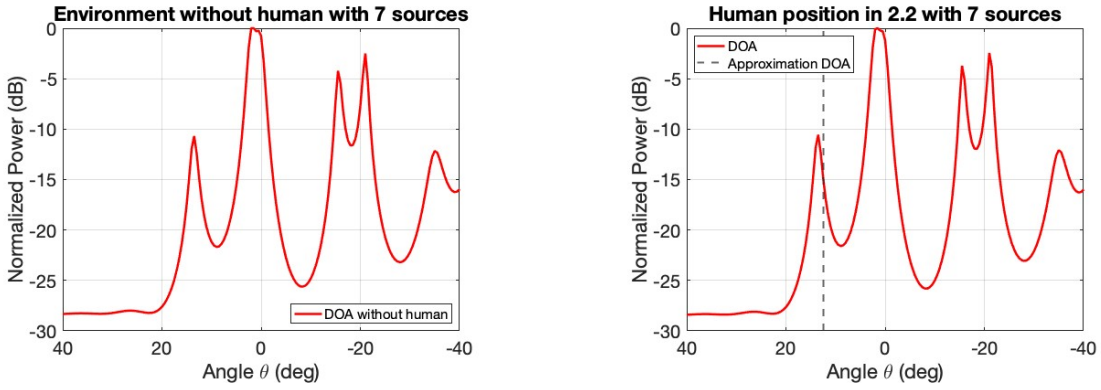
the room for two cases. Plot 4.8b shows the DoA when the human is positioned at (1,1), and 4.8a shows the DoA of the digital twin without the human. In both cases, the dominant peak corresponds to the LoS. The peak close to the approximated DoA of human has slightly smaller amplitude when the human is present, plot 4.8b, compared to 4.8a, where the human is not present.




**(a)** **(b)**

**Figure 4.8:** Two MUSIC plots of the digital twin, one when human is not present and one with the human at position (1,1).

Figure 4.9 present the MUSIC spectra obtained from the digital twin of the room, with the human positioned at (2.2) in 4.9b and without the human in 4.9a. In both cases, the highest peak corresponds to the LoS. When the human is present, and close to the LoS, the peak close to the approximated DoA of human has a slightly bigger amplitude compared to when no human is present, and is almost perfectly aligned with the approximated DoA of human.

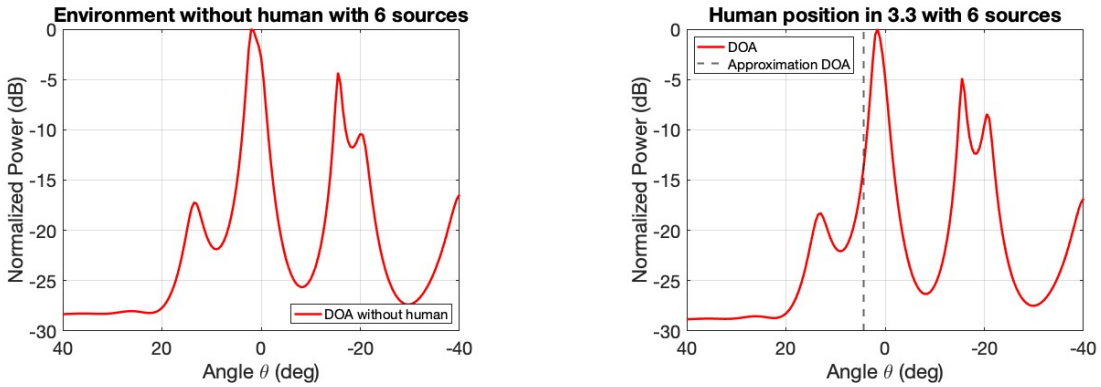



**(a)** **(b)**

**Figure 4.9:** Two MUSIC plots of the digital twin, one when human is not present and one with the human at position (2,2).

The two plots in Figure 4.10 show the MUSIC spectra obtained from the digital twin of the room, where 4.10a is without the human and 4.10b is with the human positioned at (3,3). There is a clear dominant peak in both cases, indicating the LoS. A secondary peak is present close to the approximated DoA of the human, but they do not completely align with each other. The peak that is close to the

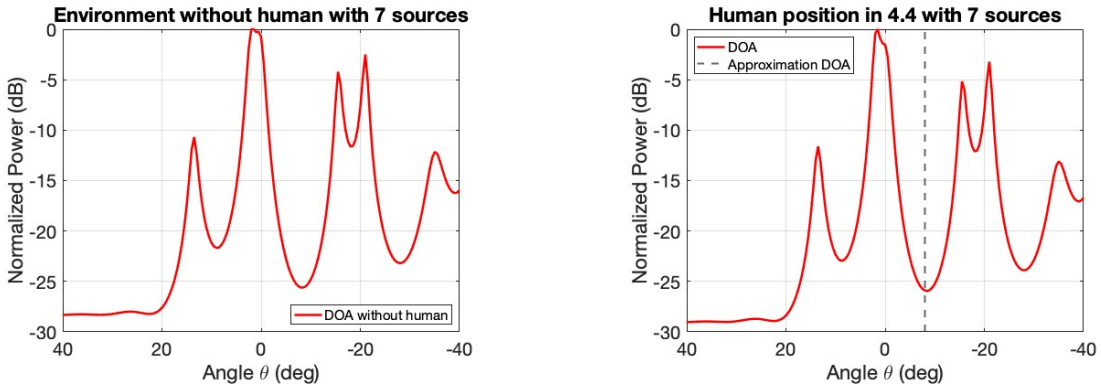
approximated DoA of human is sharper and has a slightly bigger amplitude when the human is not present, compared to the when the human is present.




**(a)** **(b)**

**Figure 4.10:** Two MUSIC plots of the digital twin, one when human is not present and one with the human at position (3,3).

Figure 4.11 present the MUSIC spectra obtained from the digital twin of the room when the human is positioned at (4.4) in plot 4.11b and without the human in 4.11a. The highest peak corresponds to the LoS, which can be noted in both plots. The approximated DoA of human does not align with any peak, but there is a slight change in the highest peak when comparing the two plots.




**(a)** **(b)**

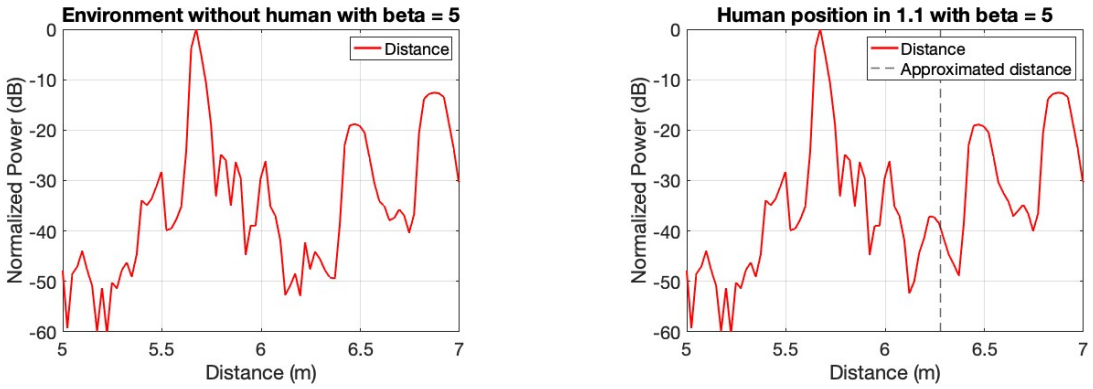
**Figure 4.11:** Two MUSIC plots of the digital twin, one when human is not present and one with the human at position (4,4).

### 4.1.2 Simulation results with IFFT

In this section, the simulated data is analyzed in the distance domain using IFFT. The IFFT enables estimation of the propagation distances associated with different signal components. The resulting plots show the normalized power in dB on the y-axis as a function of the distance in meters on the x-axis. The pointed gray line is the expected distance of the human.

Figure 4.12 shows the distance profiles using IFFT for the two cases when a human is present at position (1,1) in 4.12b, and without a human in 4.12a. In both cases there

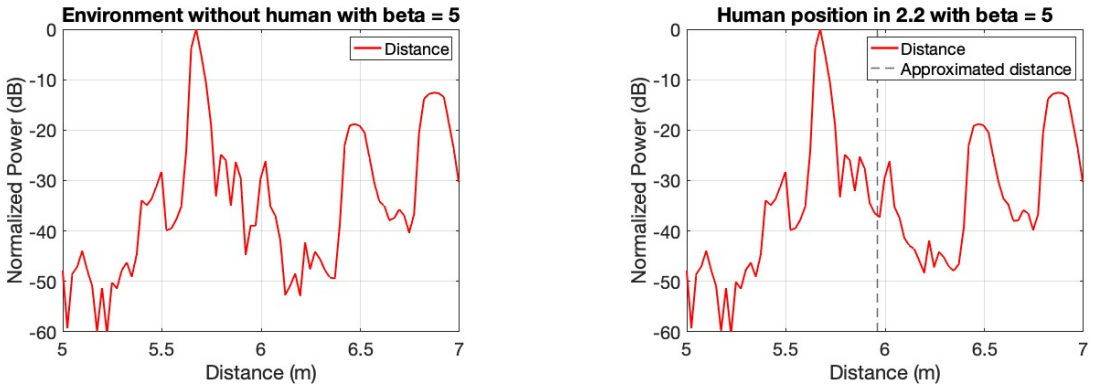
is a dominant peak at approximately 5.7 to 5.8 m, indicating the LoS. There is a slight change in the amplitude in the range between 6-6.5 m where the approximated distance is located. In 4.12b a larger amplitude can be observed in that, range compared to 4.12a, where the human is not present. One can also notice that 4.12b only has one big peak, while 4.12a has multiple small peaks.




**(a)** **(b)**

**Figure 4.12:** Two plots of IFFT on the digital twin, with one being without the human and one with human at position (1,1).

Figure 4.13 shows the IFFT-based distance profiles for the digital twin without the human in 4.13a and when the human is positioned at (2,2) in 4.13b. Similar to before the highest peak can be observed at approximately 5.6-5.7 m in both cases, corresponding to the LoS. Differences between the two cases are visible in the region between 5.8 m and 6.3 m. When the human is present, this region becomes more irregular and contains several local peaks, suggesting additional contributions to the signal. In particular, the variations around 6 m are more pronounced compared to the environment without human, which aligns with the approximated distance of the human.

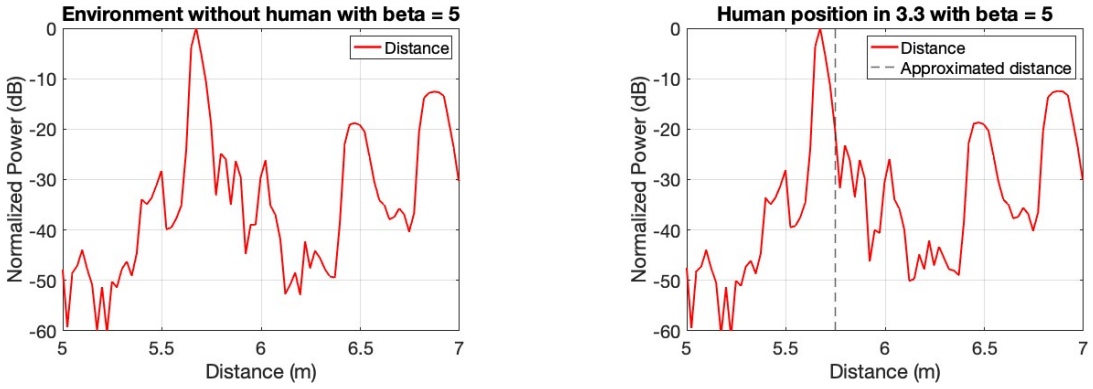



**(a)** **(b)**

**Figure 4.13:** Two plots of IFFT on the digital twin, with one being without the human and one with human at position (2,2).

The IFFT plots for the human positioned at (3,3) and for the environment without a human are shown in Figure 4.14. In both cases, the dominant peak appears around

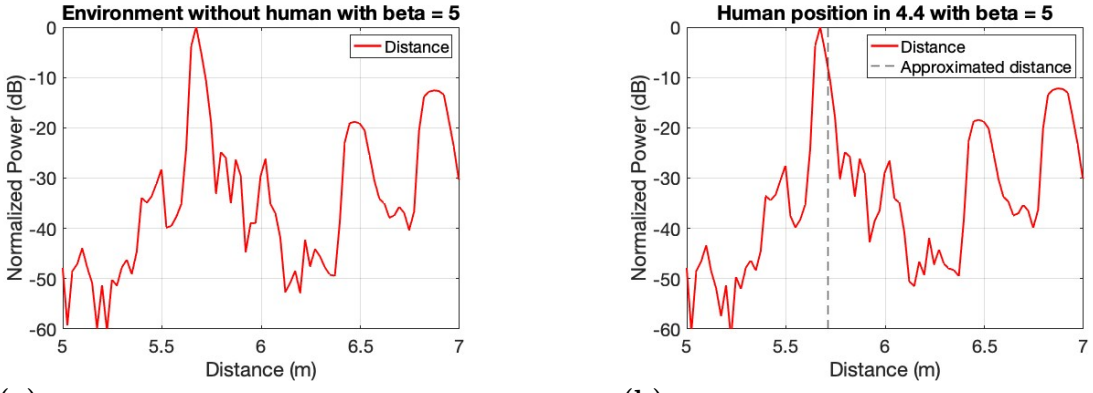
5.6 to 5.7 m like before, indicating the LoS. More noticeable differences occur in the interval between 6.2 m and 6.4 m. When the human is present in 4.14b, this region exhibits a more regular pattern, with the lowest peaks having higher amplitude compered to the environment without human in 4.14a. It is also a slight change in the amplitude of the secondary peak that is located around 5.8 m, which aligns with the approximated distance which shows the presence of the human.




**(a)** **(b)**

**Figure 4.14:** Two plots of IFFT on the digital twin, with one being without the human and one with human at position (3,3).

Figure 4.15 shows the IFFT for two cases, where 4.15a shows without the human and 4.15b shows the human placed at position (4,4). Similar to before, there is a dominant peak at approximately 5.7-5.8 m in both cases, indicating the LoS. There is a slight change in the peaks at the range between 6.1 to 6.5 m, with 4.15b having three relatively uniform small peaks in that range compared to 4.15a where the peaks are less regular and more uneven in amplitude. There is also a slight change at 5.7 m in the secondary peak next to the approximated distance, where the peak is shorter in plot 4.15b.



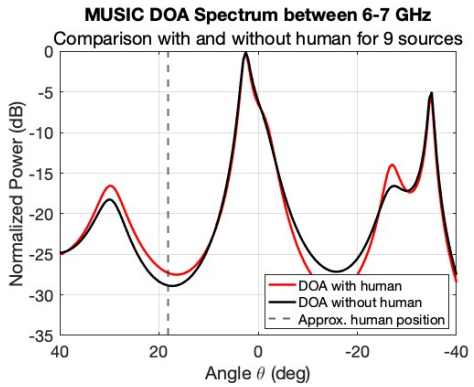

**(a)** **(b)**

**Figure 4.15:** Two plots of IFFT on the digital twin, with one being without the human and one with human at position (4,4).

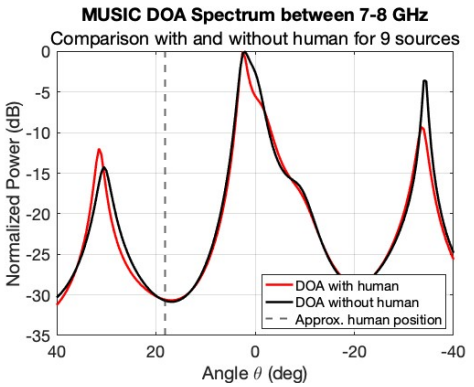
### 4.1.3 Results after dividing the frequency band into sub-bands.

This section present the results obtained from the digital twin when observing sub-bands with different number of sources. In the plots, the red curve shows the DoA when a human is present and the black curve shows the DoA obtained in the environment without a human. The gray dotted line is the approximated position of the human.

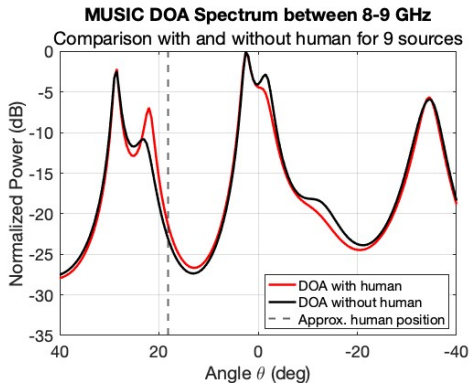
The MUSIC spectra for different frequency sub-bands between 6 and 10 GHz are shown in Figure 4.16, with the human positioned at (1,1), and 9 assumed sources. In 4.16a, the two spectra largely overlap, with only a small difference observed near the approximated human direction. The human contribution is therefore weak and difficult to distinguish in this frequency range. In 4.16b more overlap occurs with almost no difference, which also indicates that the human contribution is difficult to distinguish. The variations becomes evident in 4.16c, where there is a clear difference at the peak that is near the approximated human distance. In 4.16d, there is still a difference around the approximated human distance, but it is less strong. Among the considered sub-bands, the 8-9 GHz interval provides the most distinguishable differences between the cases with and without the human for 9 sources.



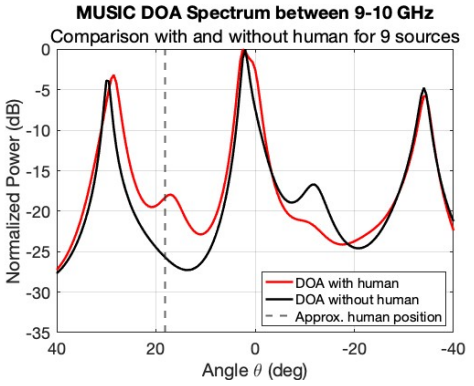

**(a)** Sub-band 6-7 GHz with 9 sources.




**(b)** Sub-band 7-8 GHz with 9 sources.



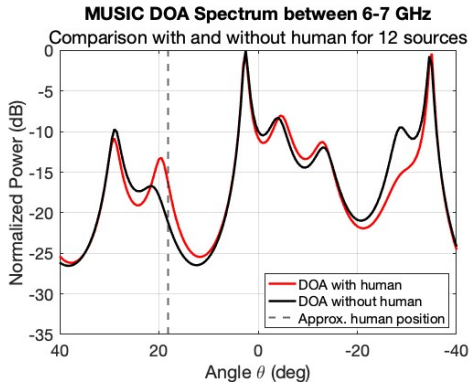

**(c)** Sub-band 8-9 GHz with 9 sources.



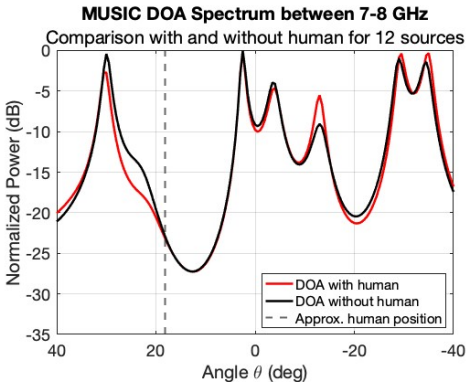

**(d)** Sub-band 9-10 GHz with 9 sources.

**Figure 4.16:** MUSIC plots with different sub-bands frequency from 6 to 10 GHz, with 9 number of sources, when the human at position (1,1).

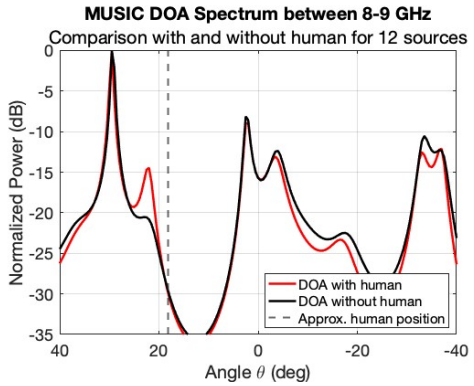
The corresponding MUSIC results for 12 assumed sources are shown in Figure 4.17 for the same case as previously, where the human is positioned at (1,1). In 4.17a, the 6-7 GHz sub-band exhibits the most pronounced difference between the two responses. A peak can clearly be observed at the approximated position of the human. In 4.17b the two spectra largely overlap, with only a small difference observed near the approximated human direction. In 4.17c the spectra becomes more complex, with multiple peaks and increased variability. There is a clear peak at the approximated DoA but the LoS, that should be around zero degrees, is not the highest peak anymore making the plot inaccurate. In 4.17d the two spectra slightly overlap and there is no clear peak at the approximated position for the human. Overall, for 12 sources, the 6-7 GHz sub-band provides the clearest plot and separation between the two cases.



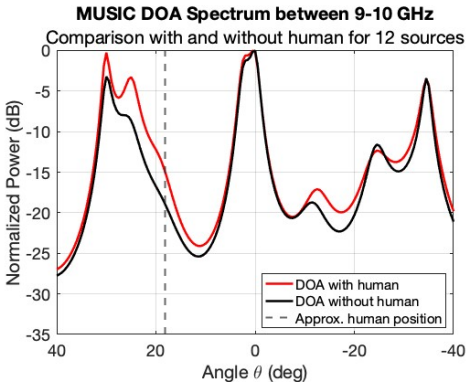

**(a)** Sub-band 6-7 GHz with 12 sources.




**(b)** Sub-band 7-8 GHz with 12 sources.




**(c)** Sub-band 8-9 GHz with 12 sources.



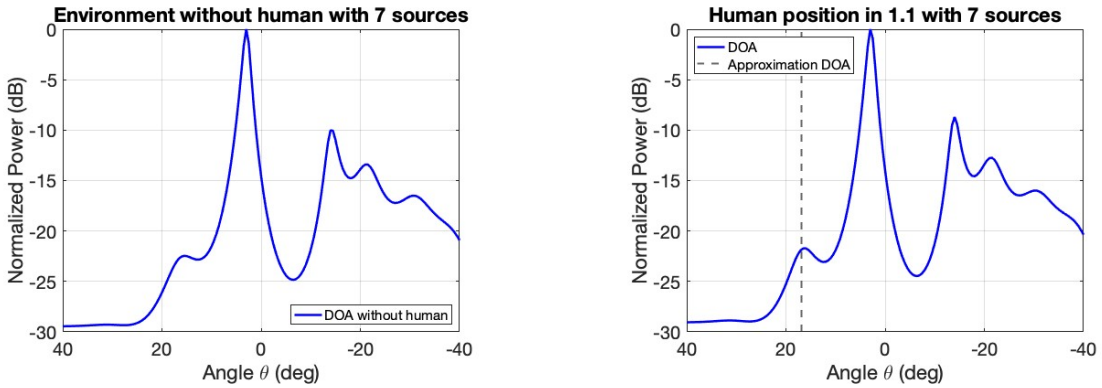

**(d)** Sub-band 9-10 GHz with 12 sources.

**Figure 4.17:** MUSIC plots with different sub-bands frequency from 6 to 10 GHz, with 12 number of sources, when the human at position (1,1).

The results for sub-bands in position (4,4), with the same number of sources as (1,1) as in Figure 4.16 and 4.17, did not give clear detection. These plots can be viewed in the appendix chapter in Figures A.1 and A.2 .

### 4.1.4 Physical results with MUSIC

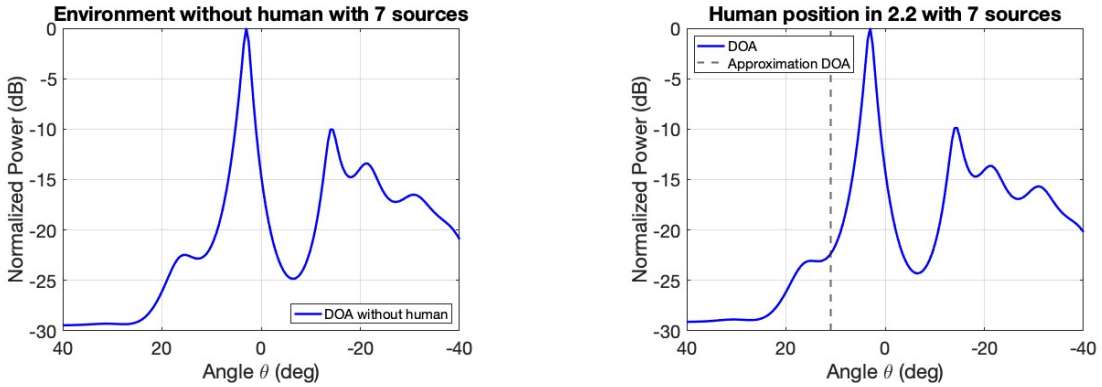
The results from the physical environment using the MUSIC algorithm are presented in this section. The positions that were observed are the positions on the diagonal between (1,1) to (4,4). Figure 4.18 shows when the human is positioned at (1,1).




**(a)** Empty environment. **(b)** Physical environment.

**Figure 4.18:** MUSIC plots when a human is present and not between 4-16 GHz.

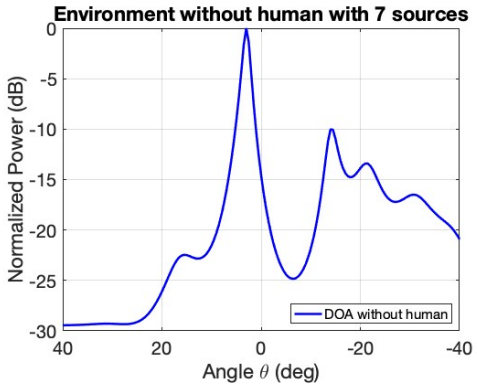
When comparing 4.18a with 4.18b, it can be observed that the same peaks are present with a slightly larger peak where the human is positioned. In addition, the peak around -15 degrees have a higher normalized power when the human is presence, which could point at the detection of a human. In Figure 4.19, the human is positioned at (2,2).



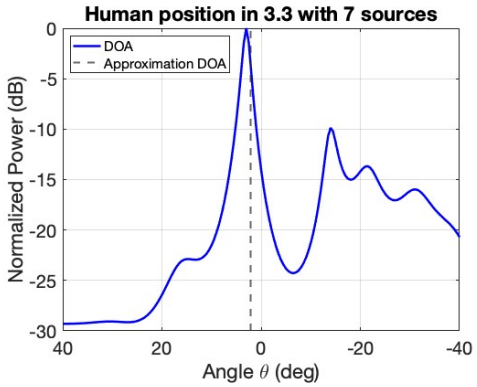

**(a)** Empty environment. **(b)** Physical environment.

**Figure 4.19:** MUSIC plots when a human is present and not between 4-16 GHz.

In this case, the human has moved closer to the LoS. A lower peak at 17 degrees can be observed and a higher peak at -30 degrees for 4.19b, compared with 4.19a. In Figure 4.20, the human has the position (3,3).



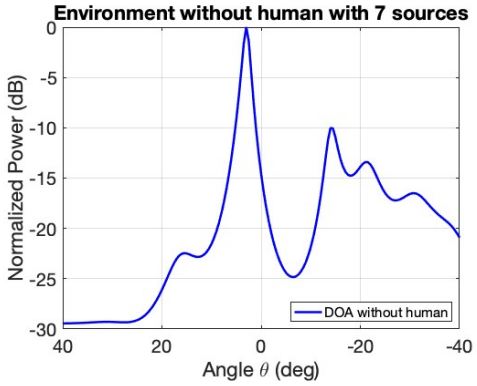

**(a)** Empty environment.



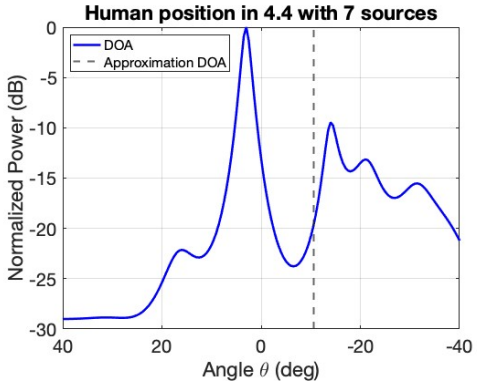

**(b)** Physical environment.

**Figure 4.20:** MUSIC plots over when a human is present and not between 4-16 GHz.

The human is now positioned under the LoS, which can be indicated by 4.20b, and similar peaks can be observed around 17 degrees and -30 degrees as in 4.19b. The last position in the diagonal, (4,4), can be seen in Figure 4.21.




**(a)** Empty environment.




**(b)** Physical environment.

**Figure 4.21:** MUSIC plots over when a human is present and not between 4-16 GHz.

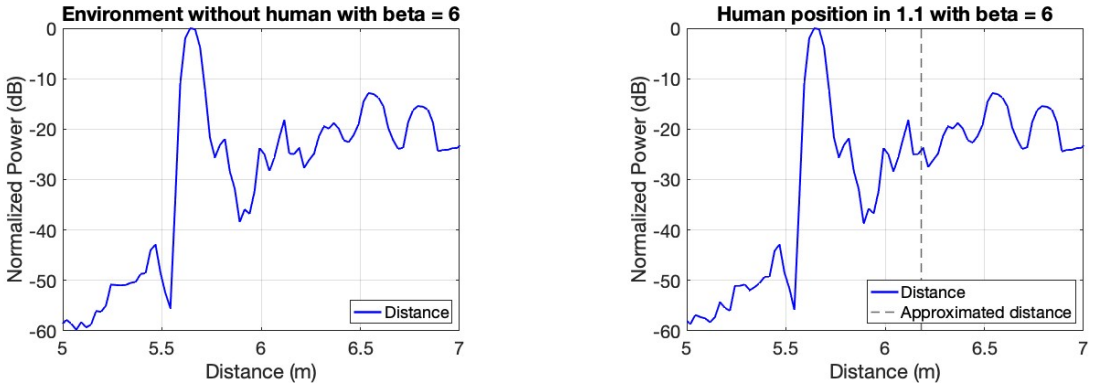
The human in 4.21b resulted in a peak at -17 degrees, compared to 4.21a. The fourth peak is also higher, which is similar to when the human was positioned at (2,2) and (3,3).

From the results for the diagonal some changes can be observed in the spectra, which indicates detection of a human. It is most noticeable at the positions (1,1) and (4,4), which are the furthest away from the LoS. It can also be observed that the positive direction of arrivals, e.g the ones that are on the left side of the LoS, have a greater impact than the right side, where the metallic anechoic chamber is present. The number of sources are stabilized around 7 sources from the implemented algorithm.

### 4.1.5 Physical results with IFFT

The distance of the human was also approximated with IFFT, and is presented for the positions on the diagonal of the grid.

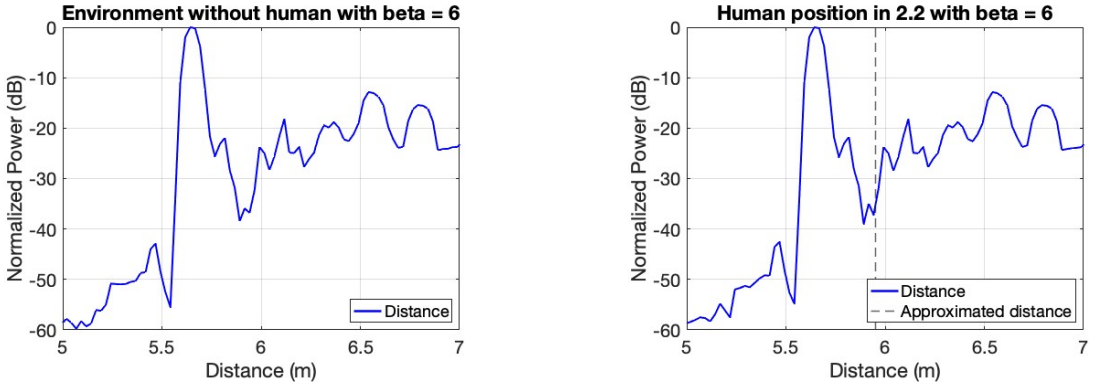
Figure 4.22 shows the human positioned at (1,1).




**(a)** **(b)**

**Figure 4.22:** Two plots of IFFT on the digital twin, with one being without the human and one with human at position (1,1).

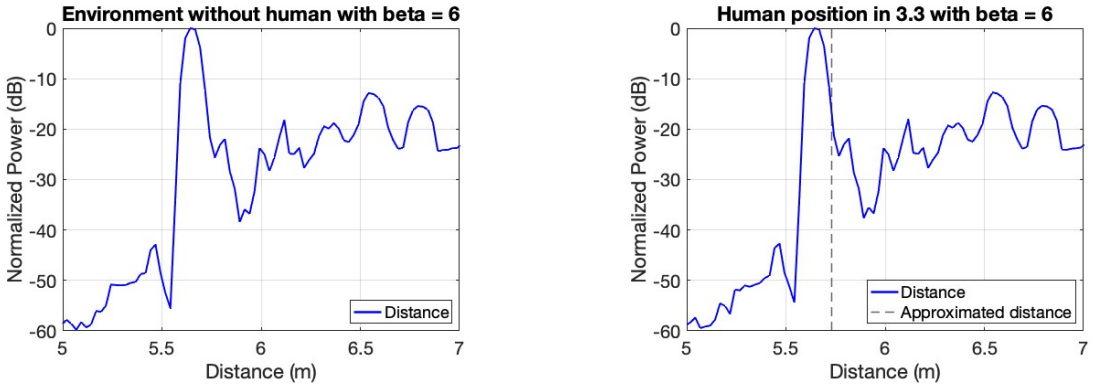
A comparison between 4.22a and 4.22b shows no significant change around the approximated distance line. However, when the human is in position (2.2), as seen in Figure 4.23, there is a change.




**(a)** **(b)**

**Figure 4.23:** Two plots of IFFT on the digital twin, with one being without the human and one with human at position (2,2).

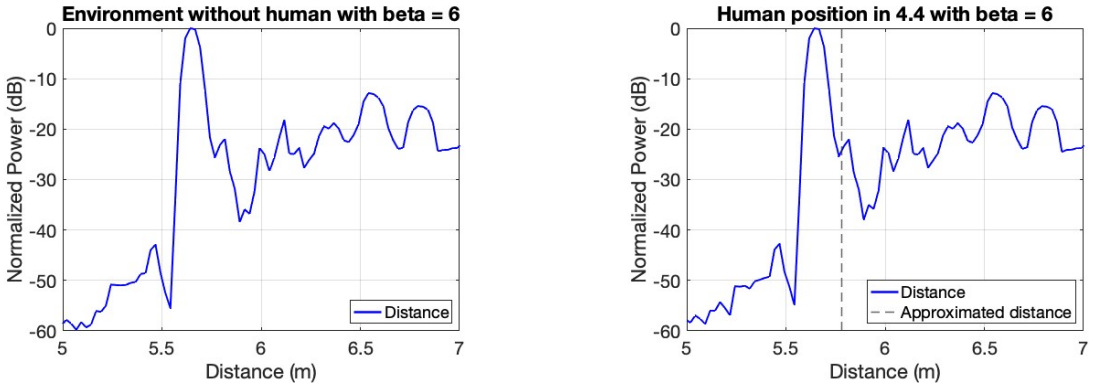
It can be observed that around the approximated distance, the normalized power are slightly stronger in 4.23b, where the human is present, compared to 4.23a, where the human is not present. When the human is in position (3,3) as seen in Figure 4.24, the result is similar to when the human is in position (1,1).




**(a)** **(b)**

**Figure 4.24:** Two plots of IFFT on the digital twin, with one being without the human and one with human at position (3,3).

When comparing 4.24a with 4.24b, there is no significant change in the normalized power. Finally, at position (4,4), the same results occur, which can be seen in Figure 4.25.



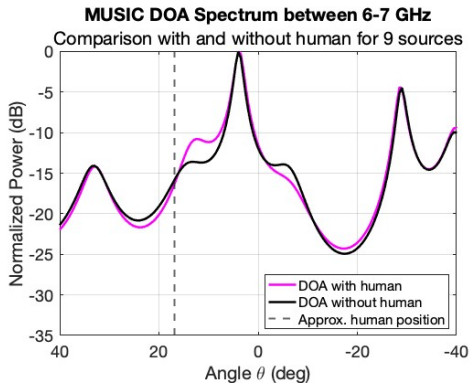

**(a)** **(b)**

**Figure 4.25:** Two plots of IFFT on the digital twin, with one being without the human and one with human at position (4,4).

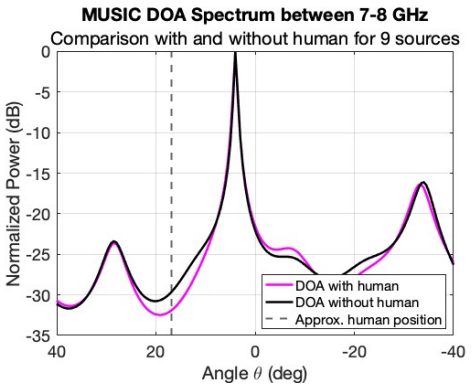
The result between 4.25a and 4.25b are too similar to deduce a difference. From the plots of position (1,1) to (4,4) it can be concluded that position (2,2) was the plot that gave a significant change, while the others stayed the same. Similar findings was also observed when changing the $\beta$ from 3 to 12.

### 4.1.6 Results after dividing the frequency band into sub-bands

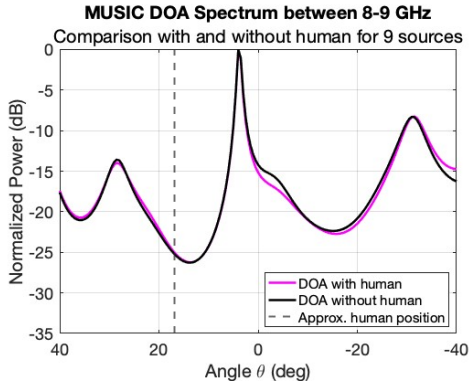
Similarly to the post-processing on the digital twin data, the data of the full bandwidth of the physical measurements are post-processed into 1 GHz sub-bands. Given the same number of sources as the sub-bands from the digital twin, the plots in Figure 4.26 are retrieved, where the magenta curve represents the MUSIC spectra for the DoA with a human, while the black curve represents the absence of a human.



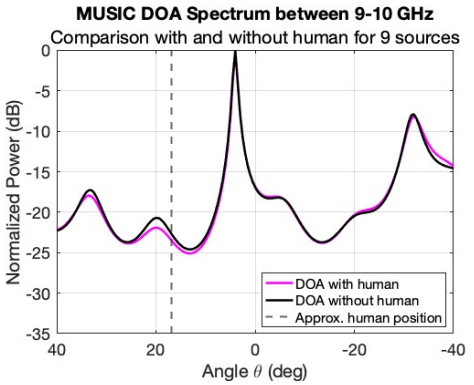

**(a)** Sub-band 6-7 GHz with 9 sources.




**(b)** Sub-band 7-8 GHz with 9 sources.




**(c)** Sub-band 8-9 GHz with 9 sources.

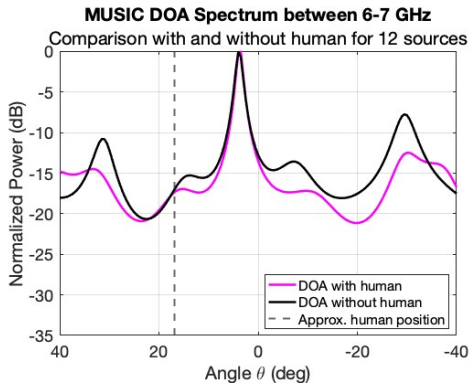



**(d)** Sub-band 9-10 GHz with 9 sources.

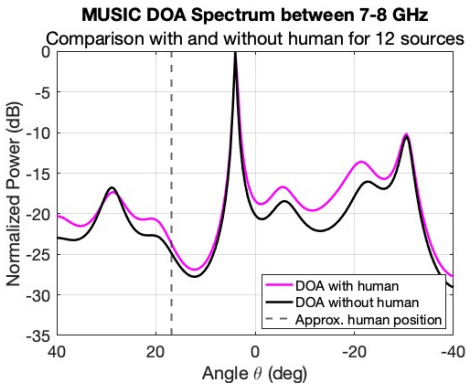
**Figure 4.26:** MUSIC plots with different sub-bands frequency from 6 to 10 GHz, with 9 number of sources, when the human at position (1,1).

A small difference can be observed in 4.26a, except around the area where the human should be present. However, since the peak is not aligned with the approximation of the human position, it could indicate that other objects in the environment are stronger. In 4.26b, no significant detection of the human's present is shown, instead the curve with the human has a lower normalized power compared to the empty environment. This could be due to the MUSIC algorithm itself. In 4.26c and 4.26d, there is no significant difference between when a human is present or not.

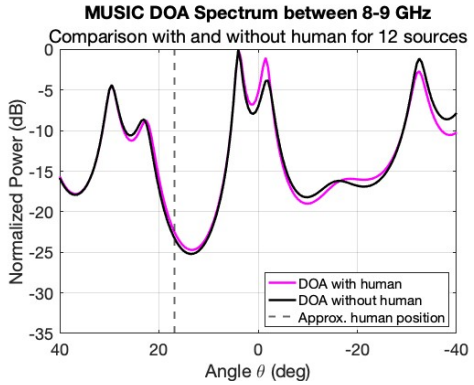
When the number of sources increased to 12 sources, the plots with a human and without a human starts to diverge more compared to when 9 number of sources were used, which can be observed in Figure 4.27.



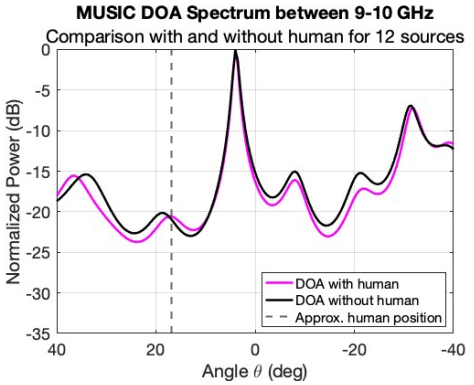

**(a)** Sub-band 6-7 GHz with 12 sources.




**(b)** Sub-band 7-8 GHz with 12 sources.




**(c)** Sub-band 8-9 GHz with 12 sources.

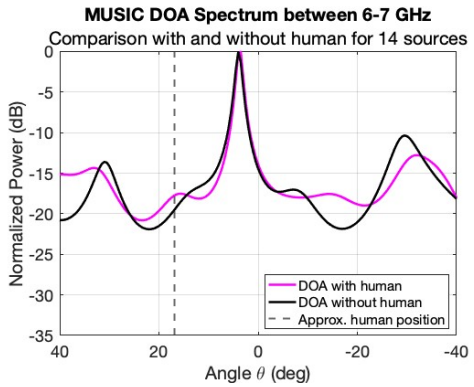



**(d)** Sub-band 9-10 GHz with 12 sources.

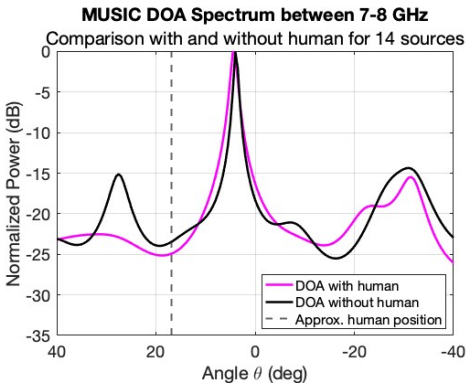
**Figure 4.27:** MUSIC plots with different sub-bands frequency from 6 to 10 GHz, with 12 number of sources, when the human at position (1,1).

When the number of sources increased to 12 sources, 4.27a was affected similarly as the sub-band for 7-8 GHz with 9 number of sources. In 4.27d, there is a peak near the estimated human position that begins to slightly differ from the physical environment. However, similar to 4.27c, the overall curve remain similar for both the cases when a human is present and not present. A greater distinction between when a human is present or not, is shown in 4.27b, but similar to 4.26a, it has a close peak, but not aligned with the approximated human position, which can be the result of other objects.

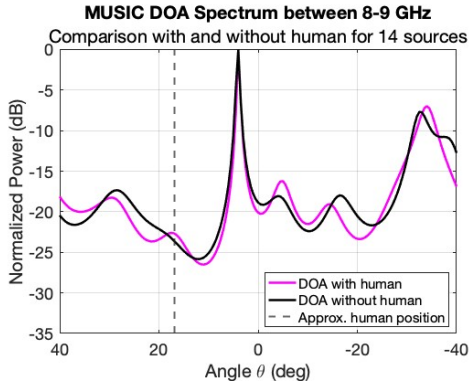
When increasing the number of sources to 14 sources the distinction between the plot with a human and without a human becomes more differentiated, which is observed in Figure 4.28.



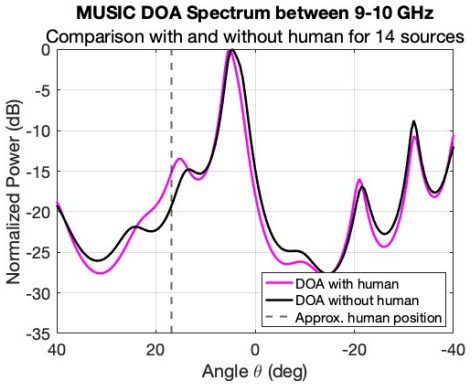

**(a)** Sub-band 6-7 GHz with 14 sources.




**(b)** Sub-band 7-8 GHz with 14 sources.




**(c)** Sub-band 8-9 GHz with 14 sources.

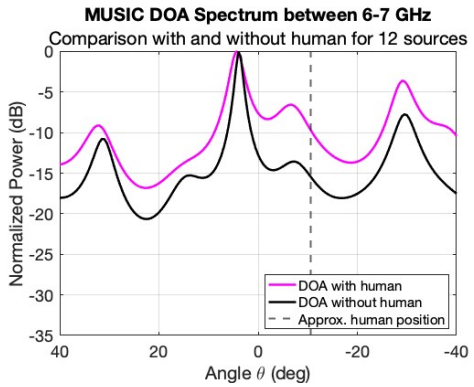



**(d)** Sub-band 9-10 GHz with 14 sources.

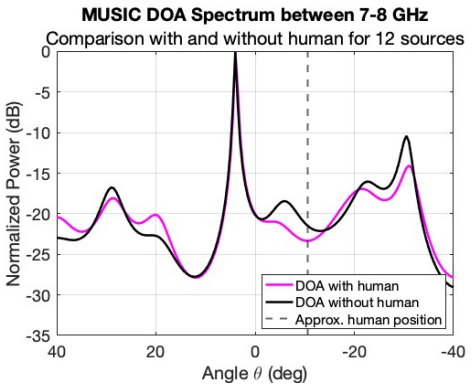
**Figure 4.28:** MUSIC plots with different sub-bands frequency from 6 to 10 GHz, with 14 number of sources, when the human at position (1,1).

From the four subfigures between 6-10 GHz, it is only 4.28b the has no peak around the approximated human position. Both 4.28a and 4.28c have a slight peak at the position where the human should approximated be. Of the four subfigures, 4.28d has the greatest difference between normalized power when a human is present and not present.

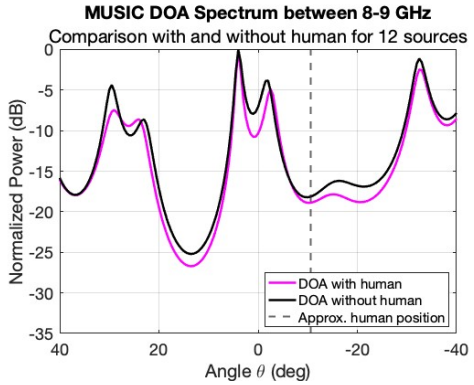
When comparing with the sub-bands for position (4,4) with 12 and 14 number of sources, it can be observed that for 12 number of sources it becomes as can be seen in Figure 4.29.



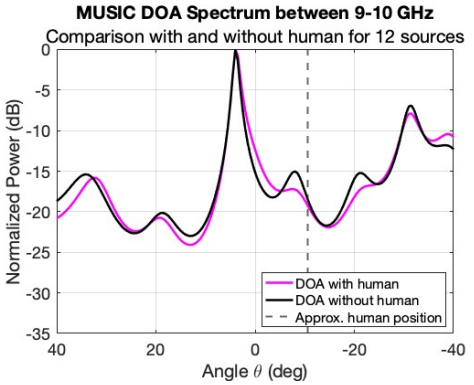

**(a)** Sub-band 6-7 GHz with 12 sources.




**(b)** Sub-band 7-8 GHz with 12 sources.



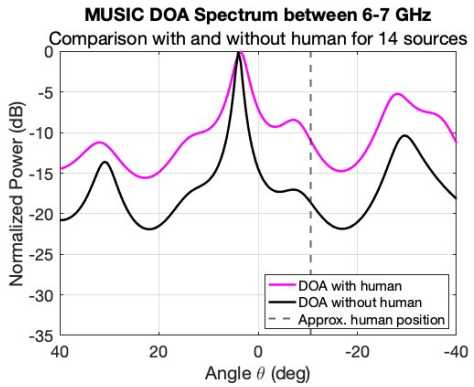

**(c)** Sub-band 8-9 GHz with 12 sources.



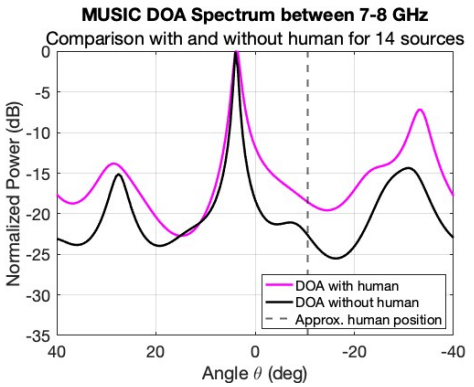

**(d)** Sub-band 9-10 GHz with 12 sources.

**Figure 4.29:** MUSIC plots with different sub-bands frequency from 6 to 10 GHz, with 12 number of sources, when the human at position (4,4).

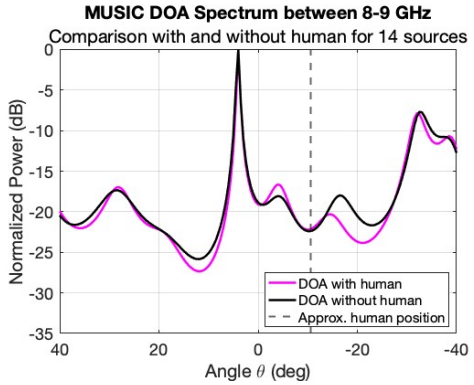
It can be observed that there is a peak around the approximated human position in 4.29a, compared to the curve with no human. However, it is not aligned. In 4.29b and 4.29c no significant peak is observed, and in 4.29d a lower peak is observed when the is human present. When increasing the number of sources to 14, the curves changes as presented in Figure 4.30.



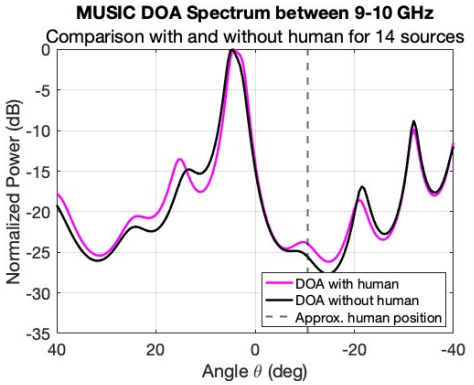

**(a)** Sub-band 6-7 GHz with 14 sources.




**(b)** Sub-band 7-8 GHz with 14 sources.




**(c)** Sub-band 8-9 GHz with 14 sources.




**(d)** Sub-band 9-10 GHz with 14 sources.

**Figure 4.30:** MUSIC plots with different sub-bands frequency from 6 to 10 GHz, with 14 number of sources, when the human at position (4,4).

In comparison to using 12 sources, it can be observed in 4.30a that there still is a significant peak around the approximated human position but still not aligned with the approximated position of the human. In 4.30b and 4.30c there is more separation between the curves. A peak can be observed in 4.30c, but since it is not on the approximated position line for the human, it could be due to other objects in the environment. In 4.30d a significant peak can be seen on the approximated line for the human position.

## 4.2 Machine learning

Different combinations of epochs with a learning rate of $5 * 10^{-4}$ for the simulated data and $10^{-3}$ for the real data resulted in the metrics seen in table 4.1.

**Table 4.1:** Comparison of performance metrics after simulated training and after fine-tuning on real data for different epoch configurations.

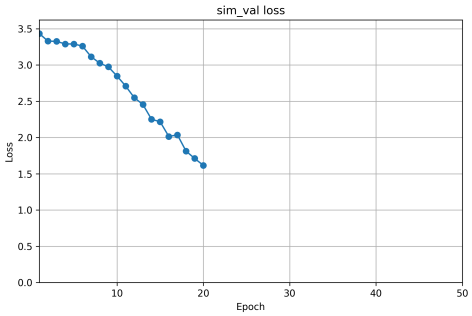
| Epochs | Human/Empty | | Position | | AvgDist | |
|---|---|---|---|---|---|---|
| | Train | Fine-tune | Train | Fine-tune | Train | Fine-tune |
| 20 sim / 50 real | 94.12% | 94.32% | 36.25% | 42.50% | 1.43 | 1.05 |
| 20 sim / 100 real | 94.12% | 98.86% | 36.25% | 43.75% | 1.43 | 1.16 |
| 50 sim / 50 real | 94.12% | 90.91% | 96.25% | 40.00% | 0.04 | 1.02 |

A mixed training set consisting of 80 epochs, with a ratio of 70% of real data, and training only on real data with 70 epochs resulted in the metrics seen in Table 4.2.

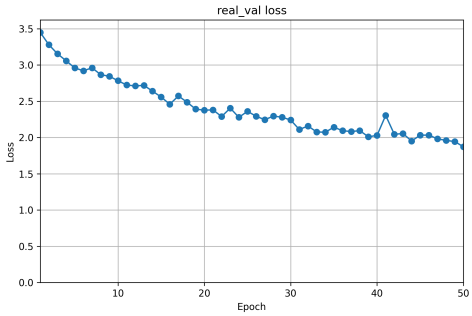
**Table 4.2:** Metrics after training on mixed data and real data.

| | Human/Empty | Position | Average distance |
|---|---|---|---|
| Mixed training set | 95.45% | 55.00% | 0.81 |
| Real training set | 98.86% | 45.00% | 1.04 |

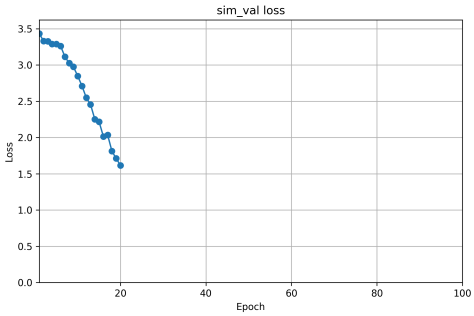
In Figure 4.31, which relates to Table 4.1, the loss from each combination of simulated and real epochs can be seen. In Figure 4.32, which relates to Table 4.2, the loss from the mixed dataset and only real dataset, can be seen.



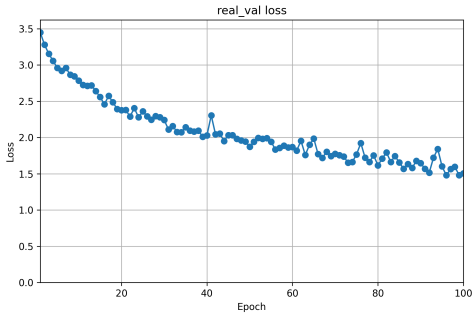

**(a)** 20/50: simulated validation loss.



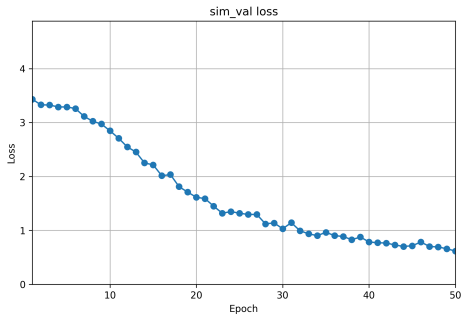

**(b)** 20/50: real validation loss.



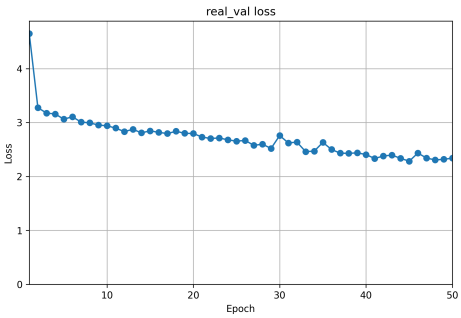

**(c)** 20/100: simulated validation loss.




**(d)** 20/100: real validation loss.

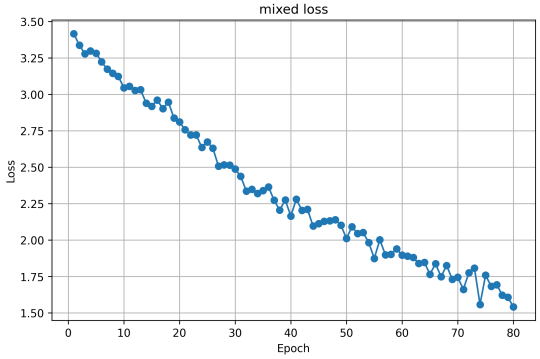



**(e)** 50/50: simulated validation loss.

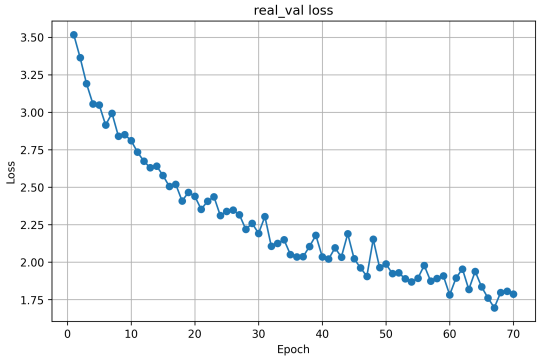



**(f)** 50/50: real validation loss.

**Figure 4.31:** Validation loss curves for different combinations of real and simulated training data.




**(a)** Mixed validation loss.




**(b)** Real validation loss.

**Figure 4.32:** Validation loss curves for the mixed dataset model and the model trained only on real data.

# 5

# Discussion

This section aims to analyze and discuss the results and methods utilized to achieve them, as well as investigate what could be improved in a future study and ethics. The result from the physical and simulated measurements with the MUSIC algorithm and the result from machine learning will be discussed and compared in this section.

## 5.1 Physical Measurements

Looking at the results from Chapter 4.1.5, using IFFT on the physical data did not provide any clear results. In Figures 4.22 to 4.25, the difference between the environment without a human and with is barely distinguishable. This is likely due to the large amount of objects in the physical environment, where nearly all of which reflect some amount of signal to the receiving antenna, which may as well be noise when trying to discern the human. This is one reason for the use of the MUSIC algorithm, as it removes the noise sub-spaces, giving a clearer result. However, the MUSIC spectrum, unlike the IFFT-plots, does not show distance. Combining these plots could lead to data that are easier to discern with the human eye. However, this was not done for the purpose of this thesis, as the neural network was deemed the more optimal way to find the human.

During the gathering of physical data, the impact of the cables became evident. At first, a faulty cable was used from the VNA to the transmitting antenna. This led to a high amount of noise and likely reflections within the cable, which became evident as post-processing was performed on the data. For example, when the IFFT of the faulty data without a human was analyzed, two large peaks were displayed at about 3.5 meters and 7 meters, as opposed to 4.22a, where a large peak is present around 5.7 meters. A large peak at 3.5 meters should in theory be impossible, as that would mean that the electromagnetic waves were traveling faster than the speed of light. In addition, the amount of noise was much greater. This clearly shows the impact of choosing the proper and non-faulty cables.

In addition to the impact of the cables, other potential sources of errors could have been present in the measuring process. The main source was interference from other electromagnetic sources in the vicinity. Aside from the equipment used for the necessary measurements for this thesis, other pieces of measurement equipment, communication devices, and other various electrical devices were present in the

laboratory. The electromagnetic energy these sources radiate could have affected the results.

Another potential source of errors would be changes in the environment being measured. As not all measurements were done in a single day, others with access to the laboratory may have moved objects in the room between measurements. Although no large pieces of furniture or similar were moved between measurements, smaller objects on tables, in shelves etc. could have been moved. Despite these objects being small they could affect the measurements, changing reflections and absorptions in the room. However, based on the results in Tables 4.1 and 4.2, a high precision was achieved despite these potential sources of errors, and the effect they had could likely be considered negligible.

## 5.2 Simulations

Looking at the results from Chapter 4.1.2, using IFFT on the simulated data gave clearer results compared to using it on the physical data. In Figure 4.12 for example, a peak very close to the approximated human distance is created when the human is introduced. This is a large contrast to IFFT used on the physical data, where the human was barely distinguishable. A probable cause for this is the absence noise in the simulation compared to the physical measurements, both due to fewer objects in the model and other sources of noise that were present in the physical twin. The fact that the IFFT was clear with the simulated data but not with the physical indicates that the data itself was the main factor for the vagueness of the results, not the IFFT itself. To remedy this, lowering the amount of noise in the physical data would be beneficial.

In Section 4.1.1 it can be seen that the MUSIC algorithm is able to differentiate the human's presence in the different positions, but with varying precision. It can be said that the simulated human produces detectable reflections to the received signal. Instead of creating its own peak, the human's presence is often depicted as a slight change of height in the smaller peaks or as a broadening of the main peak. This was somewhat expected, since the LoS signal is supposed to be stronger than human reflection. However, the chosen location of the antennas could alter the clarity of the result. It can be seen that the graphs depicting some of the locations further away from the LoS have peak alterations closer to the approximate position of the simulated human. This suggests that moving the antennas, and therefore the LoS, to the edge of the grid could make the grid positions produce more usable results.

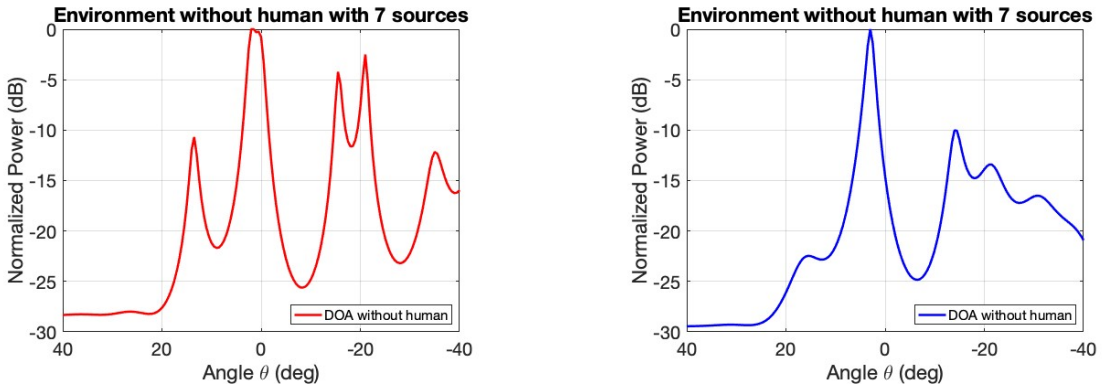
The results suggest that the number of sources are vital in the level of detail one can visualize the room in. A higher number reveals more peaks, which could help pinpoint the position of the human but also details more unnecessary peaks, which can make the main interpretation more challenging. It should be noted that further optimizing the number of sources could yield clearer results in future pursuits.

It is also worth noting the differing behavior when walls are introduced, as shown in Figure 4.5. This seems to add several reflections and multi-path propagation.

While LoS is still the most prevalent peak, the simulated human becomes even less prominent and aligned with the approximation. The changes in the nearby peaks and overall level seem to suggest the human presence.

## 5.3 MUSIC

Comparing the MUSIC algorithm's DOA estimation for physical and simulated data, they differ noticeably. For example, compare the physical lab with the simulation of the digital twin, both without any human presence in Figure 5.1.



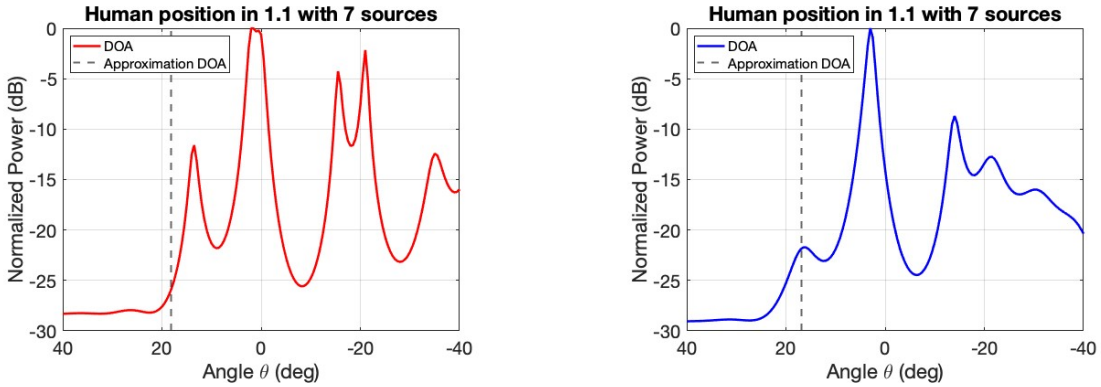

**(a)** **(b)**

**Figure 5.1:** Two MUSIC plots without a human, one from the digital twin (left) and one from the physical room (right). These are retrieved from 4.8a and 4.18a in the result section.

The MUSIC plot of the simulated data in 5.1a has a much larger variation between the peaks than the plot from physical data in 5.1b, which can be observed around 15°. In the simulated example, it peaks around -11 dB, while the same peak in the physical data peaks at about -23 dB. A probable reason for this, as well as the inconsistencies of the other peaks, is the material choice in the simulation. In Figure 3.9, the model for the digital twin is shown. There, the metallic surfaces correlate to the high peaks to the right and left of LoS in the plots of Figure 5.1.

In the simulation, metallic surfaces are considered perfect reflectors. They therefore reflect all incoming energy unlike real-life metals and thus the surfaces in the lab. Unlike in the simulation, these surfaces are not completely flat, which disrupts the reflections in the physical twin. The right wall can be seen in Figure 3.2.

It is also evident that the part of the room to the right of the receiving antenna is more similar between simulated and physical data than the left. The difference of the left peak between simulated and physical data is more than 10 dB, while the difference between the peaks to the right of LoS is between 5-10 dB, depending on the peak. This indicates that the cabinet to the left perhaps shouldn't be estimated as a perfect reflector, while that approximation is more accurate in the case of the right side. This larger variation in the simulated data is still present when introducing a human in the environment, which can be seen Figure 5.2.




**(a)** **(b)**

**Figure 5.2:** Two MUSIC plots with a human at position (1,1), one from the digital twin (left) and one from the physical twin (right). These are retrieved from 4.8b and 4.18b in result section.

Comparing these with the environment without a human in Figure 5.1, the introduction of the human at position (1,1) shifts the peaks slightly. However, note that the change is in opposite directions. For the simulated data the peak shifts downward, while for the physical data the peak shifts upward. This seems to indicate that while the human is less reflective than the surface closest to it in the simulation, it is more reflective than in reality. It could also be due to the different properties between the human model in the simulation and the real human in the physical measurements.

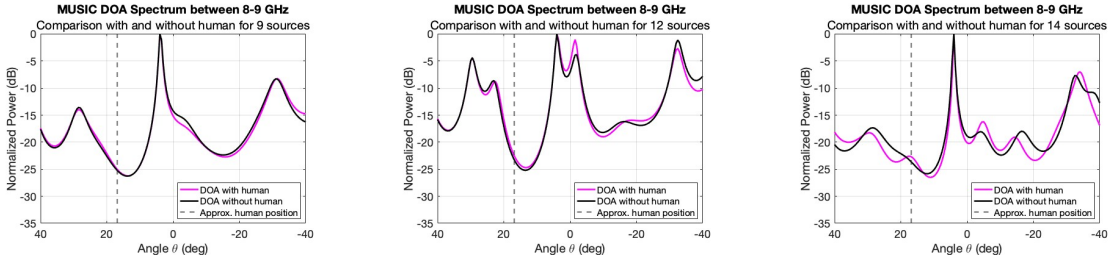
The number of sources chosen in the MUSIC algorithm is one of the most important parameters. It governs how many transmitters and reflectors of the signal that the algorithm should show as the output, considering the rest as noise that should not be included. As the number of sources directly correlates to the amount of peaks shown in the plot, choosing a lower amount of sources may omit important information.

In Figure 4.3, which has a completely open environment except for the human and a metallic cylinder, the difference of going from 3 to 4 sources is very noticeable. Although the difference would be less pronounced in a more complex environment with more objects reflecting the signal, this clearly shows the importance of not selecting an amount of sources that is too few. Choosing an excessive number of sources is also not preferable however, as then it would run the risk of including the noise subspace in the output. For the 4-16 GHz band in the simulations on the digital twin and on the physical data, 6 or 7 sources were chosen. Although the physical laboratory has numerous small objects that were excluded from the digital twin that do function as reflectors and thus sources of signal, the amount reflected from these is so small that it can be considered as noise. The physical data then does not need a higher number of sources than the simulated data, despite having more reflectors.

### 5.3.1 MUSIC on sub-bands

Comparing the different sub-band MUSIC spectra, they have very distinct differences. For example, by observing Figure 4.26 it shows sub-bands taken from physical

data with 9 sources. While all four have a clear LoS, the other peaks differ significantly from each other. Figure 5.3 below shows the 8-9 GHz sub-band for 9, 12 and 14 sources in particular.



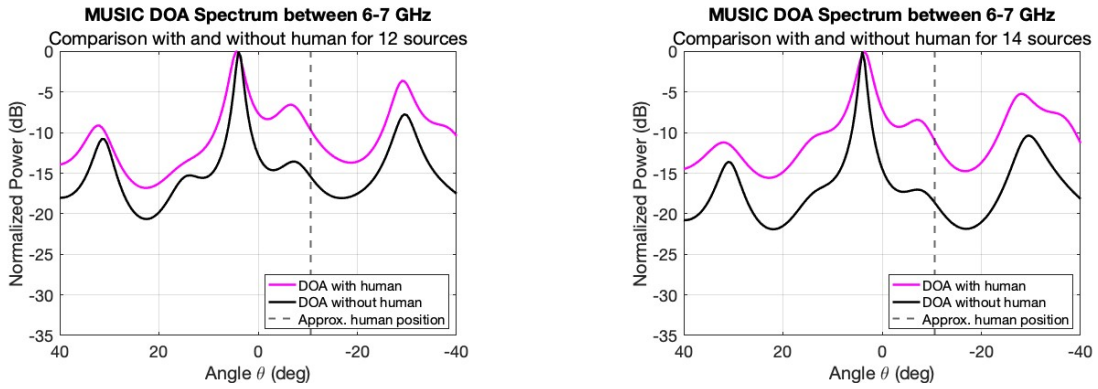

**(a)** **(b)** **(c)**

**Figure 5.3:** MUSIC plots with 9, 12 and 14 number of sources, in the 8-9 GHz sub-band, with the human at position (1,1). These are retrieved from 4.26c, 4.27c and 4.28c.

In 5.3a, the MUSIC spectrum changes very slightly around the approximated human position. Similar behavior is present in the case with 12 sources in 4.27c. It is evident that this frequency band is particularly poor at reflecting off of humans. Increasing the number of sources to 14 as in 5.3c, there is now a clearer peak where the human is, but if possible a lower amount of sources is preferable. The 8-9 GHz band is thus suboptimal for human detection, requiring a higher number of sources to discern the human.

It is notable however that the human seems to appear in the sub-band 8-9 GHz in the simulated data for 9 sources as in 4.16c, while the human is not distinguishable in the 7-8 GHz sub-band as in 4.16b. Increasing the number of sources to 12 in Figure 4.17, there is now a slight difference in the 7-8 GHz sub-band, but it is not as striking as the 8-9 GHz sub-band. A probable cause for this is the properties of the human model in the simulation and how it differs from reality. As was stated in Chapter 3.2.2, the human has the same properties as sea water in the simulation. Although this is an acceptable approximation, in the case of the sub-bands some flaws seem to show themselves.

One sub-band that is of interest is the 6-7 GHz sub-band when a human is standing in the (4,4) position, see Figure 5.2.




**(a)** **(b)**

**Figure 5.4:** MUSIC plots of the 6-7 GHz sub-band with a human at position (4,4), one with 12 sources and the other with 14. These are retrieved from 4.29a and 4.30a.

The introduction of the human in the environment shifts the entire graph except LoS upwards, instead of just shifting a single peak as in most MUSIC spectra. As the y-axis is normalized to the highest peak, the shifting of all peaks except LoS indicates that the LoS peak has become lower in amplitude. Position (4,4) is relatively close to LoS, so the human partially blocking it is possible.

It is of note that no sub-bands above 10 GHz is shown in the results. This is due to the higher amount of noise in the physical measurements when the sub-bands frequency increases. In Figure 3.1, the $S_{11}$ response of the transmitting antenna is shown. The amount of noise present increases as the frequency increases, being especially prominent from 10 GHz to 16 GHz. This leads to less clear MUSIC spectra, as the noise overpowers other signals. Plots of the 10-16 GHz sub-band can be found in A.2.

## 5.4 Machine learning

The results show clear differences between training on simulated data, real data, and a combination of both. By comparing the metrics positional accuracy, average distance, and validation loss, it is possible to evaluate how well the model learns spatial patterns and how well the learned features transfer to real-world measurements.

The human/empty metric remained consistently high across all experiments, likely due to class imbalance in the dataset, where most samples contained a human. As shown in Table 4.1, the prediction accuracy for human/empty after training remained 94.12% for all configurations, which corresponds to 16/17 samples. This suggests that the model largely relies on predicting the dominant class and therefore predicts the presence of a human in most cases. As a result, this metric becomes less informative for evaluating the overall model performance.

Several configurations were evaluated when training on simulated data followed by fine-tuning on real data. Increasing the number of epochs for real data generally improved the position accuracy, indicating that additional exposure to real mea-

surements helps the model adapt to real-world conditions. In contrast, increasing the number of simulated epochs did not consistently improve the results. Using a large number of simulated epochs could tend to reduce the positional accuracy after fine-tuning, suggesting that the model learned patterns specific to the simulated environment that do not transfer well to real data. Better results were instead obtained when fewer simulated epochs were combined with more epochs on real data. This indicates that simulated data is useful for learning general spatial features, while real data is essential for the model to adapt to real measurements.

The results before fine-tuning further support this observation. As seen in Table 4.1, the model achieved very high positional accuracy and a very low average distance when evaluated directly on simulated data. This suggests that the model learned the simulated data very well, likely due to simulated data being cleaner and contains less noise than the real measurements. However, the performance decreased significantly after fine-tuning and evaluation on real data, where the positional accuracy dropped and the average distance increased. This indicates that a domain gap exists and that the model may overfit to the simulated data.

In addition to fine-tuning experiments, the model was also trained on only real data and on a mixed dataset containing both simulated and real data. The best overall performance was achieved using the mixed dataset. This indicates that simulated data helps the model learn broader general spatial patterns and underlying structures, while real data improves its ability to handle real-world conditions. Training only on real data resulted in poorer performance, likely due to the limited amount of data. Including simulated data therefore appears to improve the model's ability to generalize to unseen real-world environments.

The validation loss curves further indicate that the model learned meaningful features, as the curves generally decreased over the epochs. The decrease was also relatively consistent, which suggests that the training process was stable and that the chosen CNN architecture and learning rates were appropriate. A noticeable difference was however that the validation loss for simulated data decreased faster and reached lower values than the validation loss for real data, which also indicates that the simulated environment was easier for the model to learn from.

## 5.5 Comparing MUSIC and Machine learning

When comparing MUSIC with machine learning, there were several strengths and weaknesses to each approach. As could be seen in the results, both were capable of detecting changes caused by human presence in the wireless channel, but the two algorithms achieved this in fundamentally different ways.

Firstly, a significant difference was that the MUSIC algorithm was grounded in the actual physics relating to the wireless channel characteristics. The output peaks in the MUSIC spectrum could be directly linked to a physical propagation path such as line of sight or the reflection caused by a human. The machine learning algorithm did not explicitly model physical phenomena like DOA and multi-path propagation.

It instead worked by learning patterns directly from the data, and was therefore somewhat less physically interpretable compared to MUSIC regarding the process of achieving their respective results.

A limitation of MUSIC was that performance changed depending on the number of assumed sources as well as the selected sub-band. It could be concluded that MUSIC requires more prior knowledge of the physical environment in order to achieve reliable results. Machine learning instead learned statistical relationships between the $S_{21}$ data and the presence or position of a human. Its advantage was therefore that machine learning implicitly learned multi-path effects, reflections and phase distortions without needing to explicitly model them in the algorithm code. However, a significantly larger amount of data was needed to obtain a valuable result from machine learning, compared with MUSIC where only a few measurements per square in the grid were needed to ensure consistency and reliability.

The results further showed that even when machine learning made incorrect predictions, they tended to be close to the true position of the human which could be seen in the average distance results in Tables 4.1 and 4.2. In contrast, the MUSIC algorithm did not provide information about such errors when the wrong number of sources was set, making it's results more difficult to interpret when no distinct peak clearly corresponded to human presence.

Another important distinction was that the MUSIC algorithm produced a spectrum that needed to be visually analyzed to identify human presence. Machine learning provided a direct and more automated output that immediately showed whether a human was present or not according to its prediction. For real applications, machine learning may therefore be seen as more practical since it removes the need of manual analysis.

## 5.6 Future Studies

This section presents possible improvements and possibilities for future work, based on the obtained results.

Overall it can be seen from the results that the positions further away from the LoS of the receiving and transmitting antenna gave slightly stronger indications of human presence. As noted before, the receiving antenna is placed further off-center, which resulted in clearer results in position (1,1) and its column. This possibly indicates that shifting the entire LoS to the edges of the grid would yield stronger results in all positions. The shift could be relevant if any future studies are conducted within approximately the same area as in this study's scenario. However, if the area is significantly larger, the effectiveness of the grid itself may deteriorate. Another system of labeling positions might have to be implemented.

Additionally, the lab environment in which the measurements were conducted had a large amount of electronic equipment and furniture, which was superfluous in pursuing this specific purpose. The disturbances, in the shape of extra reflections and noise, that this might have caused in the results, are difficult to estimate. In

any future pursuits, it would be optimal to perform these measurements with the minimum equipment and furniture in the room. Another part of the measurements that could be improved in future studies, is having an algorithm that more accurately could estimate the number of sources especially for the sub-band measurement.

For machine learning, to further improve the model robustness and generalization capability, a larger dataset containing a wider range of measurement scenarios would be required. In addition, the simulated environments should be refined to better represent real-world conditions and thereby reduce the domain gap between simulated and measured data.

Another improvement would be to normalize both the physically measured data and the simulated data in order to mix them better. One way to achieve this could be to use the data processed by MUSIC.

## 5.7 Ethics

The main ethical concerns, if implemented as a practical system, are privacy concerns. A person's presence could be detected without their knowledge, especially when combined with third-party devices, such as smartphones. Addressing these ethical risks requires user consent, limitations on data usage, and transparency. However, from the results it can be stated that this system mainly detects if something has changed in the environment, which can be that a human has entered the room. This reduces the risk of identifying and profiling the person significantly. Another concern could be the radio waves emitted from the antennas, and its influence on the human body. However, because of the low power it is non-ionizing which makes this setup harmless to humans.

# 6

# Conclusion

The main goal of this thesis was to identify human presence in an indoor environment by analyzing data from the $S_{21}$ parameter. Utilizing the IFFT did not give clear results, mainly due to noise in the physical environment. The MUSIC-algorithm successfully produced spectra in which the human could often be seen as a slight shift in one of the existing peaks. The MUSIC algorithm was seen to be highly dependent on the chosen number of sources, which gave insight into the possibility of developing a more efficient algorithm that could correctly identify the optimal number of sources needed to find the human. Another factor that affected the ability to detect the human was the bandwidth within the MUSIC algorithm was applied. The placement of the antennas was also an important factor, since graphs with locations further away from line of sight were more accurate on depicting the humans position. Nevertheless, the results demonstrated that human-induced changes in the wireless channel could be observed and analyzed through DoA estimation.

The machine learning algorithm was another task of this thesis, which provided further understanding of how to use $S_{21}$ data for human presence detection, as well as its potential for future sensing applications. It could be concluded that simulated data was valuable for helping the model learn general spatial patterns of the environment. The approach of fine-tuning demonstrated promising potential, since training on simulated data allowed the model to learn fundamental characteristics before adapting to the real measurements. Training on a mixed dataset, where no fine tuning was involved, provided the highest accuracy, which supports the conclusion that simulated data is useful to help the model generalize to unseen data. While several improvements are still required, the neural network showed that combining simulated and measured data is a promising approach for future sensing systems.

# Bibliography


[1] Ericsson. "6g," Ericsson, Accessed: May 12, 2026. [Online]. Available: https://www.ericsson.com/en/6g.

[2] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005, ISBN: 9780521845274.

[3] J. Andrews, M. Kowsika, A. Vakil, and J. Li, "A motion induced passive infrared (pir) sensor for stationary human occupancy detection," in *2020 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, 2020, pp. 1295–1304. DOI: 10.1109/PLANS46316.2020.9109909.

[4] ETSI, "Integrated Sensing And Communications (ISAC); Use Cases and Deployment Scenarios," European Telecommunications Standards Institute, ETSI Group Report ETSI GR ISC 001 V1.1.1, Mar. 2025. [Online]. Available: https://www.etsi.org/deliver/etsi_gr/isc/001_099/001/01.01.01_60/gr_isc001v010101p.pdf.

[5] W. Saad, M. Bennis, and M. Chen, "A vision of 6g wireless systems: Applications, trends, technologies, and open research problems," *IEEE Network*, vol. 34, no. 3, pp. 134–142, 2020. DOI: 10.1109/MNET.001.1900287.

[6] International Telecommunication Union, "Framework and overall objectives of the future development of IMT for 2030 and beyond," ITU Radiocommunication Sector, Recommendation ITU-R M.2160-0, Nov. 2023, Recommendation ITU-R M.2160-0. [Online]. Available: https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M.2160-0-202311-I!!PDF-E.pdf.

[7] 3rd Generation Partnership Project (3GPP), "Study on channel model for frequencies from 0.5 to 100 ghz," 3GPP, Technical Report TR 38.901, version V17.0.0, 2022, Release 17. [Online]. Available: https://www.3gpp.org/DynaReport/38901.htm.

[8] A. Kaushik et al., "Toward Integrated Sensing and Communications for 6G: Key Enabling Technologies, Standardization, and Challenges," *IEEE Communications Standards Magazine*, vol. 8, no. 2, pp. 52–59, 2024. DOI: 10.1109/MCOMSTD.0007.2300043.

[9] A. Liu et al., "A survey on fundamental limits of integrated sensing and communication," *IEEE Communications Surveys  Tutorials*, vol. 24, no. 2, pp. 994–1034, 2022. DOI: 10.1109/COMST.2022.3149272.

[10] J. A. Mwakali and G. Taban-Wani, Eds., *Proceedings from the International Conference on Advances in Engineering and Technology*, Oxford, UK: Elsevier Science, 2006.

[11] C. Wang, S. Chen, Y. Yang, F. Hu, F. Liu, and J. Wu, "Literature review on wireless sensing-wi-fi signal-based recognition of human activities," *Tsinghua Science and Technology*, vol. 23, no. 2, pp. 203–222, 2018. DOI: 10.26599/TST.2018.9010080.

[12] W. Hong et al., "Integrated sensing and communication (isac) channel model towards 3gpp 6g standardization: Modelling, validation, and application," *IEEE Journal on Selected Areas in Communications*, 2026. DOI: 10.1109/JSAC.2026.3656257.

[13] I. Darwazeh and C. Poole, "S-parameters," in *Microwave Active Circuit Analysis and Design*, C. Poole, Ed., 1st ed., London, England: UCL, 2015, ch. 6, pp. 167–204.

[14] W. Fan, Z. Yuan, Y. Lyu, and G. F. Pedersen, "Enabling vna based channel sounder for 6g research: Challenges and solutions," in *2024 18th European Conference on Antennas and Propagation (EuCAP)*, 2024, pp. 1–5. DOI: 10.23919/EuCAP60739.2024.10501372.

[15] J. M. Blackledget, "Chapter 4 - the fourier transform," in *Digital Signal Processing (Second Edition)*, ser. Woodhead Publishing Series in Electronic and Optical Materials, J. M. Blackledget, Ed., Second Edition, Woodhead Publishing, 2006, pp. 75–113, ISBN: 978-1-904275-26-8. DOI: 10.1533/9780857099457.1.75. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9781904275268500051.

[16] R. Oshana, "4 - overview of digital signal processing algorithms," in *DSP Software Development Techniques for Embedded and Real-Time Systems*, ser. Embedded Technology, R. Oshana, Ed., Burlington: Newnes, 2006, pp. 59–121, ISBN: 978-0-7506-7759-2. DOI: 10.1016/B978-075067759-2/50006-5. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B9780750677592500065.

[17] R. N.Bracewell. "Fourier series and transforms," AccessScience. [Online]. Available: https://www.accessscience.com/content/article/a270600.

[18] S. Wei, X. Ma, and J. Wang, "Experimental teaching reform of window function design in digital signal processing course," in *2022 Asia Conference on Algorithms, Computing and Machine Learning (CACML)*, 2022, pp. 385–390. DOI: 10.1109/CACML55074.2022.00071.

[19] T. M. Inc., *Kaiser window*, 2026. Accessed: Apr. 27, 2026. [Online]. Available: https://se.mathworks.com/help/releases/R2025a/signal/ref/kaiser.html.

[20] MathWorks. "MUSIC Super-Resolution DOA Estimation," MathWorks, Accessed: Feb. 20, 2026. [Online]. Available: https://se.mathworks.com/help/phased/ug/music-super-resolution-doa-estimation.html.

[21] H. Tang, "DOA Estimation Based on MUSIC Algorithm," BSc thesis, Linnaeus University, Växjö, Sweden, 2014. [Online]. Available: https://lnu.diva-portal.org/smash/get/diva2:724272/FULLTEXT01.pdf.

[22] MathWorks. "Grating Lobes," MathWorks, Accessed: Feb. 25, 2026. [Online]. Available: https://se.mathworks.com/help/antenna/ug/grating-lobes.html.

[23] T. Salman, A. Badawy, T. M. Elfouly, A. Mohamed, and T. Khattab, "Estimating the number of sources: An efficient maximization approach," in *2015 International Wireless Communications and Mobile Computing Conference (IWCMC)*, 2015, pp. 199–204. DOI: 10.1109/IWCMC.2015.7289082.

[24] A. Stanescu, R. A. Mata-Toledo, and P. Gupta. "Machine learning," AccessScience, Accessed: Feb. 25, 2026. [Online]. Available: https://www.accessscience.com/content/article/a395250.

[25] GeeksforGeeks. "What is a neural network?" GeeksforGeeks, Accessed: May 9, 2026. [Online]. Available: https://www.geeksforgeeks.org/deep-learning/neural-networks-a-beginners-guide/.

[26] GeeksforGeeks. "Introduction to convolution neural network," GeeksforGeeks, Accessed: May 9, 2026. [Online]. Available: https://www.geeksforgeeks.org/machine-learning/introduction-convolution-neural-network/.

[27] GeeksforGeeks. "Relu activation function in deep learning," GeeksforGeeks, Accessed: May 9, 2026. [Online]. Available: https://www.geeksforgeeks.org/deep-learning/relu-activation-function-in-deep-learning/.

[28] GeeksforGeeks. "What is batch normalization in deep learning?" GeeksforGeeks, Accessed: May 9, 2026. [Online]. Available: https://www.geeksforgeeks.org/deep-learning/what-is-batch-normalization-in-deep-learning/.

[29] GeeksforGeeks. "Dropout in neural networks," GeeksforGeeks, Accessed: May 9, 2026. [Online]. Available: https://www.geeksforgeeks.org/machine-learning/dropout-in-neural-networks/.

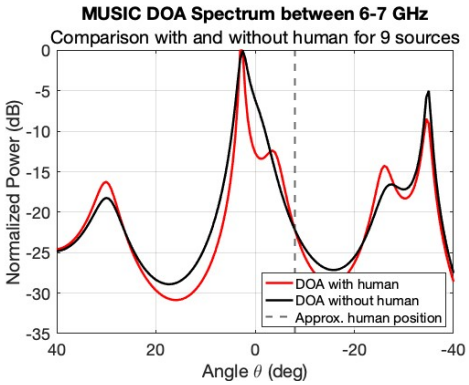
[30] Ansys Inc. "Ansys hfss | 3d high frequency simulation software. "[Online]. Available: https://www.ansys.com.

[31] E. Fredenvall, G. Ferm, M. Obaitan, A. Bjärkby, and J. Fagerstål, "Wireless channel for 6th generation networks sensing," Chalmers University of Technology, Tech. Rep., 2025.

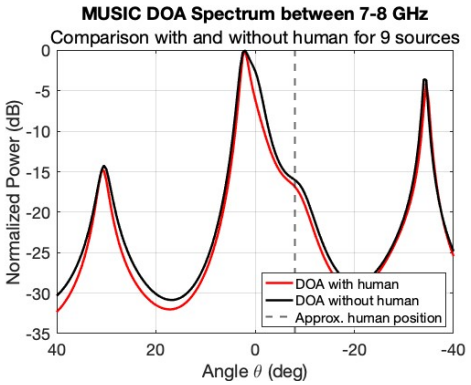
# A

# Appendix 1

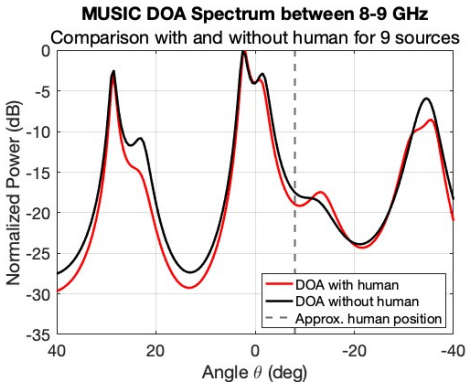
## A.1 Sub-band for 4.4 on simulated data



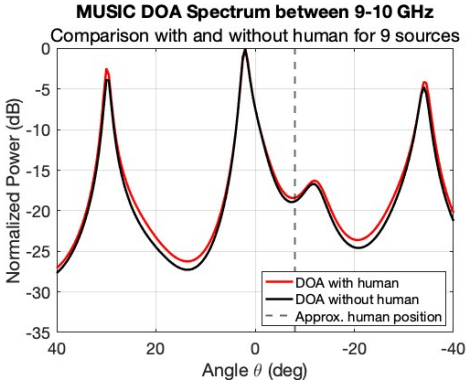

**(a)** Sub-band 6-7 GHz with 9 sources.




**(b)** Sub-band 7-8 GHz with 9 sources.

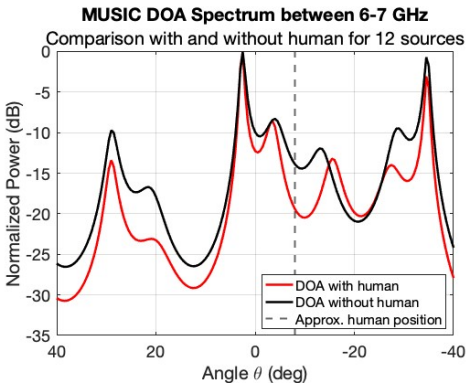



**(c)** Sub-band 8-9 GHz with 9 sources.

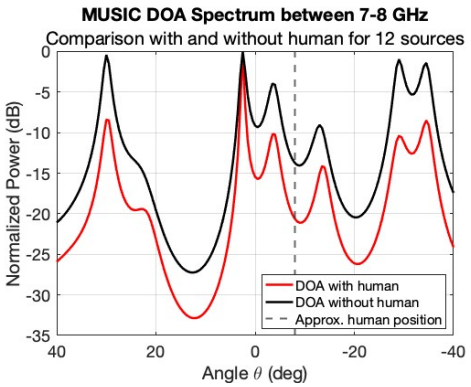



**(d)** Sub-band 9-10 GHz with 9 sources.

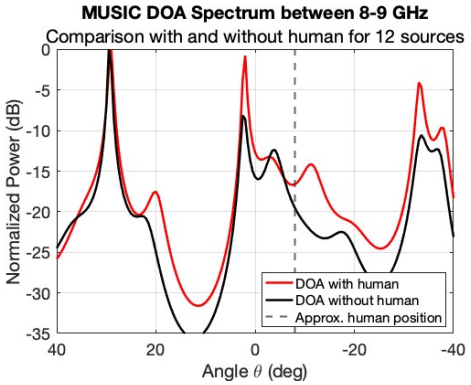
**Figure A.1:** MUSIC plots with different sub-bands frequency from 6 to 10 GHz, with 9 number of sources, when the human at position (4,4).



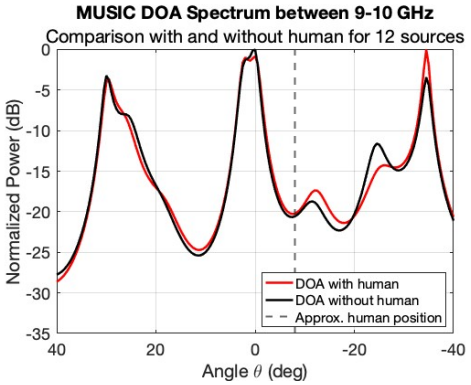

**(a)** Sub-band 6-7 GHz with 12 sources.




**(b)** Sub-band 7-8 GHz with 12 sources.



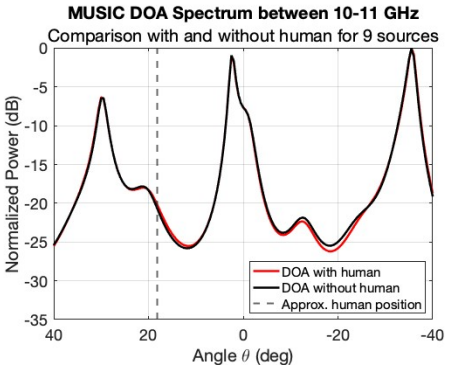

**(c)** Sub-band 8-9 GHz with 12 sources.

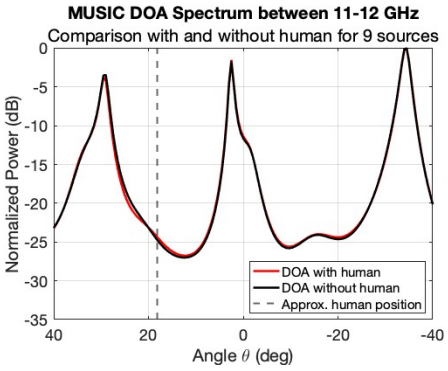


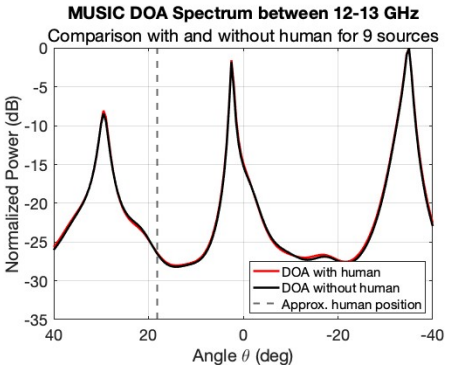

**(d)** Sub-band 9-10 GHz with 12 sources.

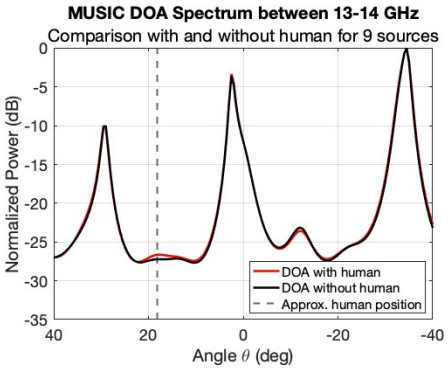
**Figure A.2:** MUSIC plots with different sub-bands frequency from 6 to 10 GHz, with 12 number of sources, when the human at position (4,4).

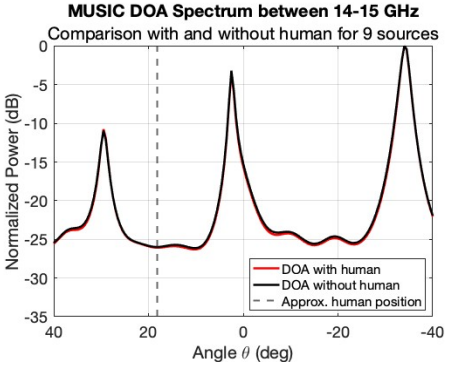
## A.2 10-16 GHz sub-band for 1.1 for simulation and physical

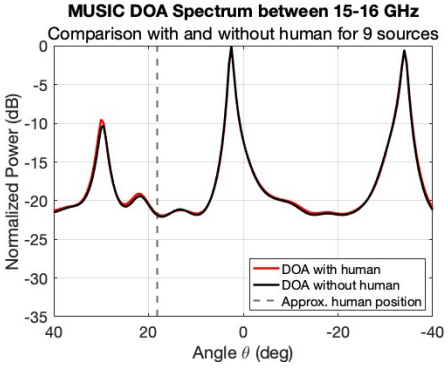


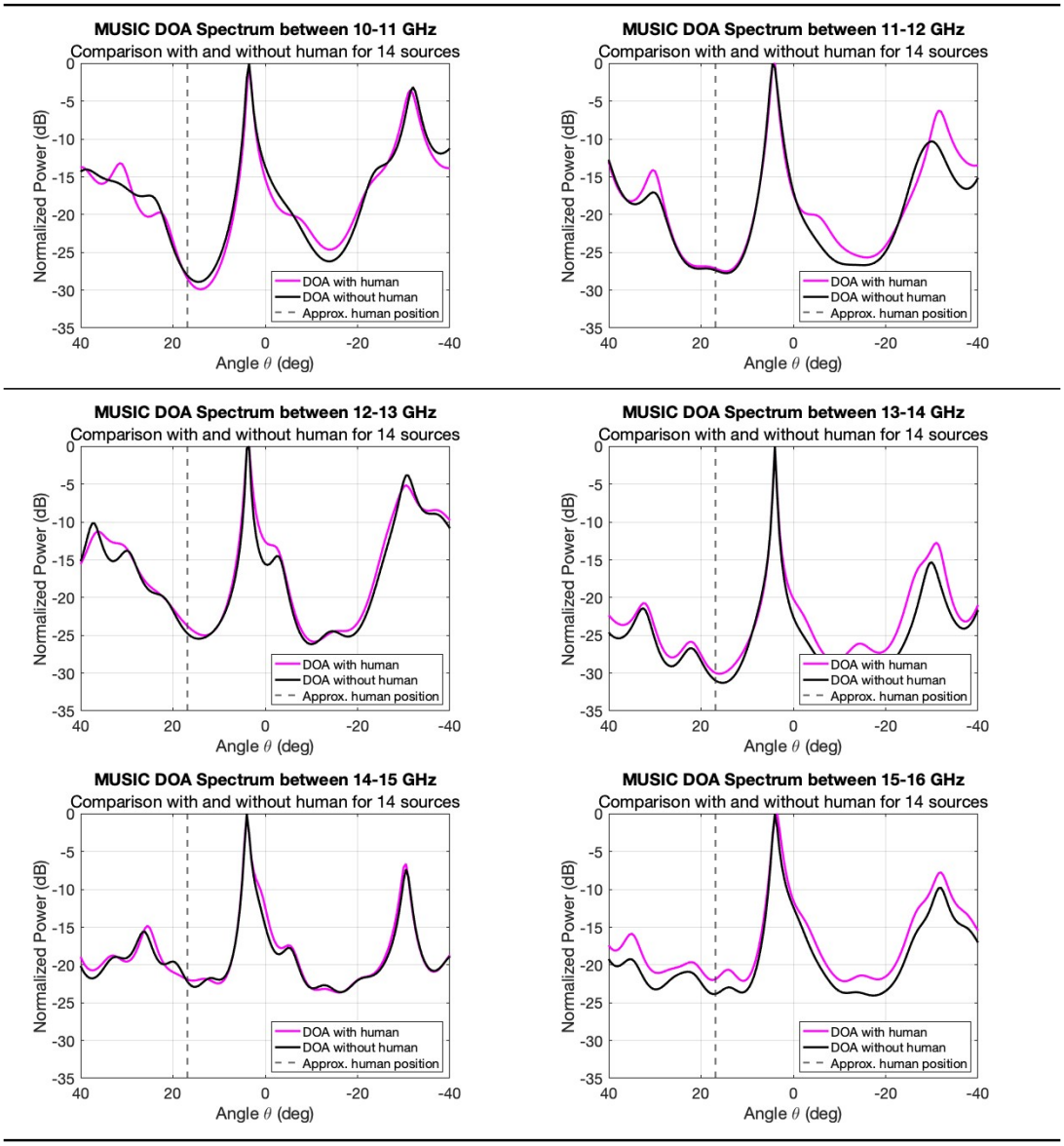






## A.3 Google Drive link to the thesis code and data

Link to Google Drive with code and data used for the MUSIC algorithm and machine learning. The link will be available for use until 2027.

`https://drive.google.com/drive/folders/1OI4mMFHlRrA4P1dVnmKabk0s_uG6fsjH?usp=share_link`

**DEPARTMENT OF SOME SUBJECT OR TECHNOLOGY**
**CHALMERS UNIVERSITY OF TECHNOLOGY**
Gothenburg, Sweden
www.chalmers.se


CHALMERS