



LUND UNIVERSITY
Faculty of Science

Detectability of fiducial markers in an MRI-only radiotherapy treatment workflow for prostate

Sevgi Emin

Thesis submitted for the degree of Bachelor of Science
Project duration: 5 months

Supervisors: Christian Gustafsson (MSc), Prof. Lars E Olsson
and Prof. Kristina Stenström

Department of Physics
May, 2017

Abstract

In a conventional prostate cancer radiotherapy treatment workflow, computed tomography (CT) and magnetic resonance imaging (MRI) are used for radiotherapy treatment planning and target delineation. MRI is necessary in the process due to better soft tissue contrast compared to CT-images. Using two different imaging modalities in radiotherapy treatment planning requires image registration, which could lead to inaccuracies in calculations. Therefore, an MRI-only based workflow for prostate cancer treatment has been suggested.

Gold fiducial markers (GFMs) are inserted in the prostate prior to treatment. GFMs are used for positioning of the prostate during treatment. GFMs may also be used for image registration between CT- and MR- images in a conventional workflow.

The fiducial markers create a high intensity signal in CT- images, but only a signal void in MRI. The identification method of GFMs in MR-images is currently a manual process requiring manual labor and therefore extra resources. An automatic identification method for GFMs has therefore been developed in a previous study. In this method, a new sequence of MR-images was used, referred to as multi-echo gradient echo images (MEGRE). The MEGRE image sequence was verified for clinical application by a human observatory study.

The human observatory study was conducted using T2-weighted (T2w) MR- and MEGRE images of 44 prostate cancer patients. The purpose of the study was to compare the performance of human observers and the automatic method in identification of gold fiducial markers to assess the accuracy of the methods.

A matrix laboratory (MATLAB) graphical user interface (GUI) was created for the human observatory study. The GUI was based on an existing GUI for the automatic GFM identification method. The study was conducted at the Department of Medical Physics at Skåne University Hospital in Lund.

Three observers conducted the study and the performance was evaluated for two different conditions. For the first condition, a true marker position was considered to be within the distance of 7.5 mm from the reference coordinates, whilst for the second one the distance was 5 mm. The results of the human observatory study showed 98.33 ± 1.53 %, 95.33 ± 4.16 %, 97.33 ± 2.31 % and 96.00 ± 1.73 %, 89.67 ± 5.69 %, 94.33 ± 3.21 % for the sensitivity, specificity and accuracy at 7.5 mm and 5 mm respectively. The internal differences average between the markers selected by the observers compared to reference markers was 1.07 ± 1.04 mm for T2w images.

The MEGRE image sequence was validated during the study and the results showed a better accuracy in MEGRE images compared to T2w images. The performance of the human observatory study in terms of sensitivity, specificity and accuracy was superior to the automatic method. However, deviations between observers existed, which is expected to be avoided by using the automatic method.

Acknowledgments

First of all, I wish to express my sincere gratitude to my supervisor Christian Gustafsson. Thank you for helping me find my way in the chaotic world of GUIDE and teaching me the basics of medical imaging in such a short time. Without your help, this project would not have been possible.

I would like to thank my co-supervisors Lars E. Olsson and Kristina Stenström, for helping me find this project and supervising my work.

I am also grateful to my classmates and loved ones for supporting me and encouraging me. Without you I would not make it so far.

Contents

Abstract	ii
Acknowledgments	iii
1 Introduction	1
2 Theory	2
2.1 Radiotherapy for prostate cancer	2
2.2 Use of gold fiducial markers in radiotherapy	3
2.2.1 Clinical identification of gold fiducial markers	4
2.3 Treatment planning	5
2.3.1 Conventional CT workflow	5
2.3.2 MRI-only workflow	5
2.4 Radiotherapy patient setup	5
2.5 CT-Imaging	6
2.6 MR-imaging	8
2.6.1 Magnetic susceptibility and its effects on MR-images	9
2.7 Automatic identification of gold fiducial markers	10
3 Methods	11
3.1 Observatory study of gold fiducial markers	11
3.1.1 Observers and patient data	11
3.1.2 Graphical user interface structure and function	11
3.1.3 Data analysis	16
4 Results	18
4.1 Duration of identification process	18
4.2 Assessment of the observer performance	19
4.2.1 Evaluation score	19
4.2.2 Gold fiducial marker spatial positions	20
4.3 Assessment of the observer accuracy	22
5 Discussion & Conclusions	24
6 Outlook	26
References	27
Appendices	29
A Manual for the human observatory study	29
B Patient cases for min and max identification duration time	39
C Patient and gold fiducial marker data	40

List of Abbreviations

CBCT Cone beam computed tomography

CT Computed tomography

FSE Fast spin echo

GFMs Gold fiducial markers

GRE Gradient echo

MEGRE Multi-echo gradient echo

MRI Magnetic resonance imaging

NMV Net magnetic vector

OAR Organs at risk

RF Radio frequency

RT Radiotherapy

RTP Radiotherapy treatment planning

T2w T2-weighted

TE Echo time

TR Repetition time

1 Introduction

Prostate cancer is one of the most common cancer types in Sweden. Almost a third of all male cancer patients have prostate cancer. A common treatment method for prostate cancer is by external application of high-energy radiation beams to destroy the cancer cells. This is called external beam radiotherapy (RT) and is given in small fractions over a period of several weeks.

Different imaging techniques are used for radiotherapy treatment planning (RTP) and two of the most common types are computed tomography (CT) and magnetic resonance imaging (MRI). In CT imaging x-rays are used and the image contrast is dependent on the tissue density and atomic number of the atoms in the tissue. In MRI, the image contrast is achieved by exploiting the magnetic properties of the tissue.

In prostate cancer RT, it is important to apply the radiation on the dedicated volume containing the cancer cells and avoid the irradiation of healthy tissue for the surrounding organs. This is achieved by fiducial markers inserted into the prostate prior to the RT treatment. Gold is often used as a fiducial maker and can be easily visualized using CT due to the difference between gold atomic number and the atomic number of surrounding tissues. The gold fiducial markers (GFMs) can also be visualized using on-board imaging techniques prior to each RT fraction on the RT machine. In this way, dose application can be limited to a certain tissue volume throughout the whole radiotherapy treatment [1].

Due to non-optimal soft tissue contrast in CT-images, in RTP both CT and MRI are used. CT is used to get the attenuation data and MRI— for soft tissue contrast. To avoid using multiple kinds of imaging techniques and image registration, an MRI-only based workflow, where CT imaging will be excluded, has been suggested [2].

However, the GFM identification on images from fast spin echo (FSE) based MR-sequences is not as simple as in CT. The GFMs appearance in FSE MR-images is only a small signal void. In images from gradient echo (GRE) based MR sequences the GFM creates a larger signal void, which increases with increasing echo time. The lack of CT in an MRI-only workflow therefore causes difficulties in localization of the GFMs in spin echo based MR-sequences.

The current clinical practice for using MR-images in RT is in a CT + MRI combined workflow. The clinical identification of the fiducial markers in such a workflow is a manual process based on CT- and single echo gradient echo images. The process is dependent on the performance of a human observer [3]. Furthermore, the workflow is not always optimal as intra-prostatic calcifications can be mistaken for GFMs in the MR-images. In an MRI- only based workflow there is no CT and it is therefore of importance to improve the detection and correct identification of GFM in MR- images.

A software for automatic detection of gold fiducial markers in a new type of MR-images is currently being developed at Department of Medical Physics at Skåne University Hospital. The software aims to be more efficient, in terms of time and resources, than the current clinical identification process and eliminate observer based inaccuracies.

The purpose of this project is to validate this new type of MR-images for identification of GFMs. The image data sets will be based on a multi-echo gradient echo (MEGRE) data acquisition sequence. The validation will be performed by comparing the performance and accuracy of human observers against the performance and accuracy of the automated software regarding the detectability of the gold fiducial markers in this new type of MR-images.

2 Theory

2.1 Radiotherapy for prostate cancer

Radiotherapy is one of the most commonly used cancer treatment methods for prostate cancer, where cancer cells are irradiated by ionizing radiation. When the radiation, i.e. high energy gamma rays, is delivered on the tissue, referred to as the target, cancer cells will be damaged due to radiolysis- splitting of molecules due to radiation. This reduces the functions of the cell, in the end causing its death. In this way, the development of cancer can be slowed down, diminished or cured [4].

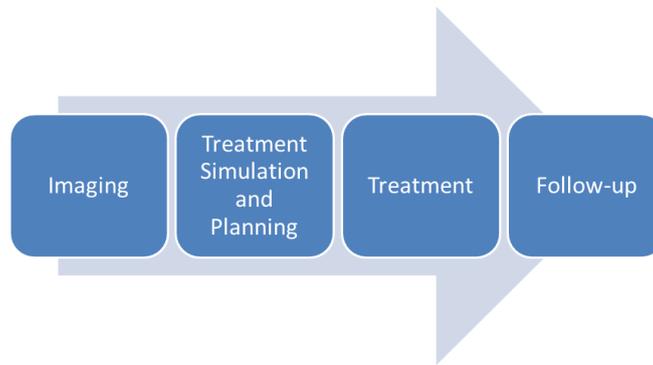


Figure 1: Workflow and different steps in radiotherapy.

As it is shown in Figure 1, the RT process can be divided into 4 main steps. In the first step, the cancer is located and examined. The treatment method is planned in the second step. CT-images are often used for treatment planning, which includes determining the necessary type of treatment and the amount of radiation. This is followed by applying the treatment using a rotating machine, shown in Figure 2. The RT machine applies beams of radiation to a selected cancer tissue of the patient. This whole process is followed up by regular follow-ups, such as imaging and physical examinations [5].

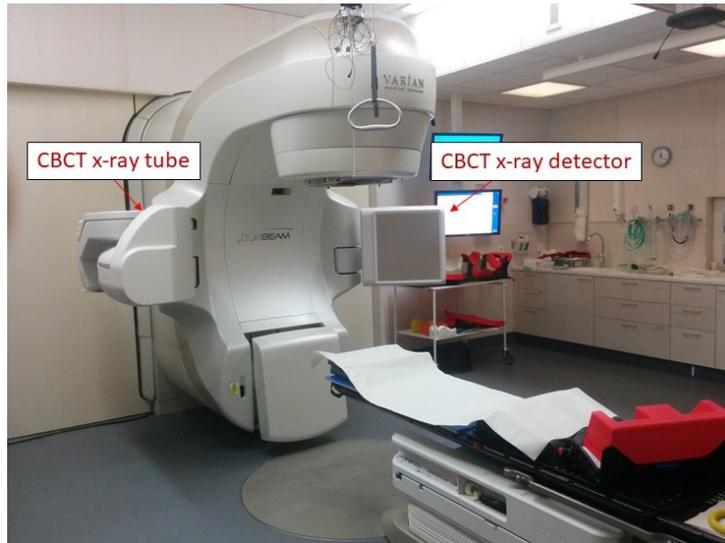


Figure 2: The radiotherapy machine and setup used at the Radiotherapy clinic at Skåne University hospital in Lund. A cone beam CT (CBCT) x-ray tube and detector positions are shown.

2.2 Use of gold fiducial markers in radiotherapy

In radiotherapy, radiation is applied to the volume affected by the cancer. In the case of prostate cancer, there are sensitive and vital organs, referred to as organs at risk (OAR), close to the prostate. Some of those organs are rectum, bladder, the reproductive system and can be damaged during the treatment. To ensure that the radiation dose is applied to the correct volume, fiducial markers are used to localize and correct the prostate position prior to treatment. This is achieved by comparing the current position of the GFMs to the original positions of the markers used in the CT dose planning. In case of differences in positioning, adjustments are made by moving the patient table on the RT machine. In this way, it is verified that the prostate position is the same for all fractions.

Another use of GFMs is during RTP, when CT+ MRI workflow is used. The CT- and MR-images are linked together by image registration, which is performed by defining three points on each image modality per image series, corresponding to three gold fiducial markers.

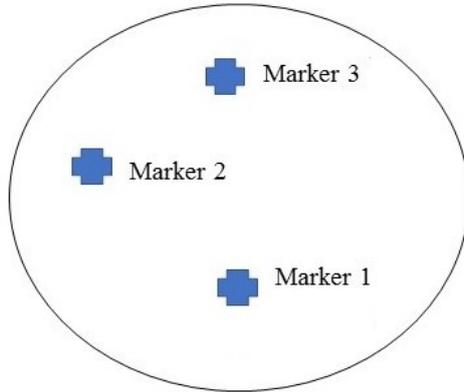


Figure 3: An example of a geometry for GFMs positions. Marker 1 is the most inferior, whilst Marker 3 is the most superior one. It is not required for the markers to be in the same plane.

The fiducial markers used at the Radiotherapy clinic at Skåne University Hospital in Lund are gold cylinders with a length of 5.0 mm and a diameter of 1.0 mm. The markers are inserted by an oncologist transrectally or perineally using preloaded, pre-sterilized needles. The process is usually not risky [3]. Typically, 3 gold fiducial markers are inserted and an example of a geometry is shown in Figure 3. Having 3 GFMs allows assessment of the prostate movement in all spatial directions [1]. Patients in Lund receive 3 GFMs, however the end result can be reduced to 2 due to fiducial marker migration.

2.2.1 Clinical identification of gold fiducial markers

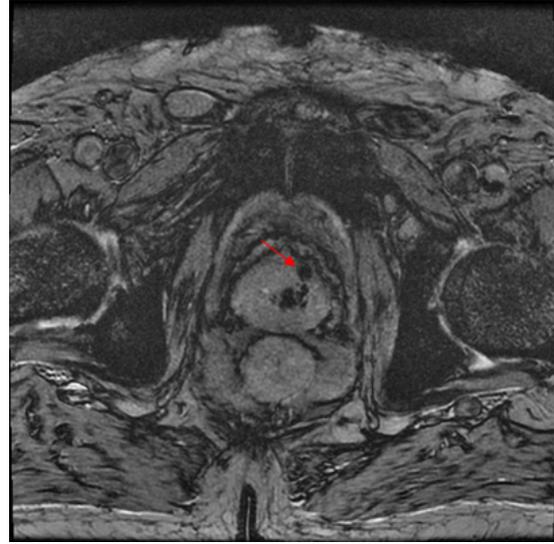
The clinical process of GFM identification in a combined CT + MRI workflow is currently based on images from the CT and images from the MRI. MR gradient echo images can be acquired for either one or multiple echoes. Echo time is the time it takes from application of a frequency pulse by the MRI machine to the measured maximum induced current signal in the scanned volume [6].

For GFM identification in a CT + MRI workflow a single-echo gradient echo image is used. In Figure 4, an example of two images is shown. In the images, one GFM and intra-prostatic calcifications, which are deposits of calcium salts in soft tissues, can be seen. The calcifications can be hard to differentiate from markers since both objects give rise to signal voids in MRI. On CT, the GFM induces streak artifacts and is easy to differentiate. This is due to the beam scattering by the gold markers, which leads to a higher signal intensity in the CT-image, appearing as bright streaks.

It can be seen that the depiction of calcifications is not optimal on the single-echo MR-image. In MEGRE images, which are a sequence of images taken at multiple echoes, the calcifications are seen as signal voids which increase in area with increasing echo time. This makes calcification identification in MEGRE images easier compared to single echo MR-images. Therefore, currently MEGRE images are clinically evaluated for use in GFM identification.



(a) CT-image



(b) Single echo MR-image

Figure 4: Images of a GFM and intra prostatic calcification. CT-image showing GFM (streak artifacts) and calcification (no streak artifact). Single echo gradient echo MR-image for corresponding slice position. Both marker and calcification are seen as signal voids. The red arrows on the images point at the GFM location.

2.3 Treatment planning

2.3.1 Conventional CT workflow

CT plays an important role in radiotherapy as it enables three-dimensional imaging and calculation of dose. Conventionally, dose calculations are performed on CT-images and MR-images are used in combination with CT for better target and OAR delineation. The target is the location of the tissue where the radiotherapy will be applied. During this process, the exposure of OAR needs to be avoided [7]. Furthermore, using both CT and MRI imaging modalities during the conventional CT workflow, requires image registration.

2.3.2 MRI-only workflow

MRI offers a superior tissue contrast compared to CT imaging for soft tissue and tumor definition. Additionally, in an MRI-only based workflow, image registration between multiple image modalities can be avoided. This can mitigate positioning errors inferred from such an image registration. Methods for using MRI for dose planning are being developed. In the MRI-only based workflow, CT-images can be substituted by synthetic CT- (sCT) images produced by obtaining the Hounsfield units (described in 2.5 *CT-Imaging*) of the tissues from MR-images [8].

There are certain aspects, such as the detectability of GFMs, of MRI which make the integration of an MRI- only based workflow hard to achieve.

2.4 Radiotherapy patient setup

Prior to treatment, the location of the GFMs in the prostate is checked in treatment set-up and the current position is compared to the reference positions in the CT-images or to the sCT-images. The on-board imaging techniques of the RT machine can be planar x-ray

imaging or cone beam computed tomography (CBCT). In Figure 2, the position of the CBCT x-ray tube and detector can be seen. The setup rotates around the patient, taking images of the prostate. The same set up is used for the planar x-ray images, referred to as kV images. These are used for patient positioning during treatment. If there are any differences between the current and the reference GFM position, adjustments are made by moving the patient table to align the target prior to dose application.

2.5 CT-Imaging

In CT imaging, a narrow rotating x-ray beam scans a part of the anatomy, creating cross-sectional images. Tissues with different atomic numbers can be distinguished. By having cross-sectional images, the superimposition of overlying tissues can be avoided, which is a drawback in conventional radiography. In conventional radiography, the image used to be presented in 2-dimensions and was developed on a radiographic film. In radiography, the contrast between tissues with similar densities, such as fluids and soft tissue, is not very different [9].

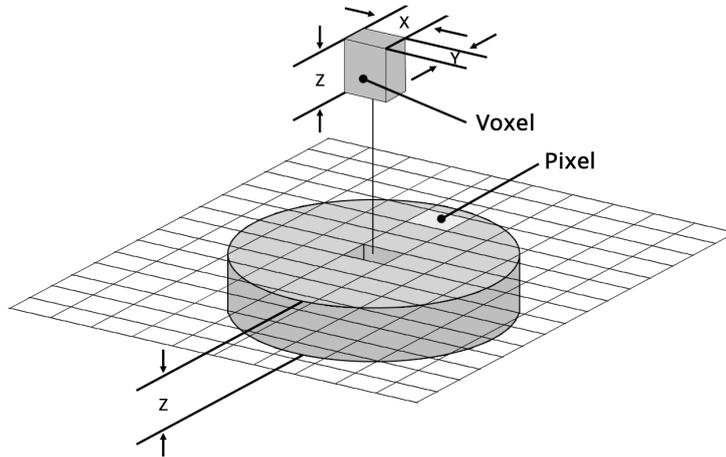


Figure 5: An example of a CT cross-sectioning. The dimensions of the cuboid voxel are given by x, y, z and its area represents a pixel. The slice thickness of the cross-section is given by z (Image reproduced from [10]).

The CT slices are acquired in axial plane, an example is shown in Figure 5. The slice thickness referred to as z can be selected manually and determines the thickness of the cross-section. Depending on the region of interest of the anatomy, the slice thickness can be adapted. The mechanically adjusted collimator determines the width of the x-ray beam and therefore the slice thickness. In this way, only a certain volume is exposed, decreasing the chances of scattering of x-rays and superimposition of tissues. In order to reconstruct a CT image, the scanned area is also divided into pixels, with width x and height y . The area of these pixels and the slice thickness form a cuboid called voxel, shown in Figure 5. When the acquired data is reconstructed and processed, the result is presented in an image matrix, usually in the size of 512×512 . Each pixel is assigned a signal value and therefore a shade of gray [10].

The phenomenon leading to different signal values in the pixels is attenuation of radiation. The x-ray beams are consisted of light quanta, called photons. Beam attenuation is the amount of reduction of photons in a beam, caused by the material it passed through. It is dependent on the thickness, the density and the atomic number of the material. The

probability of absorption is higher in denser and high atomic number materials, due to the fact that the atoms are closely packed. Linear attenuation coefficient, denoted by μ , measures the amount of attenuation per unit length. Some examples of attenuation coefficients are given in Table 1. These values are dependent on the energy of the beam. For more energetic photons the coefficient decreases, whilst for low energy photons it increases [10].

Table 1: Attenuation coefficients of different materials for a x-ray beam of 125-kVp (kilovoltage peak, this energy is often used in CT imaging) [10].

Material	μ (cm^{-1})
Air	0.0003
Water	0.180
Blood	0.182
Dense bone	0.46
Gold	88.71 [11]

A more common unit used for attenuation in CT-images is the Hounsfield unit (HU), which shows the degree an x-ray beam is attenuated by a material and is calculated by [12]:

$$HU = \frac{\mu_{material} - \mu_{water}}{\mu_{water} - \mu_{air}} \times 1000 \quad (2.1)$$

The Hounsfield unit for distilled water is assigned to 0, whilst for air it is -1000. In Figure 4, the range of the signal strength can be seen, where tissues with larger Hounsfield unit — such as bones and gold— appear bright, whilst tissues with smaller Hounsfield unit (e.g. air) appear dark.

Each photon in the radiation beam has a characteristic energy and when they come in contact with an atom this can lead to Compton scattering or to the photoelectric effect [13]. If Compton scattering occurs, the direction of the photons in the beam changes leading to reduction in beam intensity. In case of photon absorption by a tissue, an electron can be emitted, this is referred to as the photoelectric effect. The emission is dependent on the energy of the incident photons. If the photon energy is larger than the work function of the electron, defined by the binding energy, the electron will be emitted [14]. The photoelectric effect decreases the intensity of the beam and depending on the density and atomic number of the material, absorption will be different for each type of tissue. As a result, a higher number of photons will pass through some tissues, which will be represented by an area with low signal value in the CT-image, referred to as low attenuation. In cases, where the photons will be absorbed, the number will be lower, therefore it will be visible in the CT-image as a high signal value — high attenuation. In the case of GFMs, which absorb most of the photons, the position of the markers in a CT-image is represented by an area with high signal value (Figure 4).

A disadvantage of CT is the polychromatic source of x-ray beams, which produces photons at different energies — ranging from low to high values. The probability of a low energy photon to be attenuated is higher and this can lead to artifacts, which lowers the image quality. To solve this issue, the x-ray beam is incident on a shield where the low-energy photons are filtered out and a more homogeneous beam is produced [10].

Another disadvantage of CT is the lack of contrast between tissues with similar density, such as soft tissue. In order to be able to differentiate between tissues in CT-images, a density difference is necessary, otherwise the contrast will be low.

Furthermore, CT-imaging can have unwanted side-effects on the patient. X-rays are a type of ionizing radiation, exposing the patient to an additional radiation dose. In adults, this side effect does not cause a big issue, due to the fact that the CT radiation dose is negligibly small compared to the RT doses given. However, in cancer treatment of children the radiation dose given from CT can be of significance. Younger children's tissues and bones are more sensitive to radiation due to their age dependent development. This does not cause a problem in adults, since the adult tissues are not as affected by age. Late side effects in children, such as neurocognitive diseases and hindered growth of tissues and bones, can be caused due to RT. Therefore, any additional radiation besides radiotherapy is tried to be avoided [15].

2.6 MR-imaging

The spin or angular momentum of specific nuclei in tissues makes MRI possible. A nucleus of an atom is consisted of subatomic particles, referred to as nucleons— protons (positive charge) and neutrons (no charge). Each nucleon has an individual half spin, which can be either spin up (positive) or spin down (negative). The sum of all the individual spins yields the total spin of the nucleus. The protons and neutrons form pairs with opposite spins. In cases of nuclei with even number of nucleons, the resulting spin is 0, whilst the odd numbered nuclei have a net spin. Some examples of odd numbered nuclei are ^1H , ^{13}C , ^{15}N , ^{17}O , ^{19}F , ^{23}Na and ^{31}P . These nuclei can be affected by an external magnetic field and align their axis of rotation in the direction of the external field, also referred to as MR active nuclei [6].

When an external magnetic field (B_0 field) is applied, a resulting net magnetic vector (NMV) of the nuclei is created and aligns parallel to the direction of the B_0 field. This makes the magnetic moment of the nuclei follow a circular path around the direction of B_0 , referred to as precession. The frequency of the precessing magnetization is equal to the natural frequency of the nuclei, or the so called Larmor frequency:

$$\omega_0 = B_0 \times \lambda \quad (2.2)$$

where ω_0 precessional frequency, B_0 is the magnitude of the external magnetic field, and λ is the gyromagnetic ratio, which is the ratio of a particle's magnetic moment to its angular momentum. A radio frequency (RF) pulse at a variable frequency is applied during magnetic resonance imaging. The application of the RF pulse tilts the NMV to the transverse plane, perpendicular to the direction of B_0 field. The nuclei of the atoms in the tissues start oscillating, which can lead to resonance if the applied frequency of the RF pulse is equal to the Larmor frequency of the nuclei. In clinical MRI applications, the Larmor frequency of hydrogen is matched in the RF pulses to achieve magnetic resonance [6].

After application of the RF pulse the NMV realigns to its initial position, parallel to B_0 . The time it takes to regain 63 % of the magnitude of the NMV in the direction of B_0 is called T1 relaxation time. Tissues in the body have different T1 relaxation times. This can be exploited to create T1-weighted image contrast. For example, fat has a shorter T1 relaxation time than water. Fat will therefore appear brighter in the T1-weighted image compared to water.

The individual magnetic fields of neighboring nuclei interact with each other leading to loss of coherent NMV, referred to as T2 relaxation. Each type of tissue has a specific T2 relaxation time. T2-weighted images are often used for delineation of pathology in MR- images used for RTP.

The image contrast in the MR-image is determined by two MRI scanning parameters, defined on the MRI console. These are repetition time (TR) and echo time (TE). T1-weighted images have short TR and short TE. T2-weighted images have long TR and long TE.

In the detection of the MR signal, Faraday’s law of induction plays a big role. Faraday’s law states that a fluctuating magnetic field around a coil will generate voltage in the coil. A RF coil is positioned around the patient, as it can be seen in Figure 6, where the transverse precession vector induces electrical voltage [6]. This is the measurable MR signal. To determine the physical origin of the signal, time dependent linear variations of the magnetic field around the patient in the MRI scanner are introduced. An example of a MRI scanner is shown in Figure 6. The linear variations of the magnetic field are created by built in magnetic gradient coils. In this way, one can calculate the origin of the MRI signal. The information collected is represented in the k-space domain and transformed into image domain by applying an inverse Fourier transformation [6].

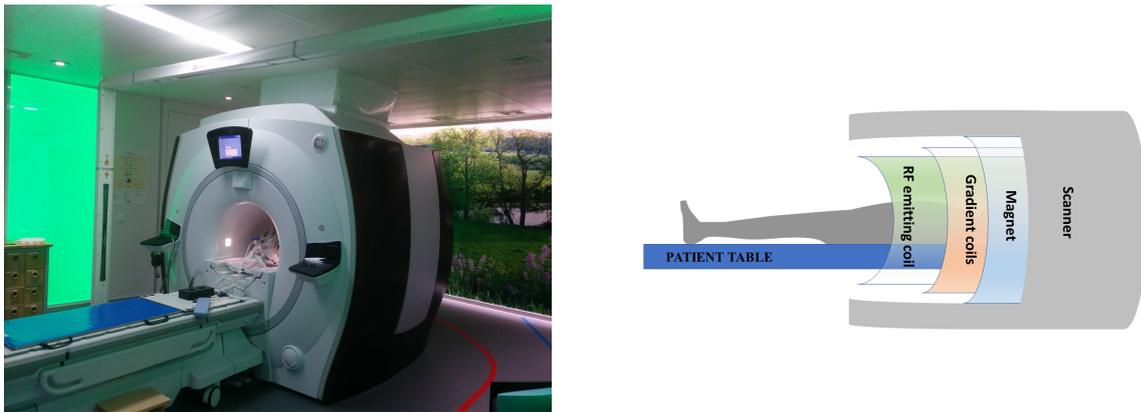


Figure 6: An example of a MRI machine used at the Radiotherapy clinic at Skåne University hospital in Lund. A diagram showing the main parts of the machine is provided.

2.6.1 Magnetic susceptibility and its effects on MR-images

Magnetic susceptibility describes the magnetization capacity of a material exposed to an external magnetic field. The value of magnetic susceptibility (χ_m) is constant for each element and it is described by the following formula:

$$\chi_m = \frac{M}{H} \quad (2.3)$$

where M is the magnetization value inside the material, whilst H is the magnitude of the external magnetic field [16].

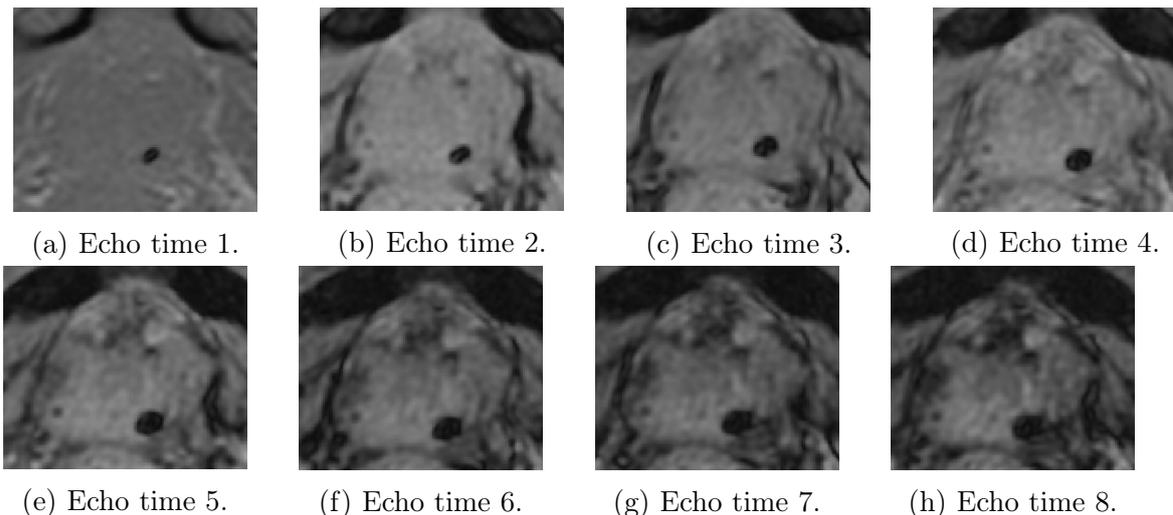


Figure 7: MR images of a prostate with a GFM at 8 different, equally spaced, echo times with the range 2.4–23.6 ms .

Image generation with MRI is dependent on a homogeneous static magnetic field (B_0). Materials with high magnetic susceptibility can induce strong internal magnetic fields along or in the opposite direction of B_0 . This will lead to reduction in the magnitude of the local field, leading to a distorted signal. An example of an image of a GFM in a prostate is shown in Figure 4. The magnetic field of gold atoms cancels out the external magnetic field, leading to a signal void in the image.

In Figure 7, MEGRE images of prostate with a GFM are shown at 8 different echo times. As it can be seen at higher echo times the artifact size increases with increasing echo time and the signal-to-noise ratio overall in the prostate tissue decreases. This relationship for the marker size is described by the linear change in the area of the artifact being proportional to B_0 and echo times [17].

2.7 Automatic identification of gold fiducial markers

The drawbacks of using CT imaging and the idea of an MRI- only based workflow were described in the previous sections. An automatic identification method for GFMs in prostate RT is aiming to provide a more efficient and reliable way of GFM identification. Resources, such as time and working staff, can be cut down and human observer based errors can be avoided.

The automatic identification method has been developed at the Department of Medical Physics at Skåne University Hospital in Lund. It was based on an algorithm acting on MEGRE images. The algorithm analyzed the size and the roundness of the GFM signal void area as a function of echo time. To discriminate against calcifications, the roundness and the change in area of the signal voids were analyzed.

The detected position of the GFM was determined directly in the target delineation sequence. The reference position of the GFM was determined using CT as a reference. If the detected position for the GFM was within 7.5 mm from the reference position it was considered as a true GFM [3].

3 Methods

3.1 Observatory study of gold fiducial markers

3.1.1 Observers and patient data

A data set of 44 prostate cancer patients with mean weight 86.6 ± 13.3 kg and mean age of 71.2 ± 5.3 years was used for the manual observatory study. The same set was also used for the automatic identification method [3], making a direct comparison between the methods possible. Permission to use the patient information has been granted by the local ethical review board and all patient data used in the study was anonymized. The results of the study will not affect the treatment of the patients.

The target delineation sequence of the study was a fast spin echo based T2w transversal MRI acquisition sequence. The GFM identification sequence was a multi echo gradient echo (MEGRE) acquisition sequence. The parameters for each sequence are presented in Figure 2.

Table 2: Parameters for MRI acquisition sequence, both for target delineation and GFM identification sequences.

Parameter	Target delineation sequence	GFM identification sequence
Sequence type	FSE	GRE
TR	9151 ms	1000 ms
TE	96 ms	2.4-23.6 ms
Number of echoes	1	8
Acquisition time	5 min	5 min

The human observatory study was conducted by 3 independent observers, who all were experienced medical physicists. The observers identified themselves as Observer 3, Observer 4 and Observer 6. Observer 4 was the founder of the automatic identification software and had seen the patient data 6 months before the study. Furthermore, he was the observer with the most experience in the MEGRE images. The work has been distributed over 4 weeks. The human observatory study was conducted using an interactive interface, which was on a remote desktop. The observer could log in and do the study whenever it was suitable for them.

3.1.2 Graphical user interface structure and function

The human observatory study was in the form of a graphical user interface (GUI), created in Matlab (version R2016b). GUI is an interactive interface, where the user can interact by using the pointer and other symbols [18]. The GUI of the study was based on the GUI developed for the automatic identification method [3]. The task of each observer was to locate the GFMs, select their positions and evaluate the certainty of their selection. This was performed in both target delineation images and GFM identification images. The positions of the selected markers were represented by Matlab coordinates.

The interface was created to make the human observatory study as simple and straightforward as possible. The workflow was dynamically color coded to show the next step the observer needed to take. Green color marked the next button the observer needed to click, and a text box was included to give information to the observer after each step.

Additionally, prior to the study the observer was given instructions verbally and in the form of an instructions manual explaining every step in detail. The manual is provided as Appendix A. Two test patients *Patient Test A* and *Patient Test B*, which can be seen in Figure 8, were included for training purposes.

The first step in using the GUI is shown in Figure 8. The observer needed to identify themselves by selecting the assigned number on a pop up list, shown after clicking "Start". Afterwards a patient was selected from the scroll list.

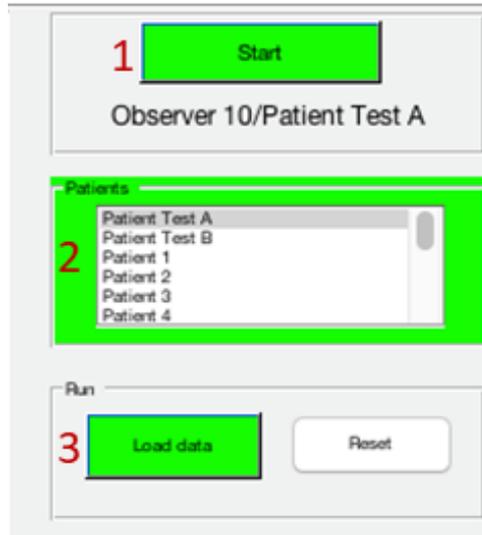


Figure 8: First step in the observatory study. The order of selection was crucial for saving the results of the study in an excel file — 1) observers started the study and identified themselves, 2) observers selected the patient, 3) patient data was loaded.

Afterwards, the observers were shown two sets of prostate images— T2-weighted (T2w), used for target delineation, and MEGRE, used for GFM identification (Figure 9). The MEGRE images had multiple echoes, meaning that each physical slice position was represented by eight images with individual echo times. The task of the observer was to browse through the images and multiple echoes to locate the GFMs. The locating was performed in both the T2w images and the MEGRE images.

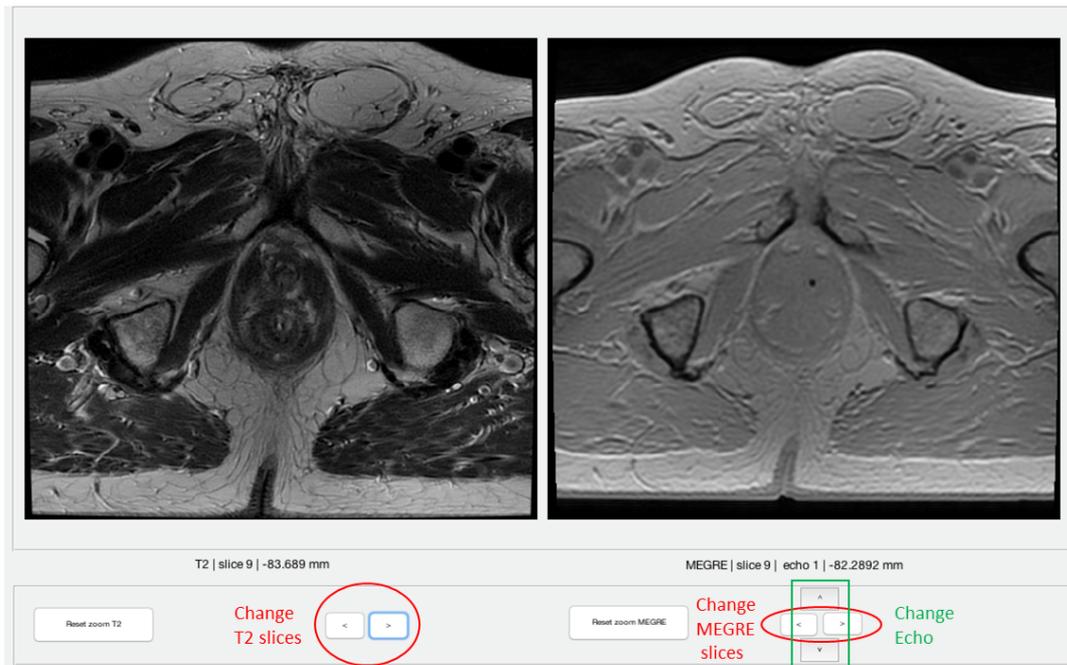


Figure 9: The observer was shown the T2w images on the left side, the MEGRE images were displayed on the right side. The slices and the echoes for the corresponding sets could be changed using the arrows, or alternatively the echoes could be changed with a mouse scroll or keyboard arrows. The slice location in mm, with respect to the MRI coordinate system, of each image was shown next to the slice number. This represented the physical position of the slice with respect to the isocenter, set for correlation.

When positions of GFMs were located, the observers could save the coordinates. The green circle in Figure 10 shows an example of a marker selection. The circle could have been shown or hidden by clicking on a button under the images. This was done to provide the observers a quick overview of their selections.

In some patients one of the markers could have migrated. Therefore, an option— "No marker found"— was included for such cases. After marking the location of the markers, the observers needed to evaluate their selections, by choosing one of the options 1 (not certain)– 5 (certain). This was performed both for T2w and MEGRE images. The process is shown in Figure 10. The markers for each image sequence were labeled 1-3, where 1 was the most inferior and 3 was the most superior one. The slices of the selected markers were shown in a table below the images, shown in Figure 10.

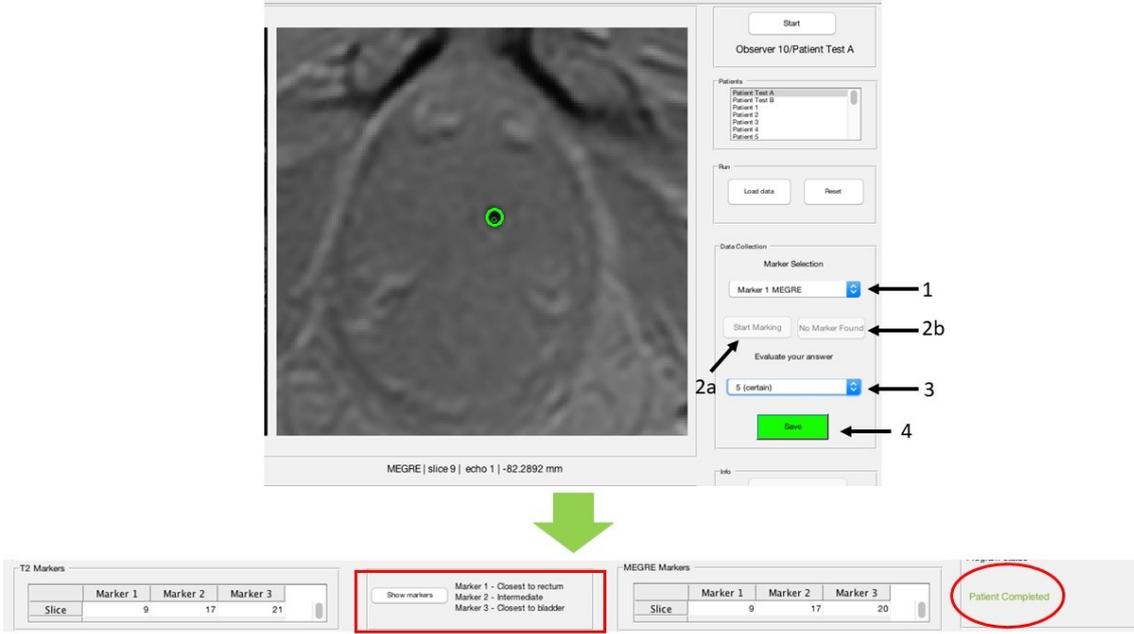


Figure 10: After locating the markers the observers needed to 1) select the located marker, 2a) mark the coordinates or 2b) click on "No Marker Found" in case of no located marker, 3) evaluate their selection, 4) save answer. The name of the completed marker in the list was updated by adding (Defined). Show markers, or the space button, could be used to hide/show selected position indicators.

Once, all the markers were defined, the patient was completed as it is shown in Figure 10. The observers were provided a list with the patients, where they could mark the completed ones.

After completion, the marked coordinates and the evaluation results were automatically saved in an excel file. Furthermore, the coordinates selected by the observer were converted from Matlab image pixels (*Row*, *Column*, *Slice*) to mm measure in the MRI coordinate system (X_{MRI} , Y_{MRI} , Z_{MRI}) using the information in Table 3 and were saved in the result file.

Table 3: Spatial resolution of the T2-weighted and MEGRE images.

	T2-weighted	MEGRE
pixel spacing- x , y (mm)	0.43	0.47
slice thickness- z (mm)	2.80	2.80

The reference coordinate system during the study was the MRI's, which was given in mm. The conversion was done using this fixed coordinate system. The MATLAB images were given in terms of rows and columns (512x512 in this case). The conversion formula was:

$$\begin{aligned}
 X_{MRI} &= X_{Image} + X_{Pixel\ Spacing} \times Column \\
 Y_{MRI} &= Y_{Image} + Y_{Pixel\ Spacing} \times Row \\
 Z_{MRI} &= Z_{Image} + Z_{Slice\ Thickness} \times Slice
 \end{aligned}$$

where $X_{Image}, Y_{Image}, Z_{Image}$ are defined by the coordinates of the upper left hand pixel of the image, which is the first transmitted voxel and the (1,1) pixel in the Matlab image. An example of a coordinate conversion is provided based on the coordinate shown in Figure 11:

$$\begin{aligned} X_{MRI} &= -6.74 = -120.36 + 0.43 \times 264.42 \\ Y_{MRI} &= 3.69 = -108.68 + 0.43 \times 261.50 \\ Z_{MRI} &= -52.89 = -80.89 + 2.80 \times 10 \end{aligned}$$

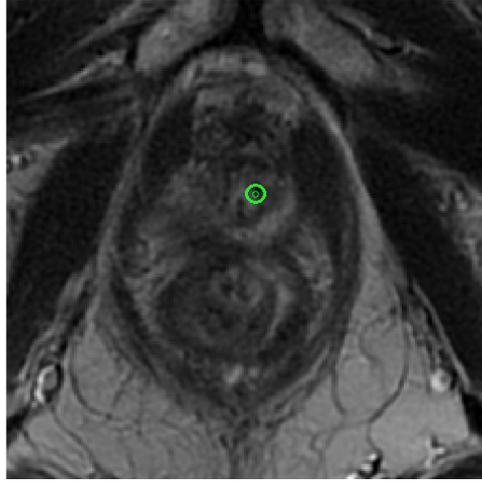


Figure 11: A marker has been selected at the coordinates [261.50, 264.42, 10] with image position [-120.36, -108.68, -80.89], which were converted into MRI coordinates [-6.74, 3.69, -52.89] in mm.

The absolute internal distances between the positions of the selected fiducial markers in mm were calculated and saved in the excel file using the formulas:

$$\begin{aligned} \Delta_{21} &= \text{Marker 2} - \text{Marker 1} = \sqrt{(X_{MRI_2} - X_{MRI_1})^2 + (Y_{MRI_2} - Y_{MRI_1})^2 + (Z_{MRI_2} - Z_{MRI_1})^2} \\ \Delta_{31} &= \text{Marker 3} - \text{Marker 1} = \sqrt{(X_{MRI_3} - X_{MRI_1})^2 + (Y_{MRI_3} - Y_{MRI_1})^2 + (Z_{MRI_3} - Z_{MRI_1})^2} \\ \Delta_{32} &= \text{Marker 3} - \text{Marker 2} = \sqrt{(X_{MRI_3} - X_{MRI_2})^2 + (Y_{MRI_3} - Y_{MRI_2})^2 + (Z_{MRI_3} - Z_{MRI_2})^2} \\ \text{Total distance between markers} &= \Delta_{21} + \Delta_{31} + \Delta_{32} \end{aligned}$$

where $X_{MRI_1}, X_{MRI_2}, X_{MRI_3}$ are the x-coordinates, $Y_{MRI_1}, Y_{MRI_2}, Y_{MRI_3}$ are the y-coordinates, $Z_{MRI_1}, Z_{MRI_2}, Z_{MRI_3}$ are the z-coordinates of Marker 1, Marker 2 and Marker 3 respectively in the fixed coordinate system given in mm.

The evaluation routine in the program was based on an ascending order of GFMs, where marker 1 was the most inferior and marker 3—the most superior fiducial marker. The manually selected GFM positions were sorted in ascending order of the slice thickness for a homogeneous structure in result calculation. Typically, each patient had 3 GFMs. As stated before, there were special cases where the patient had only 2 remaining fiducial markers. The following special cases were taken into account in developing the GUI:

- 2 markers in patient, 2 markers defined by observer. This situation was solved by allowing an option "No marker found" instead of "Start Marking";

- 2 markers in patient, 3 markers defined by observer. This situation was solved by removing the last marker. The results were flagged for attention and checked manually to verify if the remaining markers were correct;
- 3 markers in patient, 2 markers defined by observer. This situation was solved by adding not defined coordinates for the last marker, which created acceptable calculation conditions for the software. The markers were represented by matrices and calculations could not be performed if the matrix sizes did not match;

3.1.3 Data analysis

Duration of the identification process for each observer per patient has been measured during the manual observatory study. The observers were not informed about this measurement to avoid any bias that could be inferred if the observer felt stressed.

The distance between manually selected marker positions and reference coordinates was analyzed. The reference coordinates were defined in the T2-images by using a combination of MR- and CT-images. CT-images were considered gold standard. The MRI reference coordinates for the T2w images were converted to corresponding reference coordinates in MEGRE images. This conversion was not corrected for possible patient movement between the MRI acquisition of the T2w images and the MEGRE images (around 10 minutes scan time).

The difference for each spatial component (x,y,z) was analyzed. Two different analyses, where the condition for true positive (TP) was different, were done for the absolute difference between the manually selected positions and the reference coordinates. In the first analysis, the markers within the distance of 7.5 mm from the reference coordinates were considered TP, whilst in the second analysis a TP was defined as markers within the distance of 5 mm from the reference coordinates. A true negative (TN) was defined as 1 for each patient. This was done due to the fact that a fake signal void has been inserted in each patient in the automatic method. Even though this was not done for the human observatory study, it was taken into account in the result analysis to be able to have a comparison between the human observatory study results and the automatic method results. An incorrectly marked marker was defined as false positive (FP), whilst a missed marker was defined as a false negative (FN).

The detection performance metrics of the human observatory study were calculated as:

$$Sensitivity = \frac{TP}{(TP + FN)} \quad (3.4)$$

$$Specificity = \frac{TN}{(TN + FP)} \quad (3.5)$$

$$Accuracy = \frac{(TN + TP)}{(TN + TP + FN + FP)} \quad (3.6)$$

The mean values and the standard deviations of the detection performance metrics were determined using the individual results from each observer.

To determine the spatial accuracy of the GFM observatory study, the absolute difference in the distances between marker pairs was calculated:

$$\begin{aligned} & (M_2 - M_1)_{Reference} - (M_2 - M_1)_{Manual} \\ & (M_3 - M_1)_{Reference} - (M_3 - M_1)_{Manual} \\ & (M_3 - M_2)_{Reference} - (M_3 - M_2)_{Manual} \end{aligned} \quad (3.7)$$

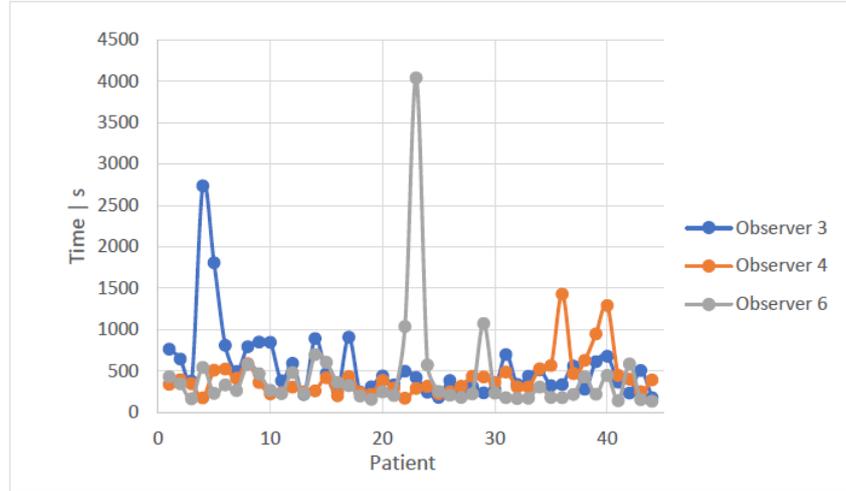
The reference pairs were referred to the markers defined by the reference coordinates using CT- and MR-images. The manual pairs were defined by the manually selected marker positions in both T2w and MEGRE images. The mean values and standard deviation for the absolute spatial difference of marker pairs were calculated.

The results from these comparisons were compared to the results obtained from the automatic identification method.

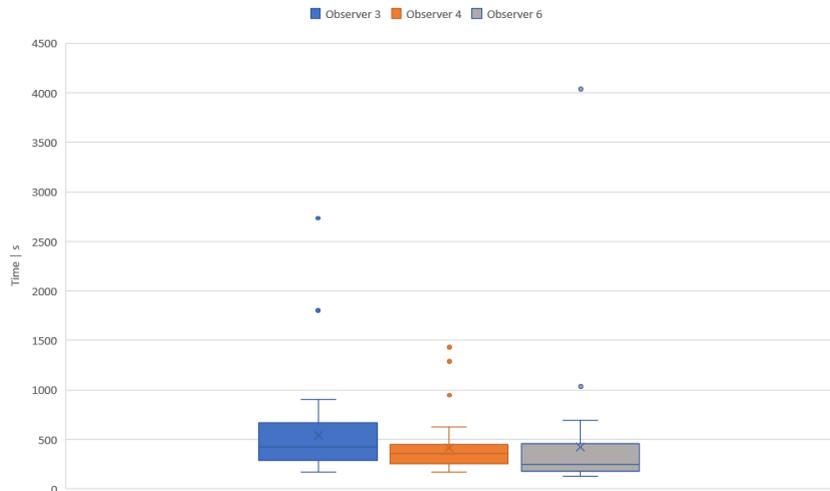
4 Results

4.1 Duration of identification process

The duration of the identification process of each observer per patient is presented in Figure 12.



(a) Duration of identification process for each patient. The y-axis represents the time in units of seconds and the x-axis is representing the corresponding patient.



(b) Distribution of the duration. The y-axis represents the time in units of seconds.

Figure 12: Distribution of identification process duration of each observer per patient.

The patients where the observers took the shortest and the longest time to define marker positions were analyzed to explain the possible causes in identification duration. An example is shown in Figure 13 and other patients can be found in Appendix B.

In Figure 13, the patient had rectal gas close to the prostate. The effect of the rectal gas on the MEGRE images is increasing with increasing echo time, whilst in the T2w image the effect is small. This is due to the fact that the T2w images are spin echo

based, whilst the MEGRE images are gradient echo based. Furthermore, the noise in the MEGRE images is increasing at larger echo times.

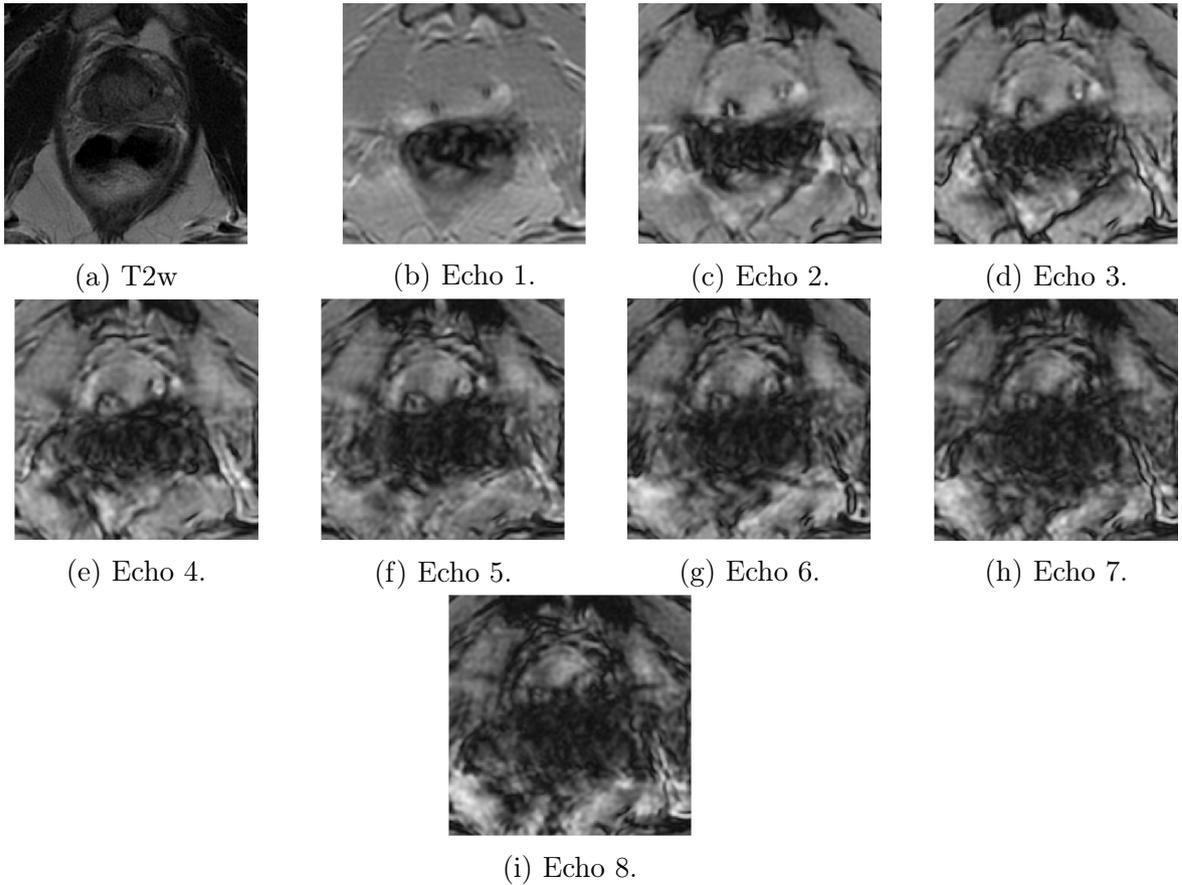


Figure 13: Images of Patient 22, where Observer 4 took the shortest time. In the echo images, image artifacts from rectal gas can be seen. The effect from the gas increases with increasing echo time. The noise in the images is also increasing with increasing echo time.

The mean values of the time distribution (shown in Table 4) for observers 3, 4 and 6 are 544.1 ± 447.4 , 415.9 ± 254.0 s and 423.4 ± 598.3 s respectively.

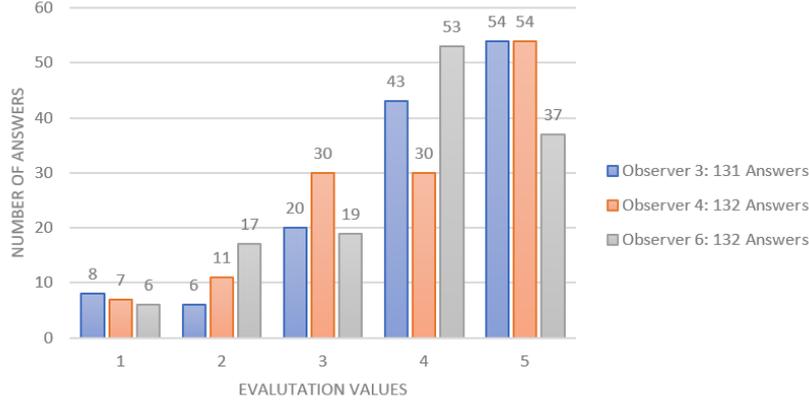
Table 4: The mean, standard deviation (σ), maximum and minimum values for the duration of each patient per observer. The values based on the data shown in Figure 12.

Observer	Mean s	σ s	Max s	Min s
Observer 3	544.1	447.4	2734.8	175.6
Observer 4	415.9	254.0	1429.3	169.4
Observer 6	423.4	598.3	4039.7	133.2

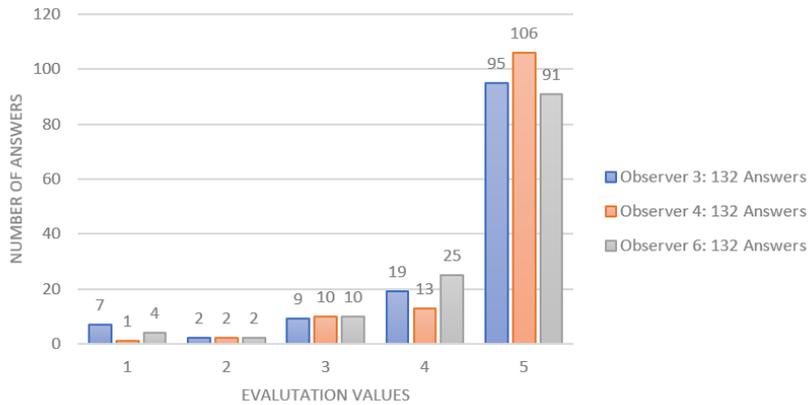
4.2 Assessment of the observer performance

4.2.1 Evaluation score

Distribution of the evaluation score, defined after each marker selection during the human observatory study, for T2w and MEGRE images, is shown in Figure 14.



(a) Evaluation scores for the manually selected marker positions in the T2w images.



(b) Evaluation scores for the manually selected marker positions in the MEGRE images.

Figure 14: Evaluation score distributions for T2w and MEGRE images from 3 observers. Score 1 was corresponding to an uncertain determination of the selected marker position, and score 5 was corresponding to a certain determination of the selected marker position.

4.2.2 Gold fiducial marker spatial positions

The absolute difference between the spatial locations of the manually selected marker positions and reference coordinates was calculated for T2w images. This analysis was not done for MEGRE images. These differences were evaluated for two different cases for each observer. First case was when a TP marker was defined as a manually selected marker within the distance of 7.5 mm from the reference coordinates, and the second was 5 mm from the reference coordinates. The number of patients who had all their GFMs correctly defined is given in Table 5 for each observer. Observers 3 and 6 mistaken calcification for a marker in 1 patient. 1 patient for observers 3 and 6 had incorrectly defined markers due to some undefined reason. Rest of the patients, whose GFMs were incorrectly defined, were due to large difference between manually selected marker positions and reference coordinates in the slice thickness. More information can be found in Appendix C.

Table 5: The number of patients who had their GFMs correctly defined for the two different analyses, with the criteria 7.5 mm from the reference coordinates and 5 mm from the reference coordinates is shown for Observer 3,4 and 6. The patients with the incorrectly selected GFMs are divided into two groups. First ones are where the mistake was due to a calcification and the second group are where the mistake was due to another reason.

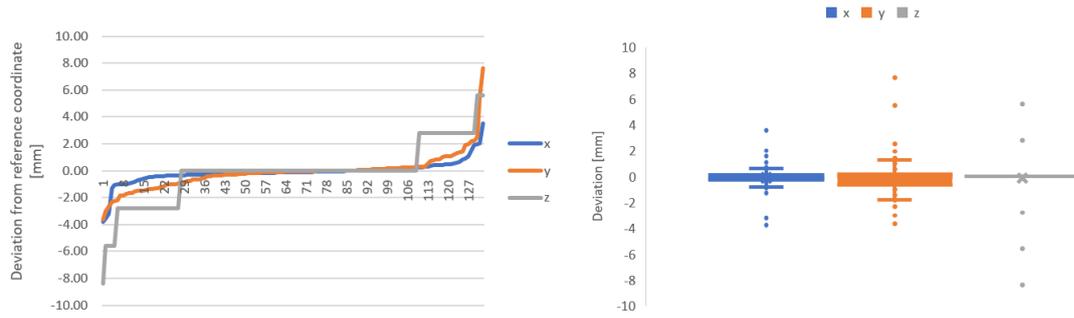
		All GFM correctly defined	Wrong due to calcification	Another reason
Observer 3	7.5	41	1	2
	5	38	1	5
Observer 4	7.5	44	-	-
	5	42	-	2
Observer 6	7.5	43	1	-
	5	39	1	4

The mean sensitivity, specificity and accuracy of the method were calculated and presented in Table 6.

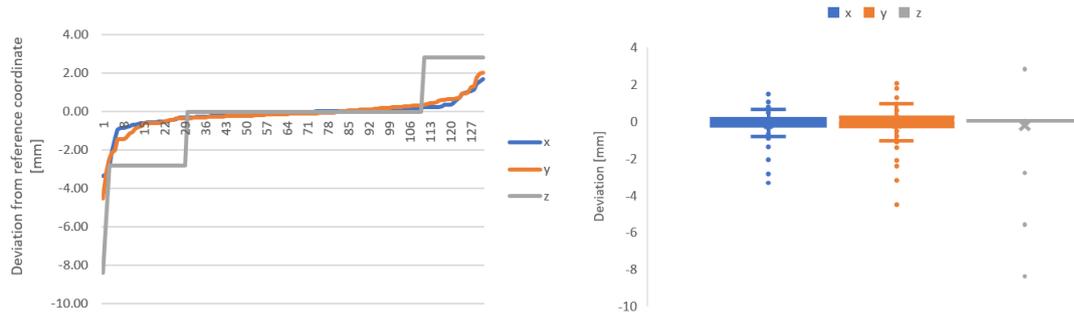
Table 6: The mean values and one standard deviation for the sensitivity, specificity and accuracy of the human observatory study for Observer 3,4 and 6. The performance matrices were evaluated for 2 different true positive criteria for T2w images.

Distance mm	Sensitivity	Specificity	Accuracy
7.5	98.33 ± 1.53 %	95.33 ± 4.16 %	97.33 ± 2.31%
5	96.00 ± 1.73 %	89.67 ± 5.69 %	94.33 ± 3.21%

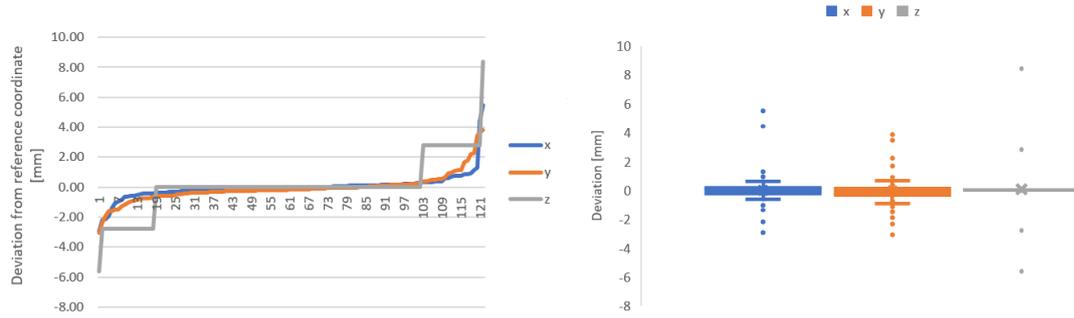
The variation in the differences between the spatial coordinates of the manually marked positions and references coordinates are shown in Figure 15 for T2w images. The differences in each spatial component (x,y,z) were calculated and combined for all observers to analyze the role of each component in determining the deviations from the reference coordinate of the answers from the human observatory study.



(a) Marker 1



(b) Marker 2



(c) Marker 3

Figure 15: The distribution of the differences between manually selected marker positions and reference coordinates in the spatial components (x,y,z) for each selected marker in the T2w images. The x-axis represents the manually selected markers in the first graph. The box and whisker plots represents the distribution of the deviations. The results are combined for all the observers.

4.3 Assessment of the observer accuracy

The absolute internal distances between manually selected marker positions were calculated and compared to the absolute internal distances of the reference coordinates for T2w and MEGRE images. The mean absolute internal distance difference between marker pairs was 1.10 ± 1.04 mm and 0.95 ± 0.87 mm for T2w and MEGRE images respectively. A more detailed account of the absolute internal distance results for each observer are shown in Tables 7 - 9.

Table 7: Mean, standard deviation (σ), maximum and minimum values of absolute internal distance differences and number of measurements (N)— dependent on the number of TP markers— for Observer 3. The reference distances were determined by using CT-images and the manual distances are from the human observatory study. The results are based on the number of TP within the distance of 7.5 mm from the reference coordinates.

	Internal Distance	Mean (mm)	σ (mm)	Max (mm)	Min (mm)	N
T2w	$(M_2 - M_1)_{Reference} - (M_2 - M_1)_{Manual}$	0.98	1.11	6.25	0.02	41
	$(M_3 - M_1)_{Reference} - (M_3 - M_1)_{Manual}$	1.31	0.96	4.35	0.02	38
	$(M_3 - M_2)_{Reference} - (M_3 - M_2)_{Manual}$	0.70	0.60	2.89	0.01	40
	All of the above	0.99	0.94	6.26	0.01	119
MEGRE	$(M_2 - M_1)_{Reference} - (M_2 - M_1)_{Manual}$	0.93	0.95	5.88	0.07	43
	$(M_3 - M_1)_{Reference} - (M_3 - M_1)_{Manual}$	1.36	1.20	4.68	0.01	40
	$(M_3 - M_2)_{Reference} - (M_3 - M_2)_{Manual}$	0.46	0.27	1.08	0.01	41
	All of the above	0.91	0.96	5.88	0.01	124

Table 8: Mean, standard deviation (σ), maximum and minimum values of absolute internal distance differences and number of measurements (N)— dependent on the number of TP markers— for Observer 4. The reference distances were determined by using CT-images and the manual distances are from the human observatory study. The results are based on the number of TP within the distance of 7.5 mm from the reference coordinates.

	Internal Distance	Mean (mm)	σ (mm)	Max (mm)	Min (mm)	N
T2w	$(M_2 - M_1)_{Reference} - (M_2 - M_1)_{Manual}$	1.04	1.12	6.19	0.03	44
	$(M_3 - M_1)_{Reference} - (M_3 - M_1)_{Manual}$	1.62	1.11	4.18	0.10	41
	$(M_3 - M_2)_{Reference} - (M_3 - M_2)_{Manual}$	0.68	0.59	2.50	0.02	41
	All of the above	1.11	1.04	6.19	0.02	126
MEGRE	$(M_2 - M_1)_{Reference} - (M_2 - M_1)_{Manual}$	0.85	0.84	3.40	0.01	44
	$(M_3 - M_1)_{Reference} - (M_3 - M_1)_{Manual}$	1.44	0.97	3.44	0.05	41
	$(M_3 - M_2)_{Reference} - (M_3 - M_2)_{Manual}$	0.51	0.36	1.53	0.02	41
	All of the above	0.93	0.86	3.44	0.01	126

Table 9: Mean, standard deviation (σ), maximum and minimum values of absolute internal distance differences and number of measurements (N)— dependent on the number of TP markers— for Observer 6. The reference distances were determined by using CT-images and the manual distances are from the human observatory study. The results are based on the number of TP within the distance of 7.5 mm from the reference coordinates.

	Internal Distance	Mean (mm)	σ (mm)	Max (mm)	Min (mm)	N
T2w	$(M_2 - M_1)_{Reference} - (M_2 - M_1)_{Manual}$	0.99	0.95	3.94	0.03	42
	$(M_3 - M_1)_{Reference} - (M_3 - M_1)_{Manual}$	1.43	1.16	6.07	0.02	40
	$(M_3 - M_2)_{Reference} - (M_3 - M_2)_{Manual}$	0.91	1.25	6.22	0.02	41
	All of the above	1.10	1.14	6.22	0.02	123
MEGRE	$(M_2 - M_1)_{Reference} - (M_2 - M_1)_{Manual}$	0.86	0.59	2.62	0.10	43
	$(M_3 - M_1)_{Reference} - (M_3 - M_1)_{Manual}$	1.39	1.03	4.20	0.02	41
	$(M_3 - M_2)_{Reference} - (M_3 - M_2)_{Manual}$	0.79	0.63	3.28	0.04	42
	All of the above	1.00	0.81	4.20	0.02	126

5 Discussion & Conclusions

During this project, the performance of human observers in detecting the position of gold fiducial markers in T2w and MEGRE images was tested. The human observatory study was conducted by 3 independent observers. The observer results were compared to the results from an automatic method. Furthermore, the MEGRE image sequence was verified for use in clinical applications for identification of gold fiducial markers.

The mean value for the identification process duration of each observer was close to each other. However, the standard deviations in the duration were large. This was due to differences in patient anatomy. The duration of the identification process for each patient is hard to predict due to the fact that each patient is a unique case. Consequently, the distribution of the time data has a large deviation. The patients which took the longest and shortest time to identify marker positions were analyzed for possible reasons. The observers took longer time for patients who had calcifications located close to the GFMs or for patients who had only 2 remaining GFMs. An example was shown in Figure 13, where rectal gas close to the markers was distorting the image, making the identification of fiducial markers at higher echoes challenging. More examples are provided in Appendix B. Additionally, in some cases the shape and the size of the calcification was very similar to a void from a GFM. Therefore, it could have taken a longer time to decide which one is the true fiducial marker. This could have also be seen in the evaluation scores, where the observer selected lower evaluation score for the more challenging markers. The cases that took shorter time did not have major or any calcifications, which made the selection process faster.

Looking at the evaluation scores shows a majority of fives, meaning that the observers were certain of their selections. The similarity in the evaluation scores between the observers, can also hint that the patient set was consistent and equally challenging. This is supported by the mean time, which is similar for all 3 observers. Furthermore, the evaluation scores for MEGRE images were higher— higher majority of fives— hinting that locating the markers was easier when multiple echoes were provided. This result was expected as the MEGRE images were aimed for GFM identification whilst the T2w images

were not. The distribution in the evaluation results shows that the MEGRE images are fulfilling their purpose. Validating the sequence for clinical application was an objective of this project, which has been achieved.

The analysis of gold fiducial marker spatial positions were done only in the T2w images. This was due to the fact that the reference coordinates for the MEGRE images were converted from the reference coordinates of the T2w image, using the MRI reference coordinates. Using converted reference positions for MEGRE could lead to additional positioning uncertainties. Two different analyses were done for the spatial positions of the GFMs. In the first case, the condition for a TP was a marker within the distance of 7.5 mm from the reference coordinates. In the second case, the condition was a marker within the distance of 5 mm from the reference coordinates. The 5 mm distance was reflecting the gold fiducial marker size, whilst the 7.5 mm distance took into account the possible errors in reference coordinate position and the potential partial volume artifact which made the marker visible in more slices. The results for sensitivity, specificity and accuracy were 98.33 ± 1.53 %, 95.33 ± 4.16 %, 97.33 ± 2.31 % and 96.00 ± 1.73 %, 89.67 ± 5.69 %, 94.33 ± 3.21 % for 7.5 mm and 5 mm respectively. The results from the automatic GFM identification method showed 84 %, 74 % and 81 % for sensitivity, specificity and accuracy respectively. [3] The human observatory study results were superior to the results of the automatic method. Furthermore, the performance of the observers had decreased during the second analysis with the condition for a TP within the distance of 5 mm from the reference coordinate. The major reason behind this is the spatial difference in the slice direction of the markers, see Appendix C. This is also shown in Figure 15, where the z spatial component is the main contributor to deviation between manually selected marker positions and reference coordinates with a maximum value of 8.40 mm. This was in agreement with the expectation, as the slice thickness was corresponding to 2.80 mm, whilst the in plane resolution was 0.43×0.43 mm which was roughly 6.5 times smaller than the slice thickness. The distribution of the variations in the x - and y - coordinates was similar for all 3 markers.

The absolute internal distance difference between the manual and reference marker pairs was calculated for the TP markers for the difference of 7.5 mm from the reference coordinates. This analysis was done for both T2w and MEGRE images, due to the fact the distance between marker pairs is not affected by the converted reference coordinates. In the human observatory study, the mean absolute internal distance difference between marker pairs was 1.07 ± 1.04 mm and 0.95 ± 0.87 mm for T2w and MEGRE images respectively. In the automatic method, the mean absolute internal distance difference between the detected GFM and the reference markers was 1.28 ± 1.25 mm. [3] The average result of the human observatory study is superior to the automatic method. Another observation made during the result analysis was that the largest internal distance difference in both the human observatory study and the automatic method was for the M_3 - M_1 case. The distance between these markers is the largest, therefore the presence of a systematic error in the detection will have the largest effect on this distance. Furthermore, marker 1 is the closest to the rectum, therefore artifacts, such as rectal gas, could have made the identification of the first marker harder.

One of the purposes of the automatic method was to eliminate the observer based inaccuracies, which are shown by the variations of the mean absolute internal distance difference for the observers. For Observer 3, 116/122 and 119/125 of the manually selected marker absolute internal differences for T2w and MEGRE images were within the distance of 3 mm from the reference coordinates, which is a clinically accepted distance. The

numbers were 117/126 and 122/126 for Observer 4, 117/125 and 120/125 for Observer 6 in T2w and MEGRE images respectively. This shows the observer based variations might exist in marker identification, which can be avoided if an automatic method is used. The results of the study are based on a limited number of observers. Therefore, the variations in the mean and the standard deviations are not big. To obtain a better overview of human observer GFM identification, the study needs to be conducted with more observers.

However, having 3 observers was enough to verify the MEGRE image sequence. The mean absolute internal distance difference performance for MEGRE images was superior than T2w images. This shows that the identification of the marker positions is more accurate in MEGRE images, which makes them suitable for clinical use.

6 Outlook

During this study the GFM identification sequence has been validated and it started being used for an MRI-only treatment at the Radiotherapy clinic at Skåne University hospital in Lund. Furthermore, the results of the study will be used for a construction of a tool which will aid the user in identifying the gold fiducial markers without the need of going through all images.

Furthermore, the human observatory study will be conducted with more observers. For this purpose, the Matlab GUI used in the study will be further improved. The workflow of the interface will be simplified by adapting it to each observer.

Additionally, the detection performance metrics of the study will be analyzed using free response operating characteristic analysis. This will improve the pixel-based analysis by including the evaluation score of the observers.

References

- [1] M. Ng, E. Brown, A. Williams, M. Chao, N. Lawrentschuk, and R. Chee, “Fiducial markers and spacers in prostate radiotherapy: current applications,” *BJU International*, vol. 113, pp. 13–20, March 2014. doi: 10.1111/bju.12624.
- [2] J. Jonsson, *Integration of MRI into the radiotherapy workflow*. Department of Radiation Sciences, Radiation Physics: Umeå University, 2013. ISBN:978-91-7459-621-2.
- [3] C. Gustafsson, “Automatic identification of gold fiducial markers in mri target delineation images intended for prostate radiotherapy without the need for image registration.” Abstract for 5th MR in RT Symposium 2017, Sydney, Australia, 22.-23.June, 2017.
- [4] “Radiation therapy.” Britannica Academic: Encyclopædia Britannica <http://academic.eb.com.ludwig.lub.lu.se/levels/collegiate/article/radiation-therapy/62401>, 2015-11-06. Accessed: 2017-04-11.
- [5] “Treatment process.” UNC School of Medicine: Department of Radiation Oncology <https://www.med.unc.edu/radonc/patient/treatments/ck/process>. Accessed: 2017-04-21.
- [6] C. Westbrook, C. K. Roth, and J. Talbot, *MRI in Practice*. Wiley- Blackwell, 4th edition ed., 2011. ISBN:978-1-4443-3743-3.
- [7] G. C. Pereira, M. Traughber, and R. F. M. Jr, “The role of imaging in radiation therapy planning: Past, present, and future,” *BioMed Research International*, vol. 2014, April 2014. Article ID 231090.
- [8] A. Johansson, M. Karlsson, J. Yu, T. Asklund, and T. Nyholm, “Voxel-wise uncertainty in ct substitute derived from mri,” *Medical Physics*, vol. 39, p. 3283–3290, June 2012. doi: 10.1118/1.4711807.
- [9] G. J. Bansal, “Digital radiography. a comparison with modern conventional imaging,” *Postgraduate Medical Journal*, vol. 82, pp. 425–428, July 2006. doi:10.1136/pgmj.2005.038448.
- [10] L. E. Romans, *Computed Tomography for Technologists*. Wolters Kluwer Health, 2011. ISBN:978-0-7817-7751-3.
- [11] M. K. Gupta, A. Dhaliwal, and K. Kahlon, “Effect of external magnetic field on attenuation coefficient for magnetic substances,” *Applied Radiation and Isotopes*, vol. 95, pp. 188–192, January 2015. Article ID 231090.
- [12] P. Mah, T. Reeves, and W. McDavid, “Deriving hounsfield units using grey levels in cone beam computed tomography,” *Dentomaxillofac Radiol.*, vol. 39, pp. 323–335, September 2010. doi:10.1259/dmfr/19603304.
- [13] B. Bransden and C. Joachain, *Quantum Mechanics*. Pearson Education Limited, 2nd edition ed., 2000. ISBN:978-0-582-35691-7.
- [14] “Photoelectric effect.” Britannica Academic: Encyclopædia Britannica <https://www.britannica.com/science/photoelectric-effect>. Accessed: 2017-05-01.

- [15] M. J. Krasin, L. S. Constine, D. Friedman, and L. B. Marks, “Radiation-related treatment effects across the age spectrum: Differences and similarities or what the old and young can learn from each other,” *Semin Radiat Oncol.*, vol. 20, pp. 21–29, January 2010. doi: 10.1016/j.semradonc.2009.09.001.
- [16] “Magnetic susceptibility.” Britannica Academic: Encyclopædia Britannica <http://academic.eb.com.ludwig.lub.lu.se/levels/collegiate/article/magnetic-susceptibility/50034>, 2017-02-03. Accessed: 2017-04-11.
- [17] J. Olsrud, S. B. Jimmy Lätt, B. Romner, and I. M. Björkman-Burtscher, “Magnetic resonance imaging artifacts caused by aneurysm clips and shunt valves: Dependence on field strength (1.5 and 3 t) and imaging parameters,” *Journal of Magnetic Resonance Imaging*, vol. 22, pp. 433–437, September 2005. doi: 10.1002/jmri.20391.
- [18] “Graphical user interface (gui).” Britannica Academic: Encyclopædia Britannica <http://academic.eb.com.ludwig.lub.lu.se/levels/collegiate/article/graphical-user-interface/109589>, 2017. Accessed: 2017-04-11.

Appendices

A Manual for the human observatory study

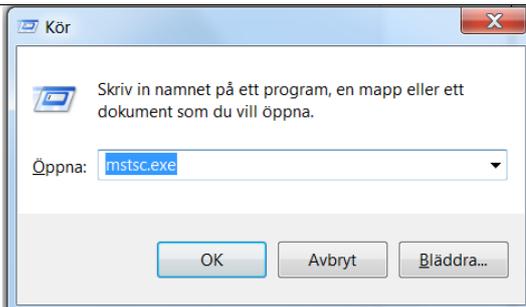
INSTRUCTIONS FOR OBSERVATORY STUDY

How well can a multi echo gradient echo (MEGRE) MRI acquisition sequence depict and differentiate intraprostatic gold fiducial markers?

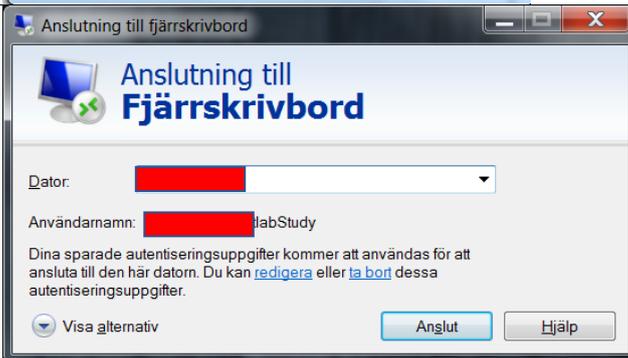
Christian Gustafsson, Sevgi Emin

Skåne University Hospital | Dept Haematology, Oncology and Radiation Physics

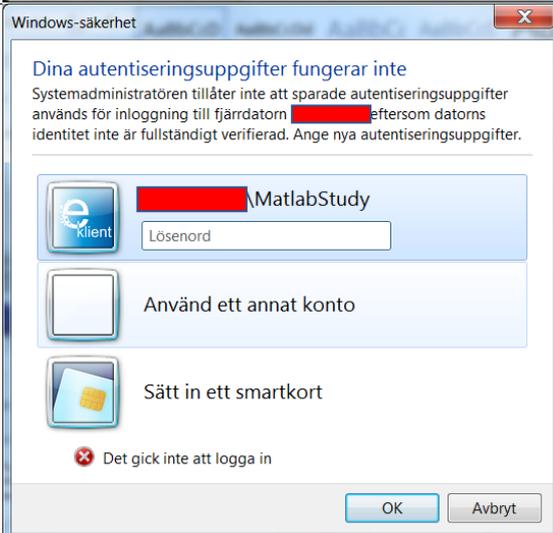
Start your PC and login with your RSID.



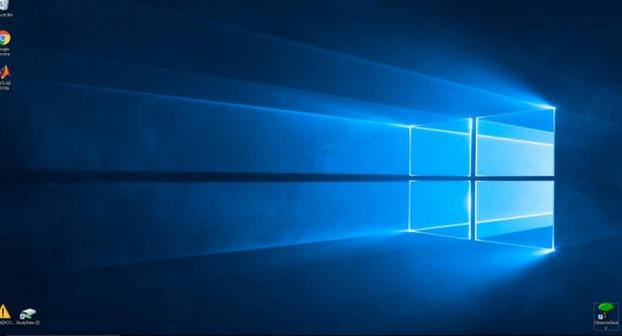
Press Windows button + R and a run dialog pops up. Write mstsc.exe



Enter [redacted] to connect to the server where the study material is found.

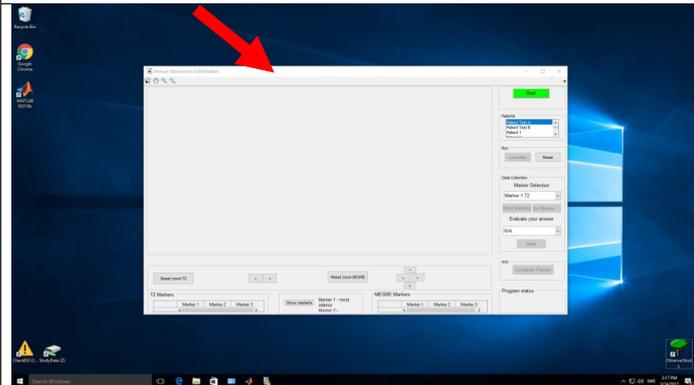


Press use another account / Använd ett annat konto

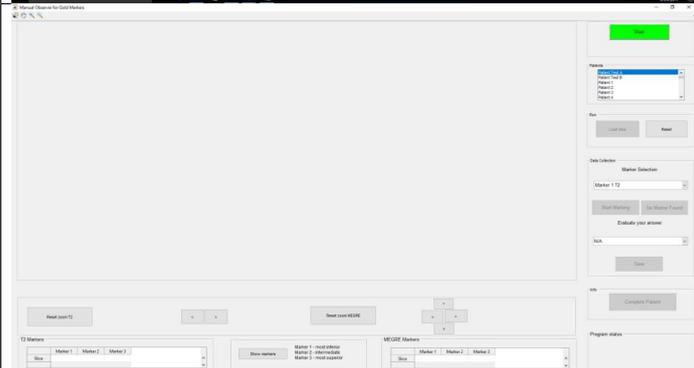
		<p>If question arise, accept certificate.</p>
		<p>Enter username: [redacted] Enter Password: [redacted]</p>
		<p>The remote desktop that is used looks like this.</p>



To start the Observatory study click on the icon.
Wait up to 30 seconds until the program is loaded



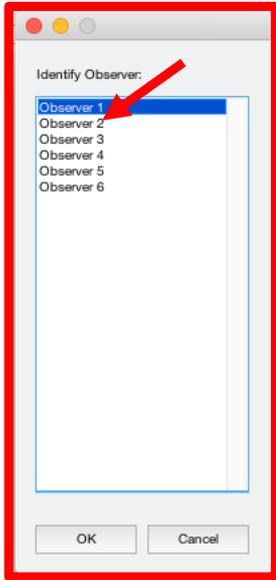
Double click on the banner to make it full-screen.
To make the work flow in the study easier, the following steps will be highlighted in green



Program in full screen mode



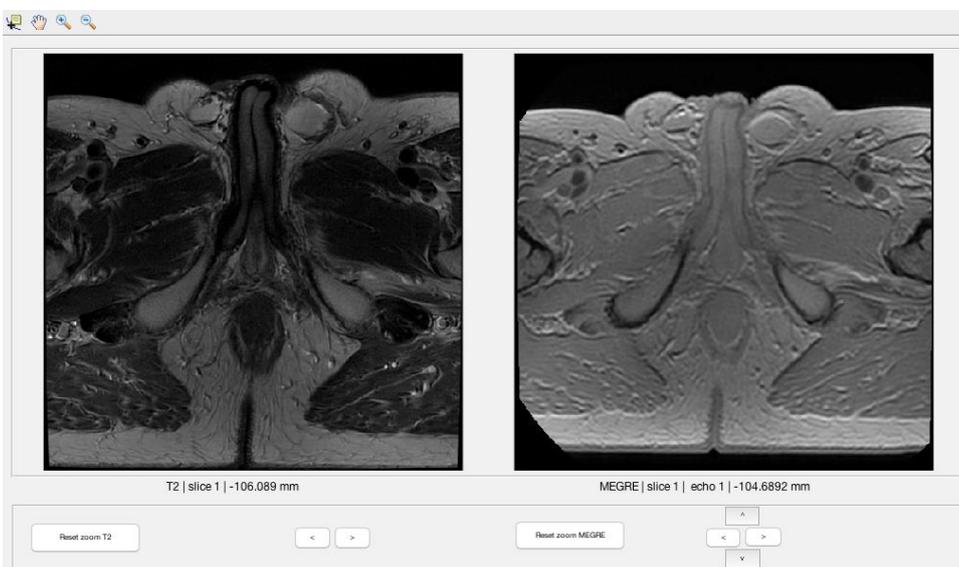
To start the study, click on START on the upper right corner.



Select the Observer number which you have been assigned and click OK. It is very important that you select the right number every time since the results are dependent on this selection.



Select a Patient from the list and click LOAD DATA. If there is a problem with the program, you can simply click RESET to restart the program.



When the patient information is loaded, the T2 image will be displayed on the left side, while the MEGRE is on the right. In the titles under the images the slice and echo information can be found.

Left/right change the slice independently in both images. Up/down changes the displayed echo in MEGRE. The echo can also be changed by using the scroll wheel on the mouse.

Not the distance reported, -106 and -104 mm in this case to see if the slices are approximately on the same spatial position. Clicking the left and right pointers under the respective image changes slice and corresponding distance for each image.

To zoom in a section of the image use the zoom icons on the top left. Click in the image to zoom. Once you select your area click one more time on the zoom icon to finish zooming. **This is important to do if you don't want the zooming function anymore!**

To reset the zoom, you can simply click on the buttons RESET ZOOM T2 and RESET ZOOM MEGRE available below the corresponding images.

Data Collection

Marker Selection

- ✓ Marker 1 T2
- Marker 1 MEGRE
- Marker 2 T2
- Marker 2 MEGRE
- Marker 3 T2
- Marker 3 MEGRE

N/A

Save

Locate your markers on the images and select a marker from the list. Marker 1 T2 is the marker closest to the rectum for the T2 images, whilst Marker 1 MEGRE is the marker closest to the rectum for MEGRE.

Consequently, Marker 2 is the intermediate, whilst Marker 3 is the marker closest to the bladder.

Data Collection

Marker Selection

Marker 1 T2

Start Marking

No Marker Found

Evaluate your answer

N/A

Save

Once your marker is selected click on **START MARKING** and mark your coordinates using the cursor.

When you click on a point a cross marker should appear.

In cases where you can't find a marker instead of clicking **START MARKING**, click **NO MARKER FOUND**.

A marker is often shown in multiple slices. Chose the most central marker.

Data Collection

Marker Selection

Marker 1 T2

Start Marking No Marker Found

Evaluate your answer

✓ N/A
1 (not certain)
2
3
4
5 (certain)

Complete Patient

When you are done with selecting the coordinates of a marker, evaluate your answer. Your answer should be based on whether you are certain that the coordinates you marked are corresponding to a marker. 1= not certain, 5= am certain.

Data Collection

Marker Selection

Marker 1 T2 (Defined)

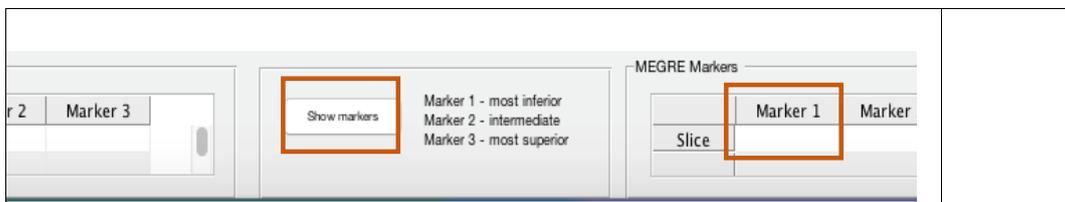
Start Marking No Marker Found

Evaluate your answer

3

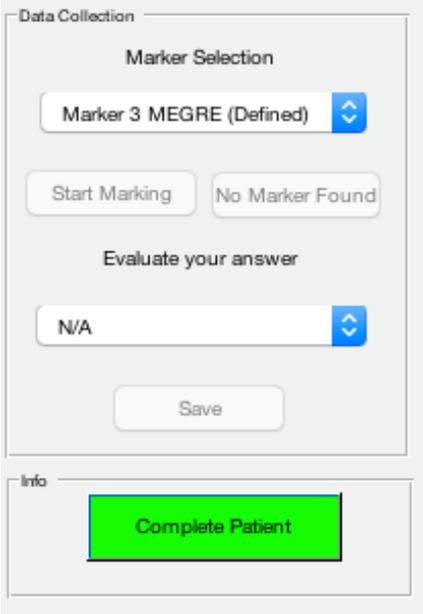
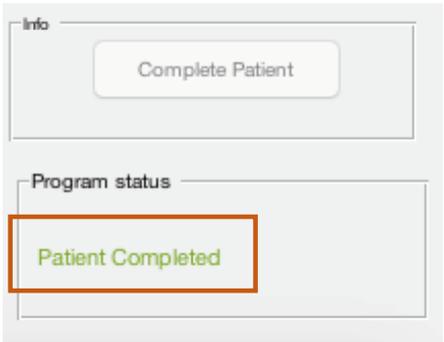
Save

Finish the current marker by clicking on SAVE and continue with the next marker. Once you are done with all the markers, move on to next step.



Every time you click on SAVE the tables for T2 Markers and MEGRE Markers will be updated with the slice number where each marker has been located by you. If you have selected NO MARKER FOUND, NaN will be displayed. Make sure that all the cells are full before you move on to the next step.

The button SHOW MARKERS can be used to display the selected locations of the markers. The same button is used to hide the circles. One can also use the SPACE key for the same purpose.

	<p>When you are done with all the markers. Finish the current patient with COMPLETE PATIENT.</p>
	<p>The patient is completed when “Patient Completed” is displayed in the Program status panel.</p> <p>Mark the finished patient in the list provided. If you have any comments please write them down.</p> <p>Chose and load the next patient.</p>

B Patient cases for min and max identification duration time

The images of the patients leading to maximum and minimum duration of the identification process. The T2w and MEGRE images at 8 different echo times are provided below.

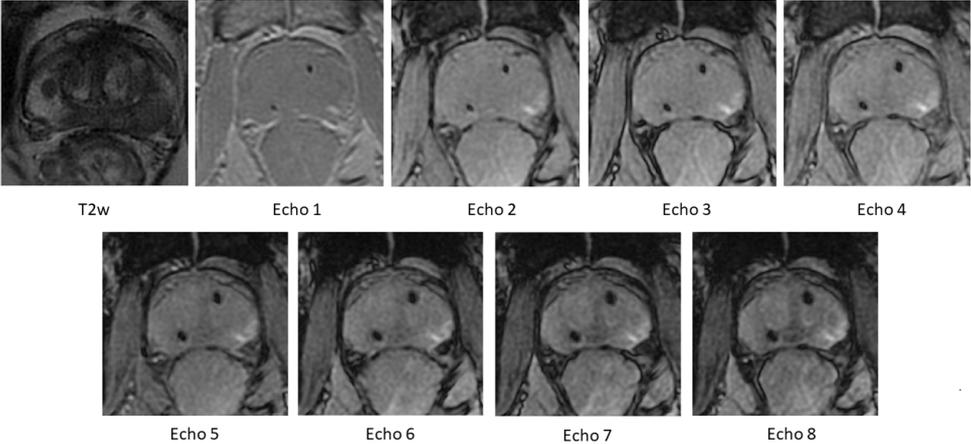


Figure 16: Images of Patient 4, where Observer 3 took the longest time. The T2w images and the MEGRE images at 8 different echo times are shown. The Patient had only 2 remaining GFMs.

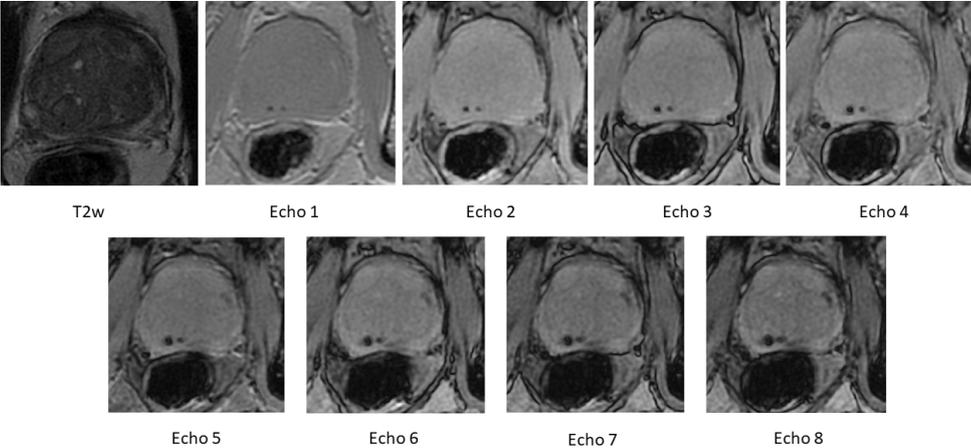


Figure 17: Images of Patient 26, where Observer 6 took the longest time. The T2w images and the MEGRE images at 8 different echo times are shown. As it can be seen in the MEGRE images, the patient has a calcification close to the marker

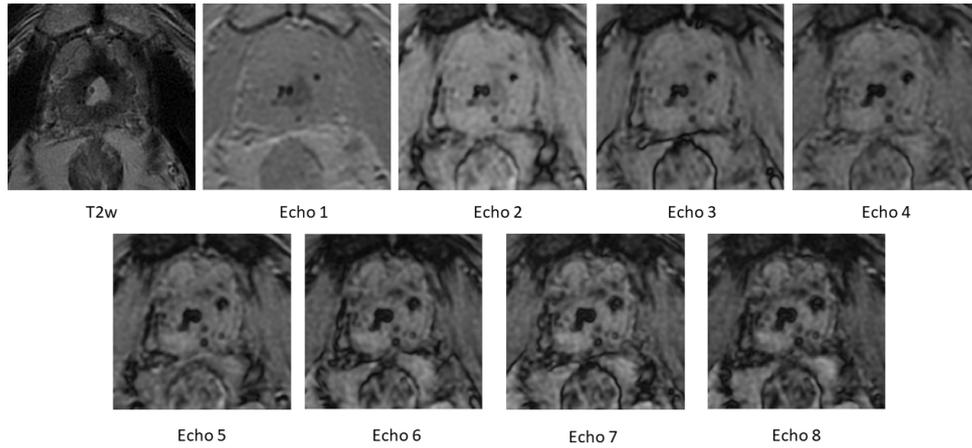


Figure 18: Images of Patient 37, where Observer 4 took the longest time. The T2w images and the MEGRE images at 8 different echo times are shown. As it can be seen the patient had major calcifications making the identification process more challenging.

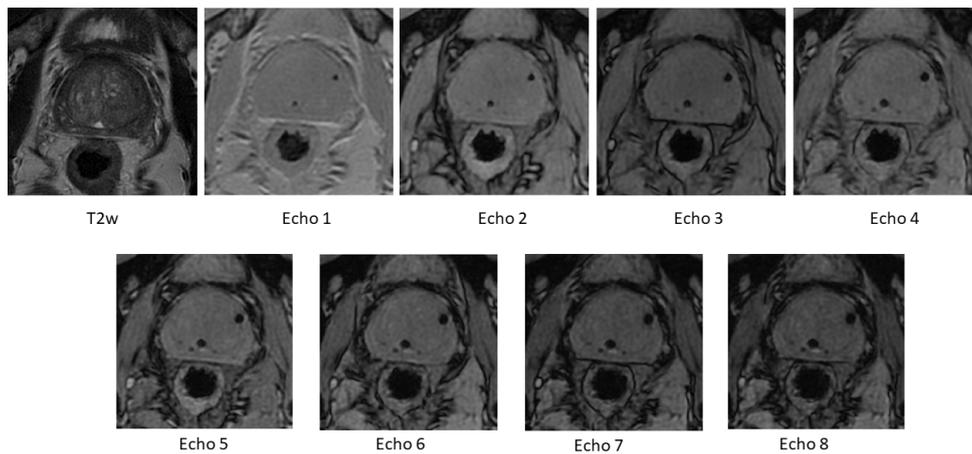


Figure 19: Images of Patient 44, where Observer 3 and 6 took the shortest time. The T2w images and the MEGRE images at 8 different echo times are shown.

C Patient and gold fiducial marker data

The number of TP, TN, FP and FN per patient of each observer are shown in the following tables. Information about number of GFMs in the patient and if the patient had calcifications is also included. The different types of reason for FP and FN marker position selections are color coded. Blue represents a selection of a third marker for patients with 2 GFMs. Orange represents mistakes in the slice selection, whilst yellow is mistakes due to calcifications and pink is other reasons.

Patient	# true of GFMs	Detected Calcification	Observer 3				Observer 4				Observer 6			
			TP	TN	FP	FN	TP	TN	FP	FN	TP	TN	FP	FN
1	3	No	3	1	0	0	3	1	0	0	3	1	0	0
2	3	Yes	3	1	0	0	3	1	0	0	0	1	3	3
3	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
4	2	Yes	2	1	0	0	2	1	0	0	2	1	0	0
5	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
6	3	No	3	1	0	0	3	1	0	0	3	1	0	0
7	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
8	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
9	3	No	3	1	0	0	3	1	0	0	3	1	0	0
10	2	Yes	2	1	1	0	2	1	0	0	2	1	1	0
11	3	No	3	1	0	0	3	1	0	0	3	1	0	0
12	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
13	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
14	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
15	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
16	3	No	3	1	0	0	3	1	0	0	3	1	0	0
17	3	Yes	1	1	0	2	3	1	0	0	3	1	0	0
18	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
19	3	No	3	1	0	0	3	1	0	0	3	1	0	0
20	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
21	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
22	3	No	3	1	0	0	3	1	0	0	3	1	0	0
23	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
24	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
25	3	No	3	1	0	0	3	1	0	0	3	1	0	0
26	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
27	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
28	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
29	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
30	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
31	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
32	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
33	3	No	3	1	0	0	3	1	0	0	3	1	0	0
34	3	Yes	2	1	1	1	3	1	0	0	3	1	0	0
35	3	Yes	2	1	1	1	3	1	0	0	3	1	0	0
36	3	No	3	1	0	0	3	1	0	0	3	1	0	0
37	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
38	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
39	3	No	3	1	0	0	3	1	0	0	3	1	0	0
40	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
41	3	No	3	1	0	0	3	1	0	0	3	1	0	0
42	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
43	2	No	2	1	0	0	2	1	0	0	2	1	0	0
44	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0

Figure 20: The number of TP, TN, FP and FN within 7.5 mm from reference coordinates per patient for each observer. Includes information about calcifications and number of GFMs in patient. Cells highlighted in blue represent a selection of a third marker for patients with 2 GFMs, orange— slice selection mistakes, yellow— mistakes due to calcifications and pink— other reasons.

Patient	# true of GFMs	Detected Calcification	Observer 3				Observer 4				Observer 6			
			TP	TN	FP	FN	TP	TN	FP	FN	TP	TN	FP	FN
1	3	No	3	1	0	0	3	1	0	0	3	1	0	0
2	3	Yes	3	1	0	0	3	1	0	0	0	1	3	3
3	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
4	2	Yes	2	1	0	0	2	1	0	0	2	1	0	0
5	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
6	3	No	3	1	0	0	3	1	0	0	2	1	1	1
7	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
8	3	Yes	2	1	1	1	3	1	0	0	3	1	0	0
9	3	No	3	1	0	0	3	1	0	0	3	1	0	0
10	2	Yes	2	1	1	0	2	1	0	0	2	1	1	0
11	3	No	3	1	0	0	3	1	0	0	3	1	0	0
12	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
13	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
14	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
15	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
16	3	No	3	1	0	0	3	1	0	0	2	1	1	1
17	3	Yes	1	1	0	2	3	1	0	0	3	1	0	0
18	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
19	3	No	3	1	0	0	3	1	0	0	3	1	0	0
20	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
21	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
22	3	No	3	1	0	0	3	1	0	0	3	1	0	0
23	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
24	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
25	3	No	3	1	0	0	3	1	0	0	3	1	0	0
26	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
27	3	Yes	2	1	1	1	2	1	1	1	2	1	1	1
28	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
29	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
30	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
31	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
32	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
33	3	No	3	1	0	0	3	1	0	0	3	1	0	0
34	3	Yes	2	1	1	1	3	1	0	0	3	1	0	0
35	3	Yes	2	1	1	1	3	1	0	0	3	1	0	0
36	3	No	3	1	0	0	3	1	0	0	3	1	0	0
37	3	Yes	2	1	1	1	2	1	1	1	2	1	1	1
38	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
39	3	No	3	1	0	0	3	1	0	0	3	1	0	0
40	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
41	3	No	3	1	0	0	3	1	0	0	3	1	0	0
42	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0
43	2	No	2	1	0	0	2	1	0	0	2	1	0	0
44	3	Yes	3	1	0	0	3	1	0	0	3	1	0	0

Figure 21: The number of TP, TN, FP and FN within 5 mm from reference coordinates per patient for each observer. Includes information about calcifications and number of GFMs in patient. Cells highlighted in blue represent a selection of a third marker for patients with 2 GFMs, orange— slice selection mistakes, yellow— mistakes due to calcifications and pink— other reasons.