# The challenges of using A-mode ultrasound for kinematic analysis

Radboudumc

Orthopaedic Research Laboratory

Nijmegen, June 2023

Student: ir. M. Devos (Utrecht University, Selective Utrecht Medical Master)

Supervisor: prof. dr. ir. N.J.J. Verdonschot (RadboudUMC, Orthopaedic Research Laboratory)

Daily supervisor: D.A. Christie MSc (University of Twente, Department of Biomedical Engineering)

Second reviewer: prof. dr. M.C. Kruijt (UMC Utrecht, Department of Orthopedics)

# Abstract

Amplitude-mode (A-mode) ultrasound (US) was proposed as non-radiant and non-invasive alternative technique for measuring bone kinematics and gait analysis. The system tracks the bone pose in 3D using optical markers and a motion tracking system, while A-mode US transducers measure the distance to the bony surface. However, detecting the actual bone depth using A-mode US is challenging, and the accuracy of the depth estimations remains unclear. Therefore, this study aimed to assess how accurate the A-mode US transducers can measure skin-to-bone distance. Multiple preliminary experiments were conducted and the outcomes were described in the appendices of this report. The main study included one right lower leg of a human cadaver, supplemented with three US holders and a total of nine US transducers. US estimated depths were calculated using the peak time in the A-mode signal and the speed of sound in soft tissues, and validated towards the ground truth depths from simultaneously conducted Computed Tomography (CT) scans at Ultra High Resolution (UHR). Two consecutive CT scans resulted in 18 CT ground truth depths, which were derived from the segmented 3D models of the tibia and the US transducers. Four measurements were excluded, and initially, for 3 of the 14 remaining outcomes (21%), the selected peak in the US signal was incorrect. After selecting the correct peaks, the US measurement overestimated the CT ground truth depth in 12/14 cases, and underestimated in the other two cases. Using the standard speed of sound in soft tissues (1540 m/s), the mean and standard deviation of the absolute difference between US estimated depth and CT ground truth depth was 0.9192 ± 0.4850 mm, whereas the best fitting speed of sound for the current setup (1424 m/s, according to one of the preliminary studies) decreased the mean absolute error to 0.6293 ± 0.6872 mm. In conclusion, if the correct amplitude peak in the US signal can be identified, the A-mode US transducers are capable of measuring skin-to-bone distance in a human cadaver leg with mean absolute errors below 1 mm.

# Contents

# 1. Introduction

Measuring in vivo bone pose and joint motion in three-dimensional (3D) space enhances making diagnoses of joint disorders, assists in the development of prosthetic implants, and enables quantitative evaluation of (surgical) treatment.[1] Contemporary methods for accurate quantification of complex joint kinematics, such as fixating intra-cortical bone pins to track the positions of bone segments with motion capture or surgical navigation systems,[2,3] and biplane X-ray fluoroscopy to capture dynamic motion of the knee,[4–6] are highly invasive or require harmful exposure to ionizing radiation, respectively.

The emerging four-dimensional (4D) dynamic imaging techniques, which include time domain data in order to track the dynamic 3D bone pose and joint kinematics inside MRI[7–9] or CT[10,11] scanners, also bear inherent limitations such as considerable workload, high cost, and unsuitability for gait analysis or evaluation of joint kinematics during activities of daily living. Furthermore, even the most commonly used non-invasive method for evaluating 3D joint kinematics in vivo, optical tracking of skin-mounted markers, poses significant challenges since non-rigid movement of the underlying soft tissues causes Soft Tissue Artifacts (STA) and impedes accurate representation of the skeletal kinematics.[12,13]

A promising non-invasive and non-radiant alternative technique involves multiple amplitude-mode (A-mode) ultrasound (US) transducers supplemented with optical markers for tracking 3D bone pose using a motion capture system. For each transducer, the distance between the skin and bone is estimated and, subsequently, the acquired points are registered to a known bone model (Figure 1). Previous studies measured tibiofemoral kinematics in a human cadaver with one-channel and multi-channel A-mode US systems, further verified its potential to overcome the issues with STA of skin-mounted markers with an in-situ comparison study on a cadaveric specimen, and demonstrated the technical and clinical feasibility of tracking joint motion during dynamic conditions in an in vivo experiment.[14–17]

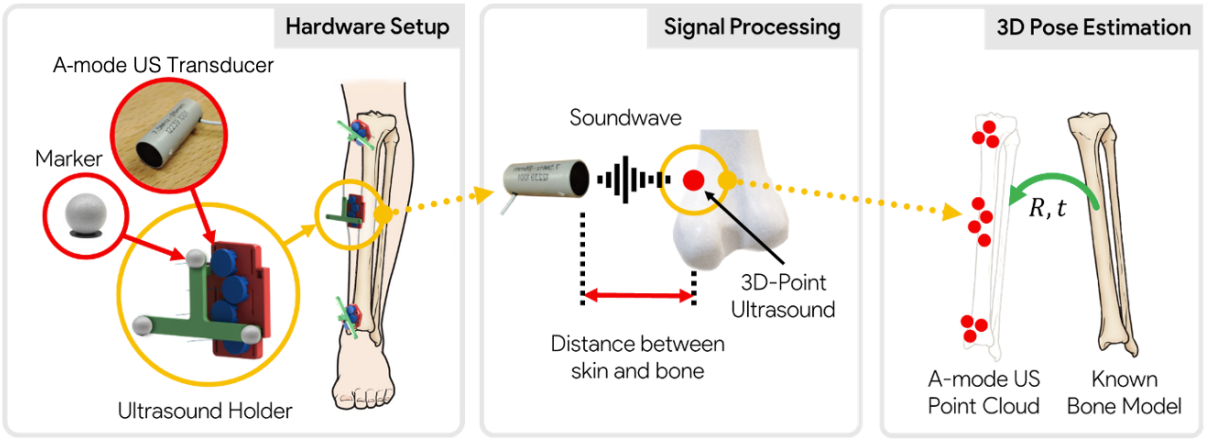



Figure 1 The first generation ultrasound (US) holders for A-mode ultrasound transducers are supplemented with optical markers (left) for motion tracking. The distance between skin and bone can be derived from the US signal (middle), and the corresponding 3D bony landmark points can be registered to a known bone model to estimate the 3D pose of the bone (right).

Despite the promising outcomes, further validation of the A-mode US system is required before clinical implementation is justified. Accurate bone depth estimation data to determine 3D

bone pose is essential for adequate kinematic analysis. However, detecting the actual skin to bone depth using the A-mode US system poses multiple challenges. Firstly, the emitted ultrasound wave passes through multiple layers of soft tissue before reaching the bone surface. The ultrasound signal is reflected and refracted at each transition, due to differences in acoustic impedances between the soft tissue layers. This can cause ambiguity in the signal received by the transducer, which potentially hides the true location of the underlying bone surface (Figure 2). Secondly, non-perpendicular orientation of the transducer with respect to the bone surface results in less prominent amplitude peaks in the US signal. Previously, the current A-mode US system's capability to accurately detect depths was only demonstrated in a phantom study involving a staircase with 24 stair steps of 2 mm height and further validation of bone depth estimation is lacking.

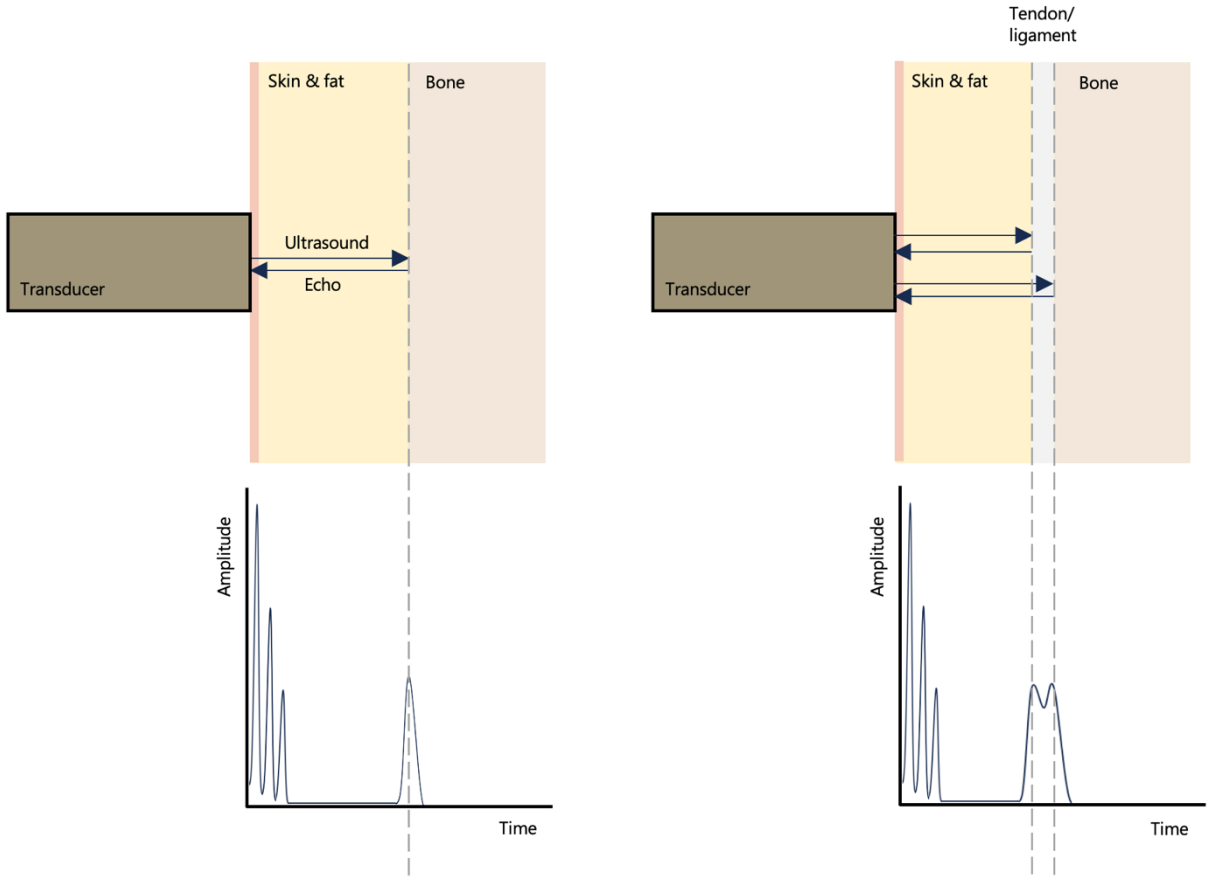



Figure 2 Illustration of an A-mode ultrasound transducer emitting ultrasound waves trough the soft tissues before reaching the bone surface (top) and the corresponding (simplified) ultrasound amplitude versus time signal (bottom). In ideal circumstances, the amplitude peak in the ultrasound signal corresponds to the bone surface (left). In case of additional layers of soft tissue, the signal received by the transducer might cause multiple amplitude peaks that complicate selecting the true bone peak (right).

Hence, the aim of this study was to assess how accurate the A-mode US transducers can measure skin-to-bone distance in a human cadaver leg. This can be done by processing the received US signal into depth estimation data (i.e., the distance between the transducer's tip and the respecting bone surface), which is validated towards simultaneously conducted CT scan images (the ground truth). CT scan data includes 3D poses of all US transducers as well as the 3D bone model of the cadaver leg, which allows for deriving the true distance between the US transducer's tip and bone surface point.

To overcome limitations of the previous US holder design, including the lack of features for adjusting individual transducers and for tracking individual transducers orientations, a new ultrasound holder prototype (gen-1) has been developed and introduced in this study. It allows for free orientation adjustment of the transducers, which means each individual transducer's orientation can now be freely adjusted by the use in the full three degrees of freedom.

# 2. Materials & Methods

## 2.1 Experimental setup

The experimental setup (Figure 3) includes one frozen right lower leg (from foot to proximal tibia) of a human cadaver, obtained from and approved by the Department of Anatomy of the Radboud University Medical Center (RUMC). An ultrasound (US) machine (Diagnostic Sonar Ltd., Livingston, Scotland) with nine A-mode US transducers (Imasonic SAS, Voray-sur-l'Ognon, France; operating frequency 7.5 MHz, diameter 6 mm, casing dimensions 20 mm by 8 mm diameter) attached (Figure 4), is connected to a PC for streaming and storing the receiving ultrasound signal. Three custom-made, 3D-printed plastic holders were developed to secure the A-mode US transducers to the cadaveric leg.

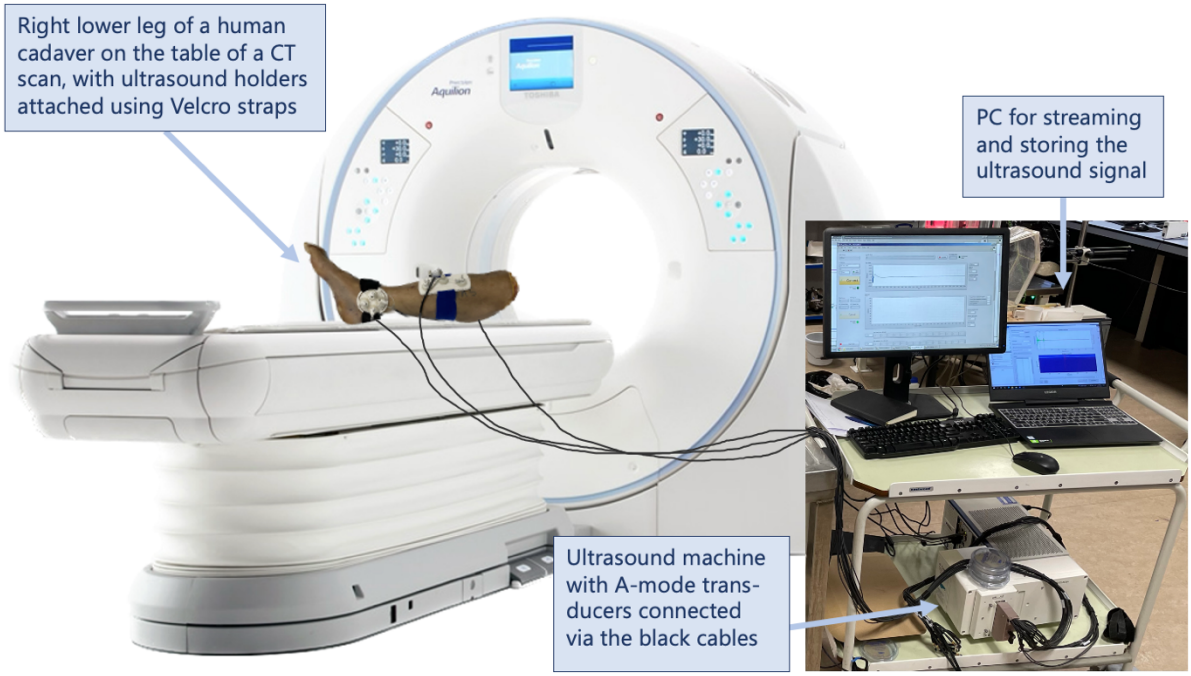



Figure 3 Schematic of the experimental setup including a human cadaveric leg on the table of a CT scan, A-mode transducers within US holders attached to the leg and connected to the ultrasound machine and external PC.

The design of the holders includes nine ball-joint casings that exactly fit the A-mode US transducers, which enable pivoting and altering the orientation of the transducers after the ball-joints are inserted into the sockets of the holders. Additionally, square polyamide platforms (15*15 mm) were 3D-printed and each platform was supplemented with four fiducial markers (tantalum spheres with diameter  1 mm). Polyamide screws connect the platforms to the ball-joint casing, and tightening of the screws further presses the transducers against the skin (Figure 5). The US setup and holders were developed at the Department of Biomechanical Engineering of the University of Twente.

## 2.2 Measurement protocol

The cadaver leg was thawed one day prior to the experiments, and the experiments were performed at room temperature. After attaching the ultrasound holders to the leg using Velcro straps, US gel was applied in the sockets of the holders. From proximal to distal, the holders are positioned (1) lateral to the tibial shaft, (2) medial to the tibial shaft, and (3) at the medial malleolus with two, three and four A-mode US transducers in each holder, respectively (Figure

6& Figure 7). The cadaveric leg was statically positioned on the table of an Ultra-High Resolution Computed Tomography (UHR CT) scanner (Aquilion Precision, Canon Medical Systems Europe BV, Zoetermeer, the Netherlands) at the radiology department of the RUMC.

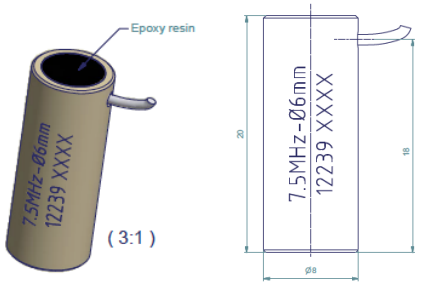



Figure 4 Schematic of one A-mode ultrasound transducer (Imasonic SAS, Voray-sur-l'Ognon, France). Dimensions are shown in millimeters.

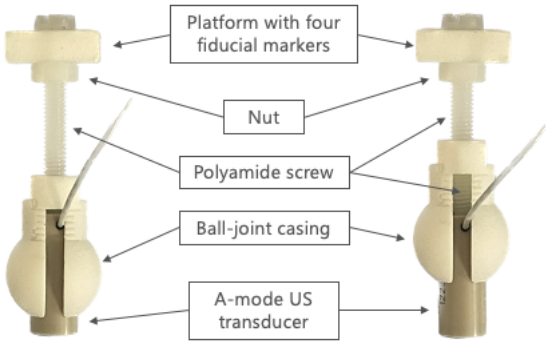



Figure 5 The A-mode ultrasound transducer is positioned within the ball-joint casing. The polyamide screw connects the platform with fiducial markers to the ball-joint casing, and further presses the transducer against the skin when tightened (right figure).

After positioning the leg on the table, the US transducers were manually pivoted, while the amplitude versus time signal of the corresponding transducer was visually checked by two observers (DC and MD) to find the optimal orientation of each transducer (i.e. the orientation that results in the clearest amplitude peak in the ultrasound signal that is presumed to reflect the bone surface). In total, the whole setup of cadaver leg with the US holders and nine US transducers was scanned in the UHR CT twice. Between the two scans, the position of the leg was slightly altered, and the US transducers were reoriented as described above. During both CT scans, the US measurements data was obtained simultaneously with CT scan imaging, yielding two data outputs; ultrasound raw signal, stored as a collection of TIFF files, and volume images, stored as a collection of DICOM files.

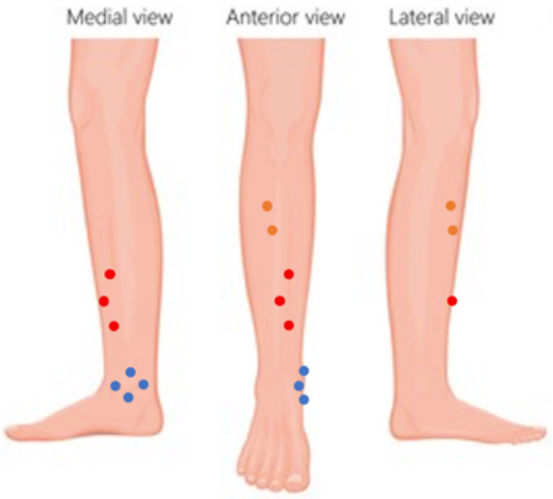



Figure 6 Ultrasound transducer positions on the lower leg. Positioning (from proximal to distal): two transducers lateral to the tibial shaft (orange), three transducers medial to the tibial shaft (red), and four transducers on the medial malleolus (blue).

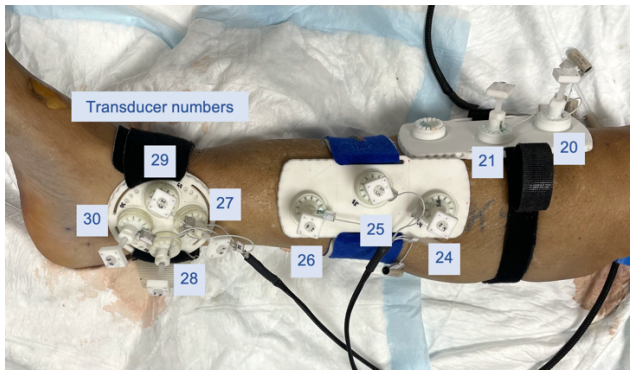



Figure 7 The nine transducers within three holders (from left to right): medial malleolus, tibial shaft distal/medial, tibial shaft proximal/lateral.

## 2.3 Data processing

### 2.3.1 Depth estimation using ultrasound

The US measurement was started before and ended after the CT scan and, as the raw ultrasound signal is encoded in TIFF files, a total of approximately 12000 TIFF files were generated and subsequently stored. A subset of all the TIFF files were selected from the middle of the data set, assuming that these were representative for the whole data set. The TIFF files are read and decoded into raw amplitude versus time data for each transducer. Raw data is then processed using Time Gain Compensation (TGC), frequency filtering, enveloping, and peak detection using an algorithm written by Dennis Christie in Matlab R2022b (Figure 8). As the algorithm requires lower and upper bounds as an input for detecting the peak, these were manually selected and agreed on by two observers (DC and MD) per transducer for each measurement.

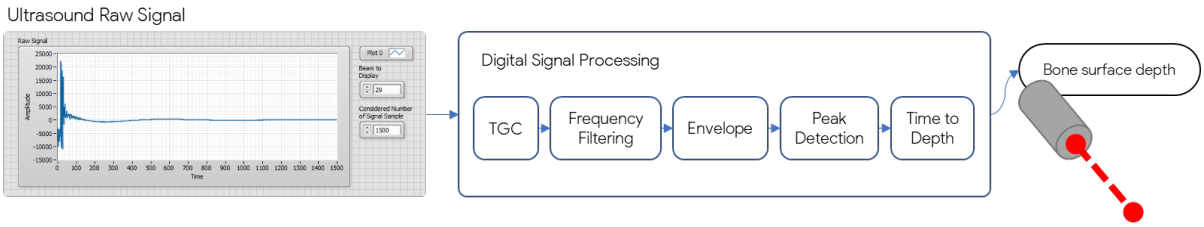



Figure 8 Workflow for computing the bone surface depth estimation from the A-mode ultrasound raw signal. TGC = Time Gain Compensation.

Figure 9 shows illustrative figures of the A-mode US results including the lower and upper bounds for four transducers, collected during preliminary experiments with the A-mode US setup and the cadaver leg. Ideally, the raw signal shows one clear, high peak which can be presumed to reflect the bone surface (Figure 9A and B). However, frequently the peak is less prominent (Figure 9C), or the signal shows multiple peaks (Figure 9D) and selecting the lower and upper bounds is not as evident. In that case, the bounds are determined such that the highest peak or, if there is not one highest peak, the farthermost peak is selected.

Finally, the time of the peak is converted into depth estimation data. As the sound waves are both emitted and received by the same US transducer, the wave travels twice the distance in the peak time. Hence:

$$s_{est} = \frac{t * v}{2}$$

where $s_{est}$ is the estimated distance between the transducer tip and the bone, $t$ is the time of the amplitude peak in the US signal, and $v$ is the speed of sound through the intermediate soft tissues. The value for speed of sound was varied during data analysis, from 1424 to 1540 m/s, based on the minimum speed of sound measured at room temperature in different pieces of cadaver soft tissues in a preliminary experiment using the current A-mode US setup (Appendix A.2), and the established value for speed of sound in soft tissues according to literature,[18] respectively. For each transducer, the subset of TIFF files yield a collection of depth estimation values, of which the median value is reported as US depth estimation.

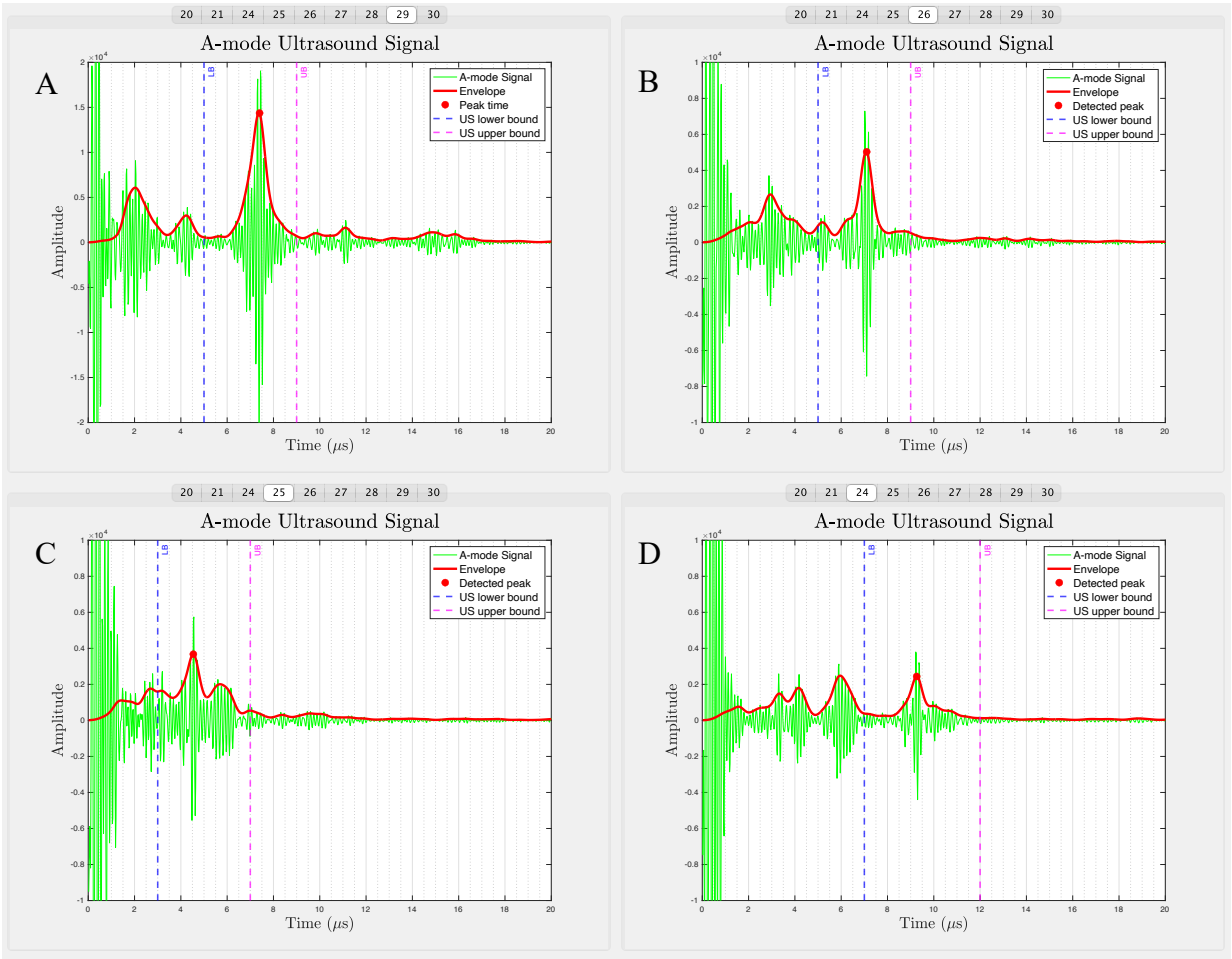



Figure 9 Four illustrative figures of A-mode ultrasound amplitude versus time results for four transducers, showing the raw signals (green), the envelope of the signal (red line), the peak detected by the algorithm (red dot), and the manually selected lower (blue dashed line) and upper (magenta dashed line) bounds for the peak detection algorithm.

### 2.3.2 Depth ground truth using CT

The CT scan Field Of View (FOV) was minimized, such that the resolution of the two scans was optimized to 0.20 * 0.20 * 0.25 mm (voxel size with 0.25 mm slice thickness). Furthermore, CT scan images were reconstructed using the Single Energy Metal Artifact Reduction (SEMAR) algorithm tool, to obtain optimal images despite the multiple metallic US transducers (Figure 10). After collecting the data, the two sets of CT images were processed to obtain two data outputs: the 3D bone model of the bone and the US transducers' 3D poses, which were used to derive the bone surface depth ground truth (Figure 11). Therefore, the tibia, nine transducers and 36 fiducial markers were segmented in both CT scans with 3D slicer[19] by one observer (MD), using thresholding and manual adjustments, with supplementary smoothing (Gaussian 0.5 mm) of the transducers. Additionally, the other observer (DC) independently segmented all elements in the first CT scan with Materialise Mimics. The resulting tibia, transducers and fiducials 3D meshes were transformed into the global coordinate system using Matlab R2022b.

The bone surface depth ground truth is assumed to be the distance between the tip of the transducer that touches the skin, and the underlying bone surface. Three different methods were applied to find this distance for each transducer, of which the 'tube method' is assumed to best represent the ground truth value of the US depth estimation (Appendix B). The method involves hypothetical US beam lines, parallel to the z-axis (long axis) of the US transducer, from

the tip of the transducer to the bone surface. The lines form a hypothetical tube with radius 3 mm (corresponding to the US transducer's tip diameter excluding the casing), around the central z-axis of the transducer (Figure 12).

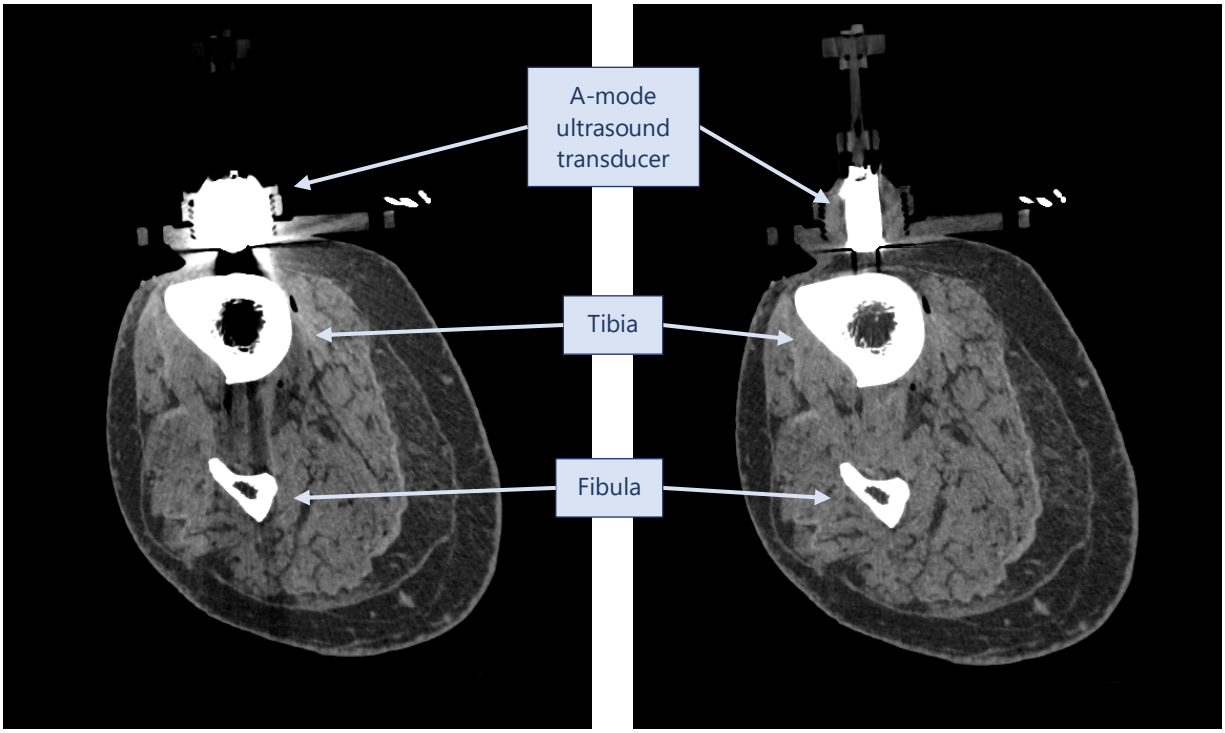



Figure 10 CT scan slice (axial) of the lower leg with an A-mode ultrasound transducer facing the tibia before (left) and after (right) Single Energy Metal Artifact Reduction (SEMAR).

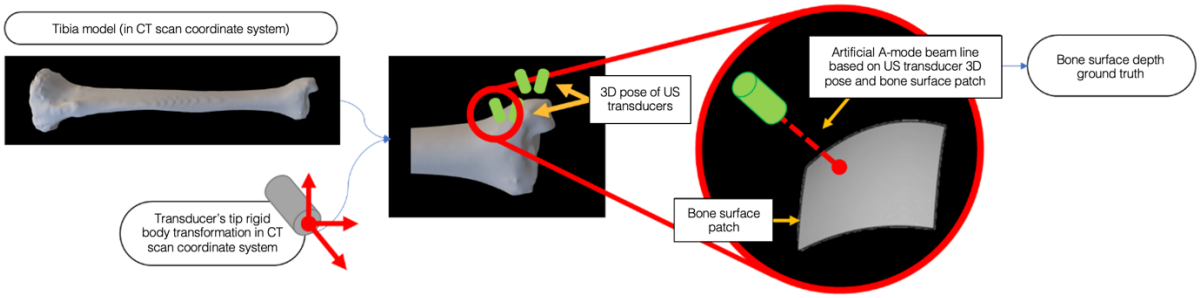



Figure 11 Workflow for computing the bone surface depth ground truth, which is defined as the distance between the tip of the transducer and the underlying tibia surface. The tibia model and US transducers are segmented from the CT scan data, and their 3D poses are used to compute the bone surface depth ground truth.

### 2.3.3 Depth validation

Using the hypothetical US beam tube method for CT scan data set 1, the ground truth was calculated using the segmentations from both independent observers (DC and MD). The ground truth distances from each observer are then compared, and the mean absolute difference is defined as the interobserver variability of the ground truth. Subsequently, the segmentations by one observer (MD) for both CT scan data sets were used to determine a total of 18 ground truth depth values. For the main comparison of the US depth estimation data and the depth ground truth data, 1540 m/s was used as speed of sound for the US depth calculations. From this comparison, the observer(s) could verify whether the manually selected lower and upper bounds in the US signal (Figure 9) correspond to the ground truth. If this was not the

case, the lower and upper bounds were updated such that the ground truth is within the range, and the US depth estimation calculations were repeated. Next, the updated estimated depths for each US transducer were compared to the corresponding ground truth depths, and the mean ± standard deviation (SD)absolute difference between the values was determined to define the overall measurement error. Finally, the speed of sound input for the US depth calculations was varied from 1424 to 1540 m/s to find the value that results in the lowest overall error between estimated and ground truth bone surface depth.

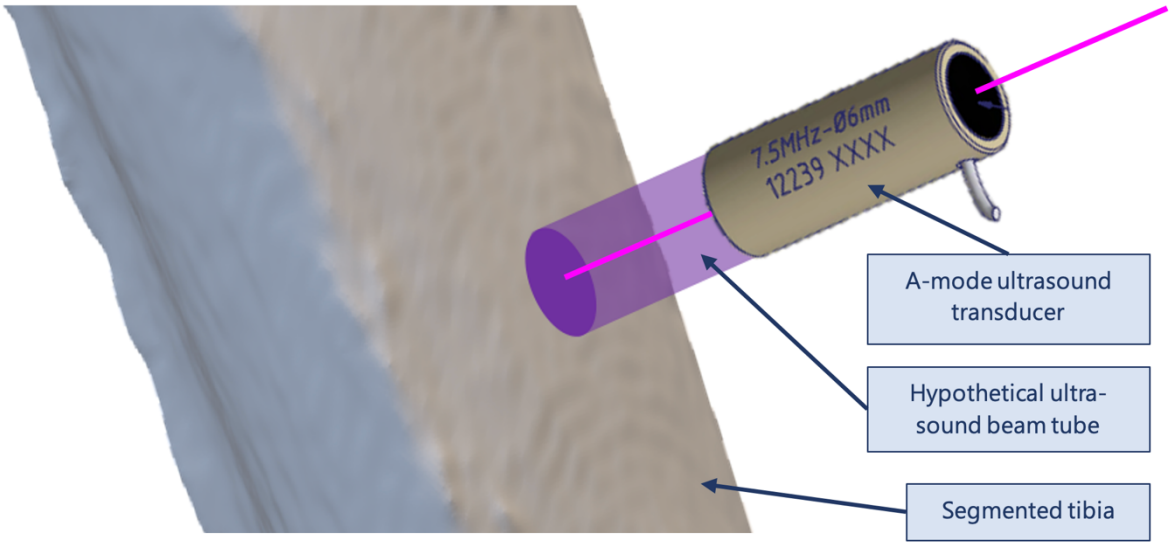



Figure 12 A hypothetical ultrasound (US) beam tube (purple) is defined as a set of hypothetical US beam lines from the tip of the US transducer to the underlying tibia surface, with a radius of 3 mm around and parallel to the central z-axis of the transducer (magenta line).

# 3. Results

## 3.1 Data collection

Data was collected for nine US transducers during the two subsequent CT scans. In the end, the results of two US transducers had to be excluded from the analysis. Firstly, the US measurements from transducer number 30 (see Figure 7) were excluded because during the experiment, a clear (bone) peak could not be found in the signal despite altering the orientation of the transducer. Secondly, the ground truth depth from the CT scan for transducer 20 (see Figure 7) could not be determined, since analysis of the CT data showed that the transducer was not facing the bone during both CT scans (Figure 13). Hence, only seven transducers remained for inclusion in the analysis for two CT scans, yielding a total of 14 values for the comparison of US depth estimation and CT depth ground truth data.

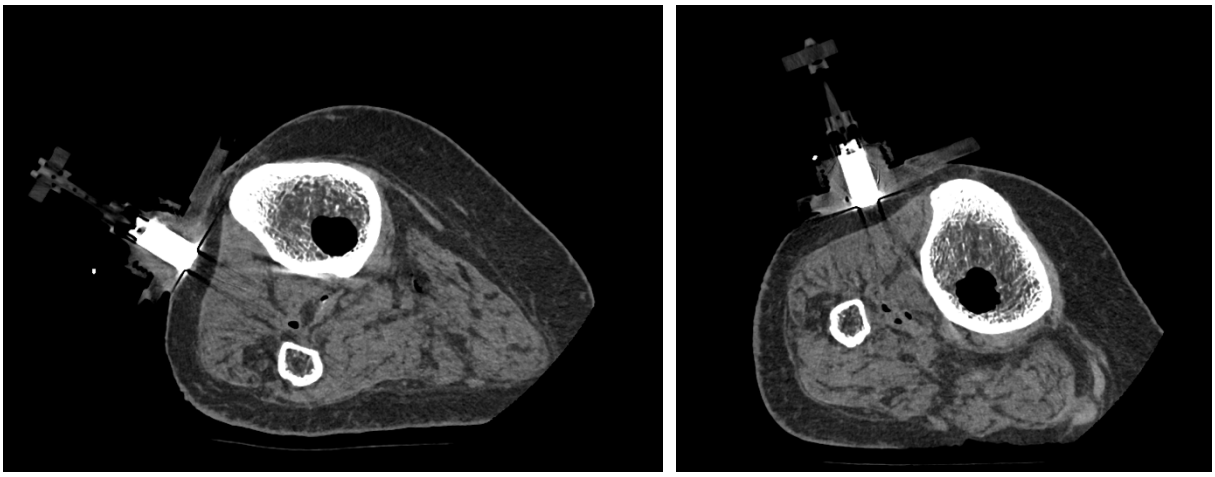


Figure 13 Axial slices of CT scan 1 (left) and CT scan 2 (right) showing the orientation of transducer number 20 with respect to the tibia and fibula. A ground truth depth could not be obtained for this transducer, since the transducer was not facing the bone (tibia).

## 3.2 Depth ground truth using CT

Depth ground truth results were obtained using the 3D models of the tibia and US transducers (Figure 14), segmented by two independent observers for CT scan 1, and one observer for CT scan 2. The interobserver variability of the ground truth depth was obtained by comparing the results of the two observers for CT scan 1 (Figure 15). The lowest difference (0.0246 mm) between the measurements from the two observers was found for transducer number 25 (4.8718 mm versus 4.8964 mm), and the largest difference (0.2198 mm) was observed for transducer 28 (14.8368 mm versus 15.0566 mm). Overall mean and standard deviation (SD) of the difference in ground truth depth was 0.1088 ± 0.0719 mm (Table C.1).

Figure 14 Screen capture of the 3D models of the tibia and the nine ultrasound transducers (segmentations from CT scan 1).

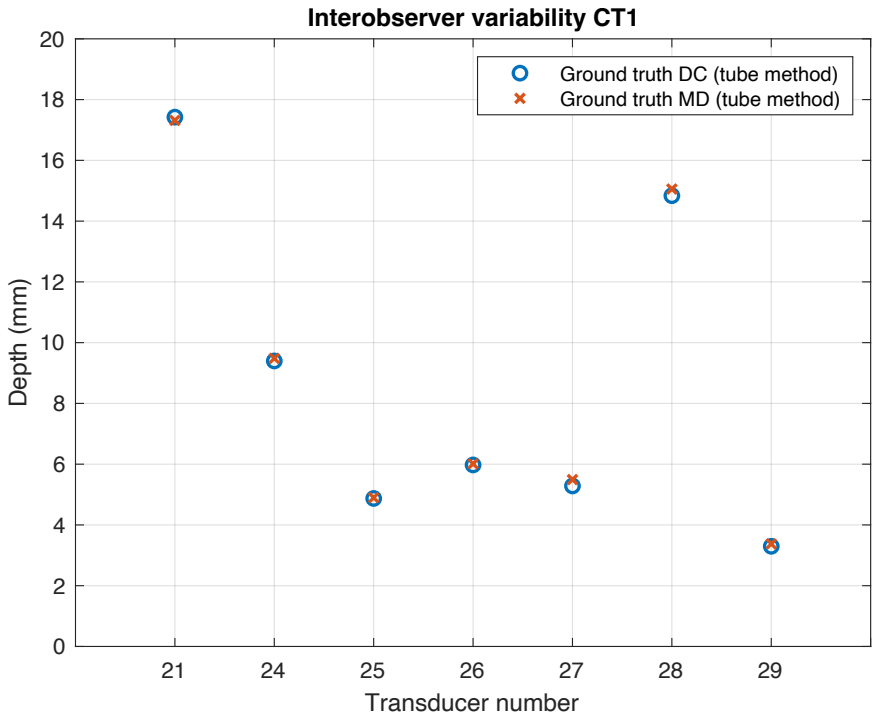



Figure 15 Ground truth distances from transducer's tip to tibia surface, using the segmentations by two independent observers (DC: Dennis Christie; MD: Maxime Devos) for CT scan 1.

## 3.3 Ultrasound peak selection

The manually selected lower and upper bounds were used as input for the US peak detection algorithm. First, the US peak times were compared to the CT ground truth depths, by converting the CT depths into time domain data using speed of sound 1540 m/s. For three out of 14 cases (21%), the CT ground truth was outside the chosen bounds for the US peak detection (Figure 16). Therefore, the US bounds for these three measurements were redefined in order to include the CT ground truth depth in the range.

## 3.4 Depth validation

The US graphs, including raw and enveloped US signal, US lower and upper bounds, and estimated peak time, were visually analyzed and compared to (1) the ground truth time value (calculated using speed of sound 1540 m/s and obtained from the data of one observer (MD)), and (2) the corresponding CT scan images (Figure C.2). Additionally, the numerical outcomes of US estimated peak times were converted into US estimated depths using speed of sound 1540 m/s, and these were compared to CT ground truth depths to find the numerical differences. The results (Table C.2) show that in 12 out of 14 cases, the US estimated depth is larger than the CT ground truth depth. Only for CT scan 2, transducer 24 and 28, the US estimated depth was an underestimation with respect to the CT ground truth depth. Overall, the absolute differences range from 0.0343 mm (CT scan 2, transducer 29) to 1.8144 mm (CT scan 2, transducer 28), and mean and SD of the absolute difference were 0.9192 mm and 0.4850 mm, respectively.

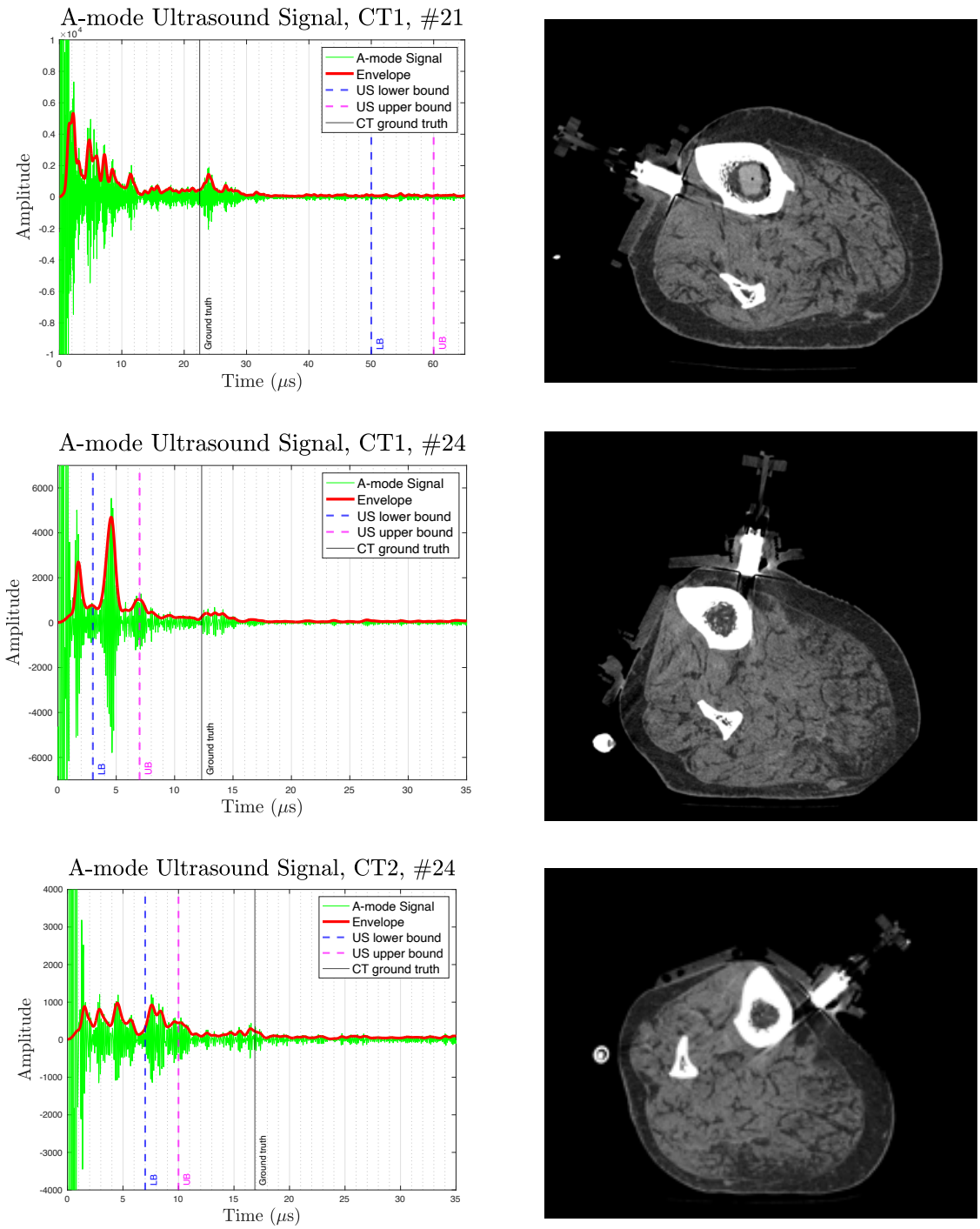



Figure 16 The ultrasound raw signal, signal envelope, US lower and upper bounds, and CT ground truths for three transducers (left) and the corresponding axial CT slice (right). For these three cases, the CT ground truth was outside the range of the manually selected US bounds.

Subsequently, the speed of sound was varied from 1424 m/s to 1540 m/s to find the speed of sound value that results in the lowest overall mean absolute difference between US estimated depth and CT ground truth depth. Using the best fitting speed of sound for the current study, 1424 m/s, the US estimated depths were recalculated (Table C.2). The resulting US estimated depths were an overestimation of the CT ground truth in 8/14 cases, and an underestimation in the remaining 6/14 cases. The absolute differences range from 0.1000 mm (CT scan 1, transducer 24) to 2.8990 mm (CT scan 2, transducer 28), and mean of the absolute differences

between US estimated depth and CT ground truth depth decreased to 0.6293 ± 0.6872 mm, as compared to the calculations with speed of sound 1540 m/s (Figure 17).

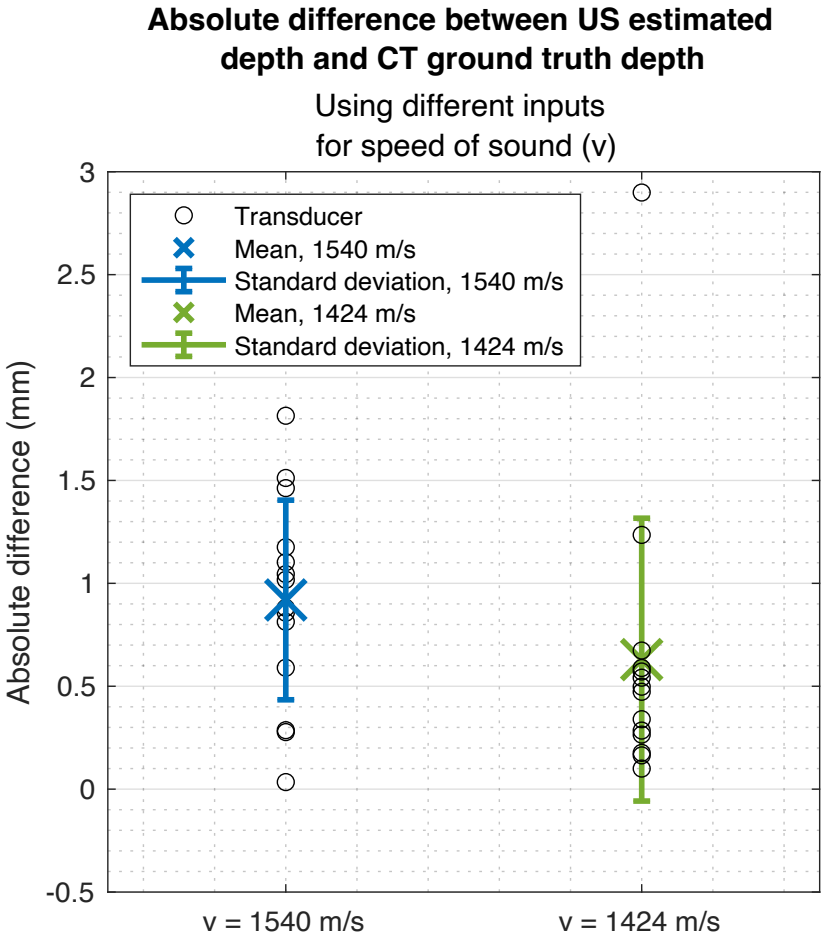



Figure 17 Absolute differences between US estimated depth and CT ground truth depth for each transducer (black circles), with two different speed of sound inputs for the US estimated depth calculations. (v: speed of sound.)

# 4. Discussion

## 4.1 Depth validation

This study aimed to asses how accurate the A-mode US system can measure skin-to-bone distance in a human cadaver leg. The depth validation results show that the mean absolute difference between the A-mode US measurement and the CT ground truth is just below 1 mm (0.9192 mm), on condition that the correct peak in the A-mode US signal is selected. For comparison, the current standard for non-invasive and non-radiant kinematic measurements with skin-mounted markers and optical tracking measures 3D bone pose with root mean square errors up to 29.3 mm.[20] The underlying issue, soft tissue artifacts, remains unsolved to date, despite extensive research and various attempts to compensate for it.[21,22] The ultimate goal is to achieve accuracies similar or superior to biplane X-ray fluoroscopy, which can capture dynamic motion of the knee with accuracies in the order of 1 mm for translations and 1 degree for rotations.[4–6] However, future steps towards 3D bone kinematic measurements, such as registration and 3D tracking of the A-mode US transducers, will further increase the A-mode US system's errors, and final accuracies in the order of 1 mm seem out of reach unless the skin-to-bone depth measurement accuracy is improved.

## 4.2 Challenges related to ultrasound

In order to improve the accuracy, the inherent challenges of measuring skin-to-bone distance using A-mode US should be addressed. The first major challenge was selecting the right bone peak in the US signal. Initially, blind selection of the peaks appeared to be incorrect for three out of 14 transducers (21%). For two out of three incorrectly selected peaks, (CT scan 1 and 2, transducer 24) the ground truth is close to multiple small peaks in the US signal rather than the highest peak, which is potentially due to multiple soft tissue transitions and therefore fewer US signal that is transmitted to the bone. For the third case, a smaller, arguable peak further away in the signal was selected. Since another observer might have selected the correct bone peak right away, this illustrates that peak selection is dependent on the observer's experience. Furthermore, one transducer (number 20) did not face the bone surface despite peaks in the US signal, which supposedly reflect transitions of soft tissues. For the transducer that lacked a clear (bone) peak in the US signal (number 30), only a small layer of soft tissue is present between the US transducer's tip and bone surface, and near field distortion of the US signal is presumed to play a role here. Overall, these findings also contribute to a better understanding of the A-mode signal.

The second major challenge of using A-mode US for estimating skin-to-bone distance is associated with the speed of sound in soft tissues. Even though 1540 m/s is known as golden standard for average speed of sound in soft tissues and even used in B-mode US reconstructions, this value was acquired in 1950 through an in vivo measurement of human biceps,[18] and speed of sound in cadaver tissue differs from human in vivo tissue (Appendix A.2). Therefore, the speed of sound was altered to find the optimal speed of sound within the range of the preliminary experiment. Apparently, the mean absolute difference between US estimated depth and CT ground truth depth is strongly dependent on the selected speed of sound, considering that using the best fitting speed of sound (1424 m/s instead of 1540 m/s) for the calculations reduced the mean absolute difference with almost 32% (from 0.9192 mm to 0.6293

mm). Thus, selecting the proper speed of sound for each US transducer location could further enhance the accuracy of depth estimations by the A-mode US system.

## 4.3 Challenges related to CT

In order to validate the US estimated depth, the tibia and US transducers were segmented from CT scan images to define the ground truth depths. Although it was initially planned to use an automatic bone segmentation algorithm[23] for tibia segmentation in CT,[24] using this neural network was not feasible due to a missing femur and patella in the CT scan. Besides, the algorithm is trained for the proximal part of the tibia whereas in the current study the distal tibia was of greater interest. Therefore, the tibia and transducers were manually segmented, and the variability of the ground truth depth was analyzed by comparing the results from two independent observers for one of the two CT scans. Due to metal artifacts caused by the US transducers, some parts of the manual segmentation were complicated, especially at locations where the transducers were close to the tibia. Nevertheless, the resulting interobserver variability of the ground truth depths was small, 0.1088 mm, which equals approximately half the size of a voxel (0.20 * 0.20 * 0.25 mm). The bone segmentation on CT scan images will remain a challenge, unless the 3D bone model is obtained from an additional CT or MRI scan without US transducers attached.

## 4.4 Strengths and limitations

In spite of the challenges, the current study had multiple strengths. Firstly, this was the first study that aimed to validate the depth estimation by A-mode US, by comparing it to the ground truth skin-to-bone depth on simultaneous CT scans. Secondly, the results from this study contribute to an improved understanding of A-mode US signals. For example, up until now, it was assumed that the orientation of the transducer should be perpendicular or at least almost perpendicular to the bone surface. However, the results from this study show that even with a non-perpendicular orientation, the A-mode US signal can show a clear peak that closely corresponds to the bone surface. Thirdly, preliminary experiments were conducted in order to define and account for potential additional inaccuracies.

One of the limitations of this study was the assumed value for speed of sound in soft tissues. Even though the realistic range for speed of sound was defined in a preliminary experiment, the accuracy of the depth validation experiment would improve if speed of sound was measured through soft tissues at the location of each transducer. The best available method for this would be to cut out the soft tissue pieces and measure speed of sound in a setup similar to the preliminary experiment. However, this procedure was not possible as the cadaveric leg would then be unsuitable for future experiments. Another limitation of this study was the method for determining the ground truth depth, as it omits the divergent character of the US beams and might therefore not exactly represent the US estimated depth. This could be corrected for in the future, but requires more detailed insight into the specific patterns of these A-mode US waves.

# 5. Conclusions and future work

## 5.1 Conclusions

The mean absolute difference between the A-mode US estimated depth and the CT ground truth depth is just below 1 mm (0.9192 mm), supposing that the correct peak in the A-mode US signal is selected. Still, selecting the right peak remains a major challenge, since it is subject to interpretation by the observer, clear (bone) peaks are sometimes lacking and in some cases, the clear high peak does not represent the bone surface. In addition, the mean absolute difference reduces considerably, from 0.9192 mm to 0.6293 mm, when the optimal speed of sound is used. However, defining the speed of sound for each transducer is challenging, and will remain a challenge in *in vivo* measurements as well. In order to qualify as accurate and reliable alternative method for 3D bone kinematic measurements, the A-mode US system's skin-to-bone depth measurement inaccuracy should be further reduced.

## 5.2 Future work

Future studies should focus on expanding the available data for validating US estimated depths towards ground truth depths. Thereby, the correlation between amplitude peaks in de A-mode US signal and the bone surface could be further investigated. For example, with additional data, a potential positive relationship between perpendicularity of the US transducer and US amplitude height or clear peaks in the US signal could be further confirmed or denied (Appendix E). To address the problem of peak selection, additional CT scan data of human subjects could eventually provide a standard for upper and lower bounds for the US signal, depending on the location of the transducer.

With regard to the speed of sound in soft tissues, future studies on human cadaver tissue could consider conducting speed of sound experiments right after the US/CT scan experiment and correlate these results to the different layers of soft tissue. Future studies on human subjects could consider measuring the different layers of soft tissue on the CT scan images. Besides, the challenge with speed of sound in vivo will be less significant than for human cadaver tissue, considering the available literature on speed of sound through soft tissues in vivo. Segmentation of the tibia and US transducers from the CT scan data could be improved by scanning the lower leg and transducers separately, which reduces the effect of metal artifacts on the segmentations and enables fitting of the segmented tibia and transducers to the other CT scan. Supposing that the accuracy of skin-to-bone depth estimation by A-mode US can be enhanced, future studies should start focusing on practicality of the system. To enable gait analyses, it should be considered that that the US transducers eventually need to be supplemented with optical trackers, which potentially requires different positioning of the US holders.

This study has shown that A-mode US is a promising tool for non-invasive and non-radiant kinematic analyses. However, further steps in the development and implementation of the technique are necessary. Although the current findings indicate that accurate skin-to-bone depth measurements by each A-mode US transducer requires correct peak selection in the US signal, the ultimate 3D bone pose reconstruction is plausibly less susceptible to wrongly selected peaks, considering that the great majority (approximately 80%) of the peaks was selected correctly. Assuming that this finding can be extrapolated for an increased amount of US transducers, future research could enhance the registration algorithm such that it finds the best

fit to the 3D shape of the bone model. Thereby, the remaining points (approximately 20%) that correlate to the wrongly selected peaks in the US signal can be identified and excluded, and the impact of selecting the (in)correct peak will be less significant. Notably, these results are based on static (cadaver) measurements. Since the dynamic aspect of kinematic analyses will increase the challenge of bone depth detection and potentially peak selection in the US signal, the current results should not be extrapolated for future studies that incorporate dynamic measurements.

## 6. Acknowledgment

The author(s) express their gratitude to Richard van Swam from the Orthopaedic Research Laboratory at Radboud University Medical Center, the Netherlands for the generous assistance during preparation and execution of the cadaveric experiments.

# References


1. Affatato S. Biomechanics of the knee. In: *Surgical Techniques in Total Knee Arthroplasty and Alternative Procedures*. Elsevier; 2015:17-35. doi:10.1533/9781782420385.1.17
2. Lafortune MA, Cavanagh PR, Sommer HJ, Kalenak A. Three-dimensional kinematics of the human knee during walking. *J Biomech*. 1992;25(4):347-357. doi:10.1016/0021-9290(92)90254-x
3. Reinschmidt C, van den Bogert AJ, Lundberg A, et al. Tibiofemoral and tibiocalcaneal motion during walking: external vs. skeletal markers. *Gait & Posture*. 1997;6(2):98-109. doi:10.1016/S0966-6362(97)01110-7
4. Li G, Van de Velde SK, Bingham JT. Validation of a non-invasive fluoroscopic imaging technique for the measurement of dynamic knee joint motion. *Journal of Biomechanics*. 2008;41(7):1616-1622. doi:10.1016/j.jbiomech.2008.01.034
5. Kozanek M, Hosseini A, Liu F, et al. Tibiofemoral kinematics and condylar motion during the stance phase of gait. *Journal of Biomechanics*. 2009;42(12):1877-1884. doi:10.1016/j.jbiomech.2009.05.003
6. Guan S, Gray HA, Schache AG, Feller J, de Steiger R, Pandy MG. In vivo six-degree-of-freedom knee-joint kinematics in overground and treadmill walking following total knee arthroplasty. *J Orthop Res*. 2017;35(8):1634-1643. doi:10.1002/jor.23466
7. Kaiser J, Bradford R, Johnson K, Wieben O, Thelen DG. Measurement of tibiofemoral kinematics using highly accelerated 3D radial sampling. *Magn Reson Med*. 2013;69(5):1310-1316. doi:10.1002/mrm.24362
8. Clarke EC, Martin JH, d'Entremont AG, Pandy MG, Wilson DR, Herbert RD. A non-invasive, 3D, dynamic MRI method for measuring muscle moment arms in vivo: demonstration in the human ankle joint and Achilles tendon. *Med Eng Phys*. 2015;37(1):93-99. doi:10.1016/j.medengphy.2014.11.003
9. Mazzoli V, Schoormans J, Froeling M, et al. Accelerated 4D self-gated MRI of tibiofemoral kinematics. *NMR in Biomedicine*. 2017;30(11):e3791. doi:10.1002/nbm.3791
10. Forsberg D, Lindblom M, Quick P, Gauffin H. Quantitative analysis of the patellofemoral motion pattern using semi-automatic processing of 4D CT data. *Int J Comput Assist Radiol Surg*. 2016;11(9):1731-1741. doi:10.1007/s11548-016-1357-8
11. Zhao K, Breighner R, Holmes D, Leng S, McCollough C, An KN. A Technique for Quantifying Wrist Motion Using Four-Dimensional Computed Tomography: Approach and Validation. *Journal of Biomechanical Engineering*. 2015;137(7):074501. doi:10.1115/1.4030405
12. Leardini A, Chiari L, Croce UD, Cappozzo A. Human movement analysis using stereophotogrammetry: Part 3. Soft tissue artifact assessment and compensation. *Gait & Posture*. 2005;21(2):212-225. doi:10.1016/j.gaitpost.2004.05.002
13. Peters A, Galna B, Sangeux M, Morris M, Baker R. Quantification of soft tissue artifact in lower limb human motion analysis: A systematic review. *Gait & Posture*. 2010;31(1):1-8. doi:10.1016/j.gaitpost.2009.09.004
14. Niu K, Homminga J, Sluiter V, Sprengers A, Verdonschot N. Measuring relative positions and orientations of the tibia with respect to the femur using one-channel 3D-tracked A-mode ultrasound tracking system: A cadaveric study. *Medical Engineering & Physics*. 2018;57:61-68. doi:10.1016/j.medengphy.2018.04.015
15. Niu K, Sluiter V, Sprengers AMJ, Homminga J, Verdonschot N. A Novel Tibiofemoral Kinematics Measurement System Based On Multi-Channel A-Mode Ultrasound System. *EPiC Series in Health Sciences*. 2017;1:166-170.

16. Niu K, Anijs T, Sluiter V, et al. In situ comparison of A-mode ultrasound tracking system and skin-mounted markers for measuring kinematics of the lower extremity. *Journal of Biomechanics*. 2018;72:134-143. doi:10.1016/j.jbiomech.2018.03.007
17. Niu K, Sluiter V, Homminga J, Sprengers A, Verdonschot N. A Novel Ultrasound-Based Lower Extremity Motion Tracking System. In: Zheng G, Tian W, Zhuang X, eds. *Intelligent Orthopaedics*. Vol 1093. Advances in Experimental Medicine and Biology. Springer Singapore; 2018:131-142. doi:10.1007/978-981-13-1396-7_11
18. Ludwig GD. The Velocity of Sound through Tissues and the Acoustic Impedance of Tissues. *The Journal of the Acoustical Society of America*. 1950;22(6):862-866. doi:10.1121/1.1906706
19. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magnetic Resonance Imaging*. 2012;30(9):1323-1341. doi:10.1016/j.mri.2012.05.001
20. Akbarshahi M, Schache AG, Fernandez JW, Baker R, Banks S, Pandy MG. Non-invasive assessment of soft-tissue artifact and its effect on knee joint kinematics during functional activity. *J Biomech*. 2010;43(7):1292-1301. doi:10.1016/j.jbiomech.2010.01.002
21. Andersen MS, Benoit DL, Damsgaard M, Ramsey DK, Rasmussen J. Do kinematic models reduce the effects of soft tissue artefacts in skin marker-based motion analysis? An in vivo study of knee kinematics. *Journal of Biomechanics*. 2010;43(2):268-273. doi:10.1016/j.jbiomech.2009.08.034
22. Richard V, Cappozzo A, Dumas R. Comparative assessment of knee joint models used in multi-body kinematics optimisation for soft tissue artefact compensation. *Journal of Biomechanics*. 2017;62:95-101. doi:10.1016/j.jbiomech.2017.01.030
23. Lessmann N, van Ginneken B, de Jong PA, Išgum I. Iterative fully convolutional neural networks for automatic vertebra segmentation and identification. *Medical Image Analysis*. 2019;53:142-155. doi:10.1016/j.media.2019.02.005
24. Tibia segmentation in CT - Grand Challenge. grand-challenge.org. Accessed October 26, 2022. https://grand-challenge.org/algorithms/tibia-segmentation-in-ct/
25. Del Grosso VA, Mader CW. Speed of Sound in Pure Water. *The Journal of the Acoustical Society of America*. 1972;52(5B):1442-1446. doi:10.1121/1.1913258
26. Sollish BD. A Device for Measuring Ultrasonic Propagation Velocity in Tissue. In: *Ultrasonic Tissue Characterization II*. Vol 525. National Bureau of Standards Special Publication; 1979:53-56. https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nbsspecialpublication525.pdf
27. Bullen BA, Quaade F, Olesen E, Lund SA. Ultrasonic Reflections used for Measuring Subcutaneous Fat in Humans. *Human Biol*. 1965;37(4):375-384.
28. Errabolu RL, Sehgal CM, Bahn RC, Greenleaf JF. Measurement of ultrasonic nonlinear parameter in excised fat tissues. *Ultrasound in Medicine & Biology*. 1988;14(2):137-146. doi:10.1016/0301-5629(88)90181-0
29. Bamber JC, Hill CR. Ultrasonic attenuation and propagation speed in mammalian tissues as a function of temperature. *Ultrasound in Medicine & Biology*. 1979;5(2):149-157. doi:10.1016/0301-5629(79)90083-8
30. Medixant. RadiAnt DICOM Viewer. Version 2023.1. Accessed April 21, 2023. https://www.radiantviewer.com
31. Legland D. *Geom3d Library, Version 1.25*. The Mathworks Inc.; 2022. https://nl.mathworks.com/matlabcentral/fileexchange/24484-geom3d

# Appendix A Preliminary experiments

## A.1 Axial registration points of the A-mode ultrasound transducers

Introduction

The ultrasound (US) estimated depth is validated towards simultaneously conducted CT scans, assuming that the ground truth depth can be measured from the US transducer's tip to the underlying bone surface. However, a potential offset in the axial registration point (Figure A.1) of the US transducers could induce additional inaccuracies. Therefore, the aim of this preliminary study was to verify whether the A-mode US signal is registered at the surface or at some point within the transducers.

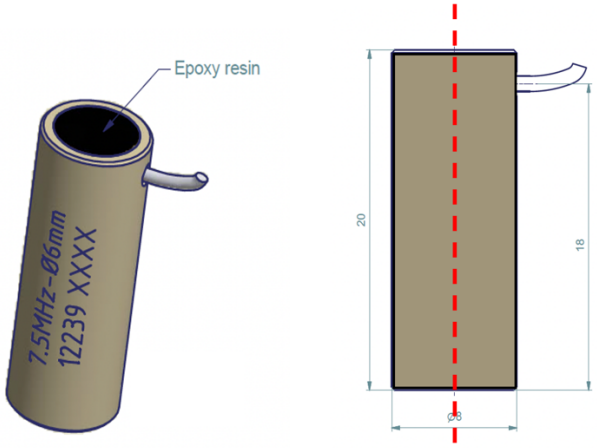



Figure A.1 The A-mode ultrasound (US) transducer 3D view (left) and side view (right). The axial registration point is defined as any point on the long axis of the transducer (red dashed line).

Methods

The current A-mode US setup includes a total of 30 transducers, and each of the 30 transducers was tested in an experimental setup (Figure A.2) with the transducer surface submerged in distilled water, facing the bottom of a glass container (Figure A.3). The setup includes a micrometer (Mitutoyo) and allows for measuring the distance from the transducer surface to the bottom of the glass with accuracy up to 0.002 mm. Using the speed of sound the time of peak ultrasound signal, the estimated distance can be calculated by

$$s = \frac{t*v}{2},$$

where $s$ is the distance in millimeters, $t$ is the time in seconds, and $v$ is speed of sound in millimeters/second within the medium. In this case, the used medium was distilled water, and in order to accurately determine the speed of sound, a thermometer was used to measure the temperature of the distilled water before each test. The thermometer measured with whole degree accuracy, hence the speed of sound wat determined as follows: for 20°C on thermometer display, 1482.343 m/s was chosen, which corresponds to 20.0°C, and for 21°C on thermometer display, 1485.372 m/s was chosen, which corresponds to 21.0°C according to *del Grosso et al.*[25]

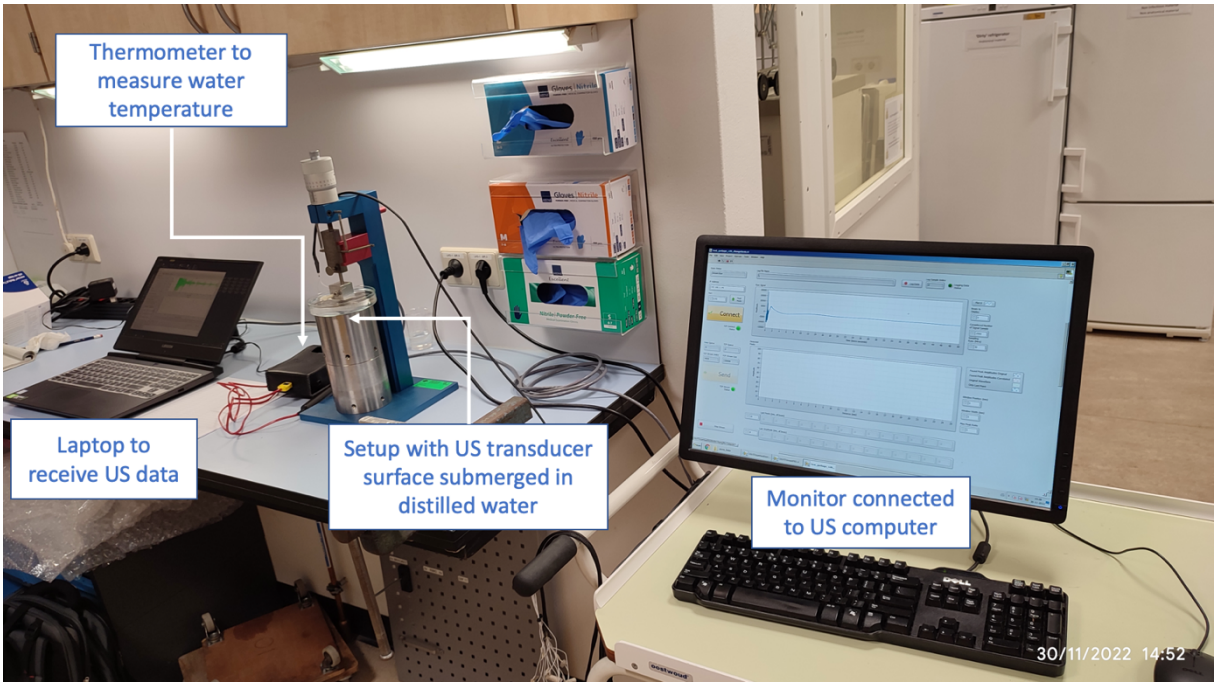


Figure A.2 Experimental setup to determine the axial registration point of the A-mode US transducers.

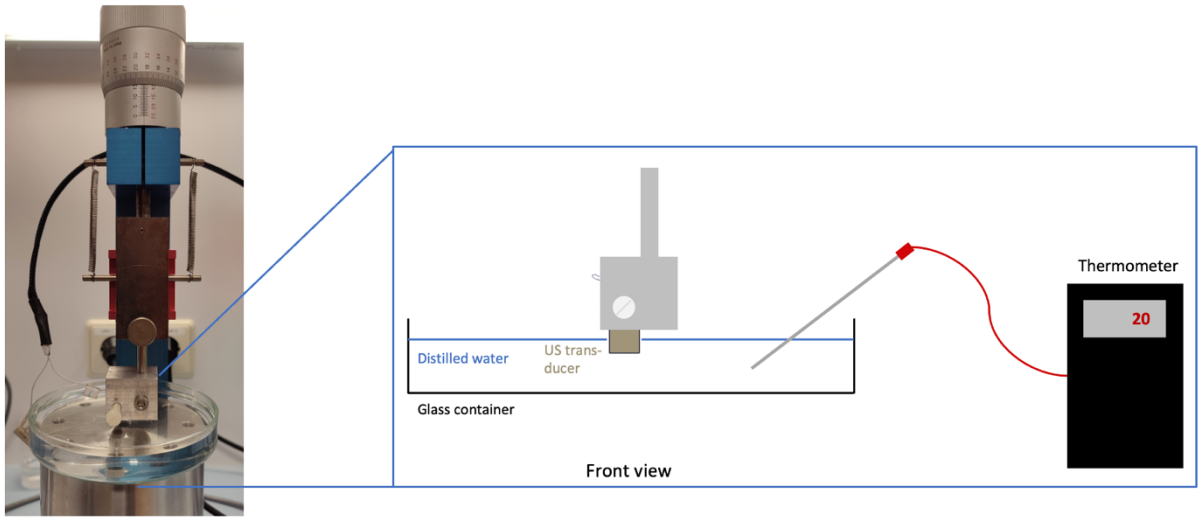


Figure A.3 A custom-made holder for the US transducer is connected to a micrometer that registers the true distance to the bottom of the glass container.

To determine the real distance from the US transducer's tip to the bottom of the glass container, the transducer was lowered until it touched the glass bottom. From that point, the distance from the transducer's tip to the glass bottom was increased to 10.000 mm, then lowered to 7.500 mm, and 5.000 mm. For each transducer, the A-mode US signal was recorded at the three aforementioned distances, for approximately 3 seconds per distance. The peak in the US was detected using the peak detection algorithm written by Dennis Christie in Matlab R2022b. The manual selected upper and lower bounds (UB and LB) for the peak detection algorithm were determined as follows: LB = 9.5 mm and UB = 10.5 mm for 10.000 mm distance, LB = 7 mm and UB = 8 mm for 7.500 mm distance, and LB = 4.5 mm and UB = 5.5 mm for 5.000 mm distance. In case the US estimated depth varied during one recording, the median of the distances was chosen. Finally, the difference between the US estimated depth and the real

distance from the US transducer's tip to the bottom of the glass contained is defined as the axial registration point (Figure A.4).

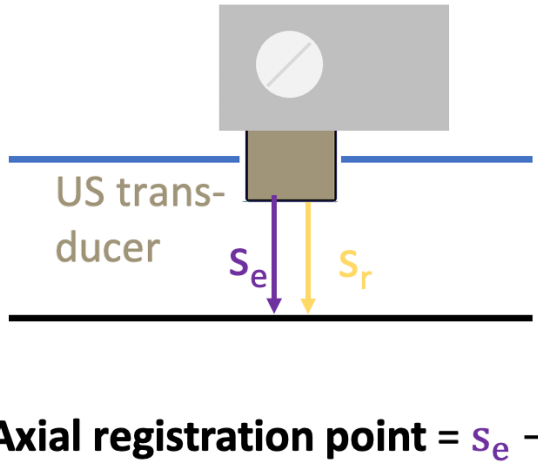



Figure A.4 The axial registration point is determined as the difference between the US estimated distance ($s_e$) and the real distance ($s_r$) from the tip of the transducer to the bottom of the glass container.

Results

Two of the 30 transducers had no signal, due to a hardware problem (transducer number 1 and number 9). The temperature of the distilled water was 20°C for all measurements except one (transducer number 30 at distance 10.000 mm). The difference between US estimated distance and real distance ranged from 0.0052 mm (transducer number 23) to 0.0941 mm (transducer number 21 and 28) (Figure A.5). The variability per transducer for the three distances was either zero or 0.0148 mm. For three transducers (number 14, 18 and 22), the US peak varied during the 3 seconds recording, resulting in a variance of 0.0148 mm.

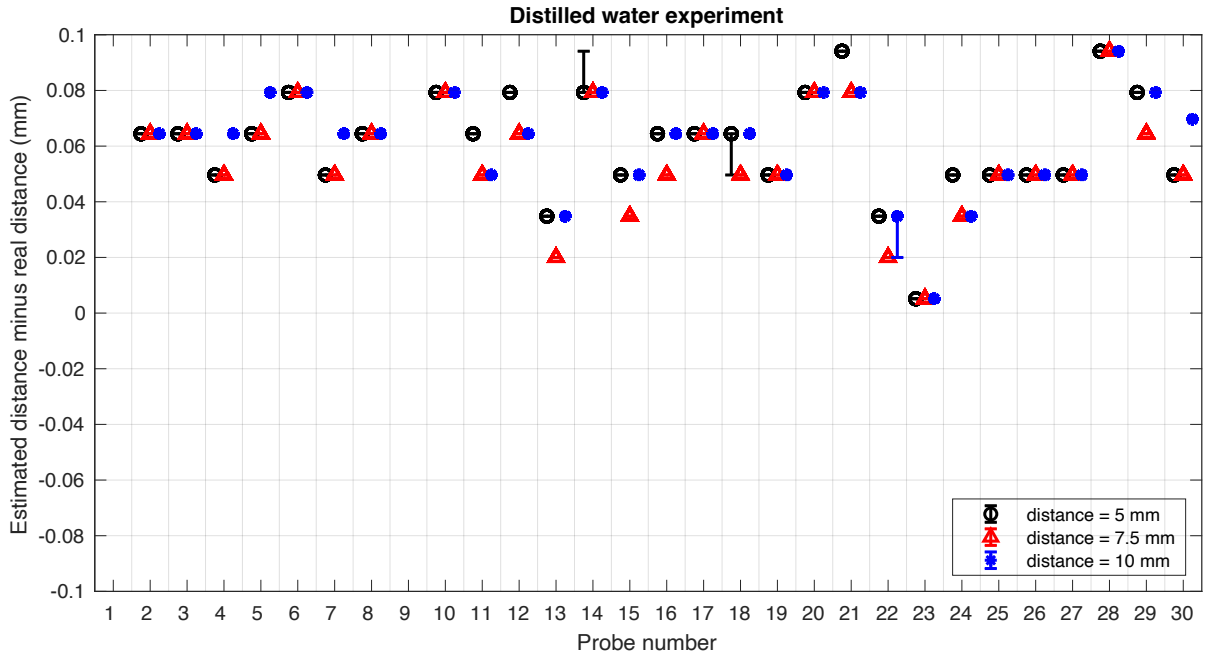



Figure A.5 Median and error bars (minimum and maximum) for the difference between US estimated distance and real distance per transducer (probe) for three distances (5 mm, black; 7.5 mm, red; 10 mm, blue).

Discussion

The aim of this preliminary study was to verify whether the A-mode US signal is registered at the surface or at some point within the transducers. All measurements show positive values, which indicated that the experiment is valid, since a negative value would represent an axial registration point outside the transducer. Overall, the axial registration point ranges from 0.0052 mm to 0.0941 mm above the transducer tip surface. One limitation of this study was the assumed speed of sound. Since the thermometer could only display whole numbers, 20°C could depict any value between 19.5°C and 20.5°C, and the speed of sound could range from 1480.798 m/s to 1483.868 m/s, respectively. Nevertheless, the introduced inaccuracies are minimal, 0.0053 mm, 0.0079 mm, and 0.0105 mm, for 5 mm, 7.5 mm, and 10 mm distance, respectively. Furthermore, the sample rate (50 MHz) of the transducers introduces a limitation in the step size of the distances. As can be seen in Figure A.5, the step size equals 0.0148 mm: $s = \frac{v*t}{2} = \frac{v*1}{2*f_s}\frac{1482.343*1}{2*50*10^6} = 0.0148\ \text{mm}$, with *s* the distance, *v* the speed of sound, *t* the peak time, and $f_s$ the sample rate.

*In conclusion,* the axial registration point is just above the transducer's tip surface and varies slightly for each transducer. With respect to the size of an A-mode US transducer (20 mm), the measured offsets in axial registration point are neglectable in the skin-to-bone depth validation experiments.

## A.2 Speed of sound in soft tissues

Introduction

The estimated skin-to-bone depth using A-mode ultrasound (US) requires a value for speed of sound as input for the calculations. The established value for speed of sound in human soft tissues is 1540 m/s, and originates from the study by *Ludwig* (1950) that measured the speed of sound through various thicknesses of calf muscles, thigh and biceps muscles.[18] Likewise, a commonly used value for speed of trough fat tissue is 1580 m/s, based on *in vivo* measurements on forearm muscles by *Sollish* (1979).[26] A standard value for speed of sound through human cadaveric tissue in vitro is lacking, and literature is scarce. *Bullen et al.* (1965) measured speed of sound through human abdominal fat tissue at 35°C, resulting in a mean value of 1476 m/s.[27] In addition, *Errabolu et al.* (1987) measured speed of sound in human fat tissues at temperatures ranging from 20°C to 37°C, resulting in a mean of 1430 m/s.[28] Nevertheless, both studies used excised fat tissues, not including the skin. Therefore, it is not representative of the A-mode US measurements in the current study with human cadaveric leg. Thus, the aim of this preliminary study is to verify what the speed of sound is in different soft tissues from a human cadaver leg.

Methods

The soft tissues of one spare human cadaver leg were used for this preliminary experiment. A total of three soft tissue pieces (approximately 5 * 5 cm) were cut out, including all the tissue up to the underlying bone tissue. The tissue pieces originate from the fibular head, medial tibial epicondyle, and lateral thigh, and composed of mixed fat and muscle, mostly fat tissue, and mostly muscle tissue, respectively (Figure A.6 and Table A.1). Using the setup with vertical micrometer of the first preliminary experiment, the distance from the transducer's tip to the bottom of the glass container is measured (Figure A.7). Next, with the peak time from the US signal, the speed of sound is calculated by:

$$v = \frac{s*2}{t},$$

where $s$ is the distance in millimeters, $t$ is the time in seconds, and $v$ is speed of sound in millimeters/second through the soft tissue.

Figure A.6 Soft tissue pieces, cut out of a human cadaver leg at the fibular head, medial tibial epicondyle and lateral thigh locations. The composition of the tissues differs slightly: mixed fat and muscle tissue, mostly fat tissue, and mostly muscle tissue.

Table A.1 Soft tissue piece numbers and corresponding locations and tissue composition

| | **Soft tissue piece 1** | **Soft tissue piece 2** | **Soft tissue piece 3** |
|---|---|---|---|
| Location | Fibular head | Medial tibial epicondyle | Lateral thigh |
| Tissue composition | Mixed fat and muscle | Mostly fat | Mostly muscle |

Transducer number 30 is used for all soft tissue measurements and each piece is measured at three different temperatures, first cold (±10°C) after ±2 hours at a 5°C fridge, then warm (±30°C) after ±30 minutes in the oven at 40°C, and finally at room temperature (±20°C). Per temperature, each piece is measured at two different spots without and with 3 mm compression of the tissue, resulting in a total of four measurements per temperature per piece. The temperature is measured using a thermometer with long metal pin, which is carefully poked in the tissue between the layers. The experiment was repeated on a separate day, and the method was adapted a little (Table A.2). The soft tissue pieces were the same for both measurements, and in between experiment days the pieces were kept in the freezer.

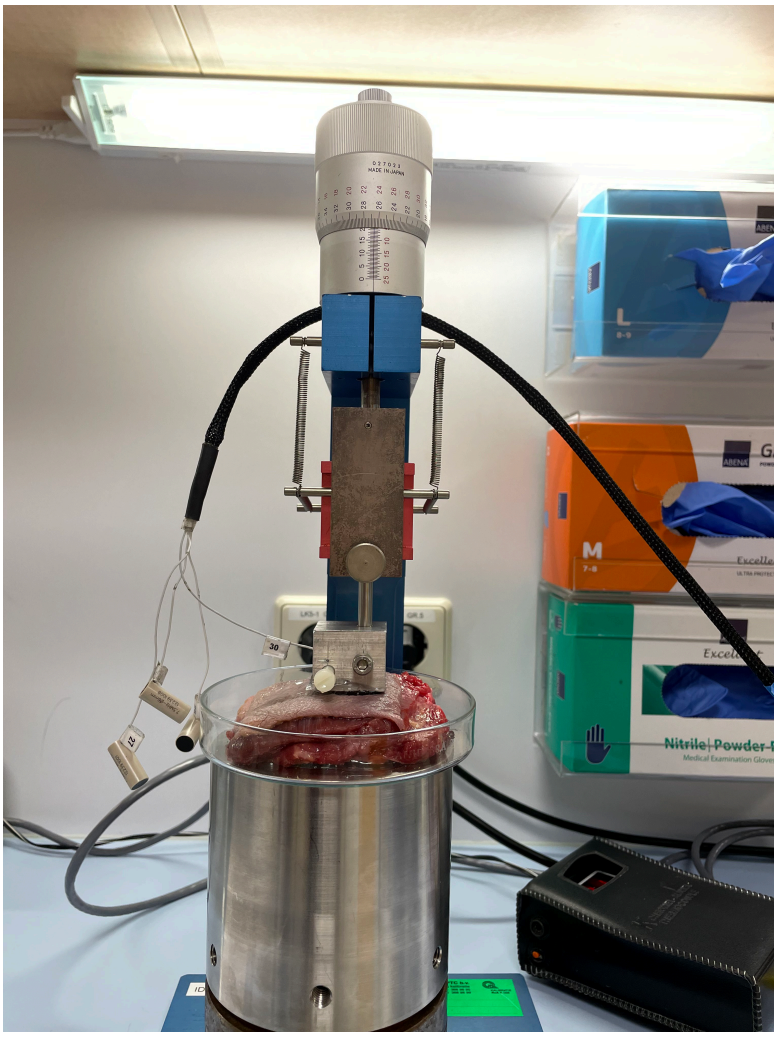



Figure A.7 Experimental setup for measuring the speed of sound through human cadaver soft tissues using an A-mode US transducer. The distance from the US transducer's tip to the bottom of the glass container is measured with the vertical micrometer.

The data is processed using Matlab R2022b, and speed of sound versus temperature results were used to compare the effect of compression. Next, compressed and non-compressed results were clustered to show median, minimum and maximum values for the speed of sound per cluster.

Table A.2 The experiment for measuring speed of sound in human cadaver soft tissues was conducted twice, on two separate days. Small changes were made to the methods.

| **Experiment 1** | **Experiment 2** |
|---|---|
| Probe 30 is used for all soft tissue measurements | Identical to experiment 1 |
| 4 measurements per piece per temperature (2 spots, first without and then with 3 mm compression) | 6 measurements per piece per temperature (3 spots, first without and then with 3 mm compression) |
| Order of measurements:<br>- Cold: after ~2h at 5 °C fridge<br>- Warm: after ~30 min at 40 °C in the oven<br>- Room: after ~1h at room temperature | Order of measurements:<br>- Cold: after ~40h at 5°C fridge<br>- Body temperature: after ~2h at 45°C in the oven<br>- Warm: right after the body temp measurements<br>- Room: after ~1.5h at room temperature |
| Temperature is measured using a thermometer with long metal pin, which is carefully poked in the tissue between the layers. | Identical to experiment 1 |

Results

For experiment 2, the additional step of measuring at body temperature (37°C ) was not feasible, since the soft tissue did not reach a temperature above 31°C despite putting the soft tissues in the oven for a longer period. The speed of sound values ranged from 1405 to 1520 m/s during experiment 1, with an outlier at 1555 m/s (Figure A.8, top), and from 1462 to 1584 m/s during experiment 2, with an outlier at 1604 m/s (Figure A.8, bottom). For experiment 1, the speed of sound was lower for the compressed measurements in all cases (18/18), whilst for experiment 2, the speed of sound was lower in most of the cases (20/27). Furthermore, median speed of sound values decreases with increasing temperature for mixed and fat tissue, and increases for muscle tissue, in both experiments (Figure A.9 top and bottom). Overall, the speed of sound values were higher during the second experiment than during the first experiment.

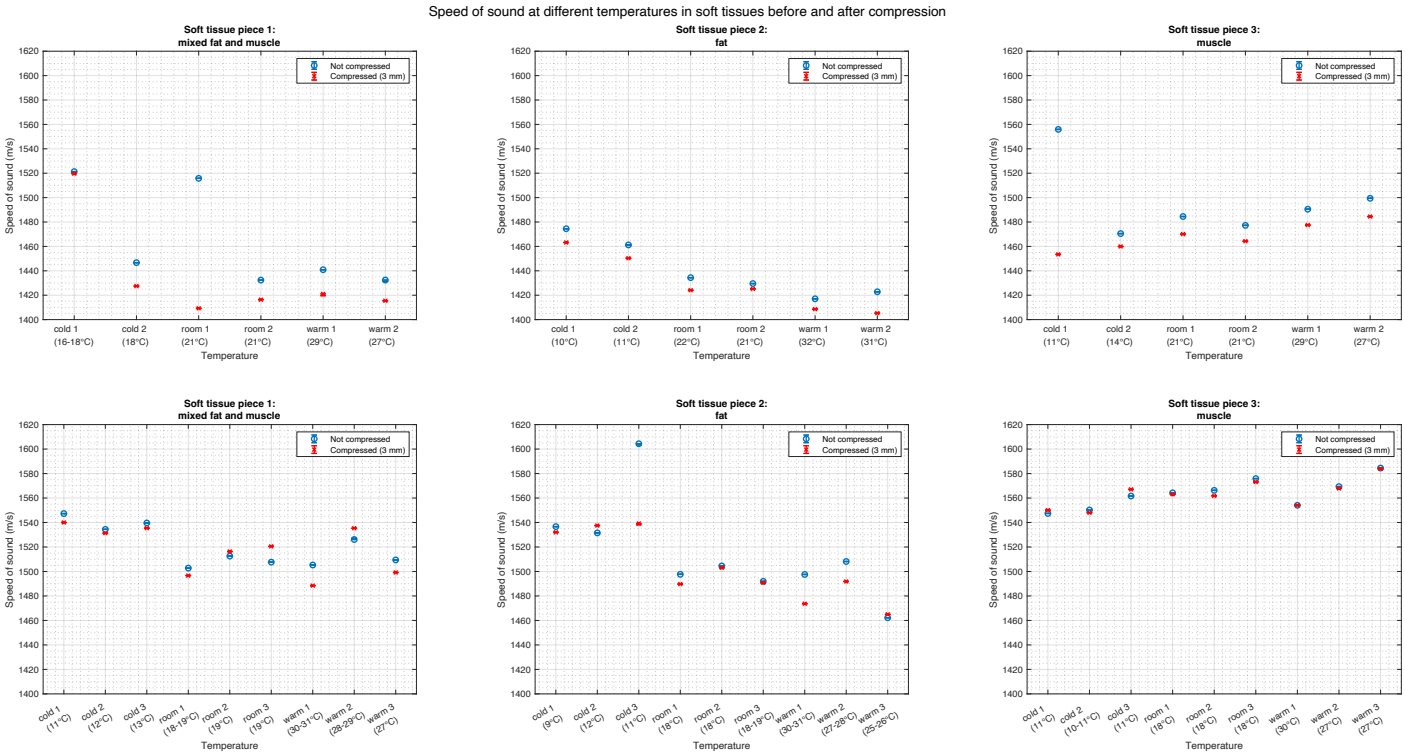



Figure A.8 Speed of sound versus temperature results per soft tissue piece (columns) for experiment 1 (top row) and experiment 2 (bottom row), comparing compressed (red) to non-compressed (blue) results.

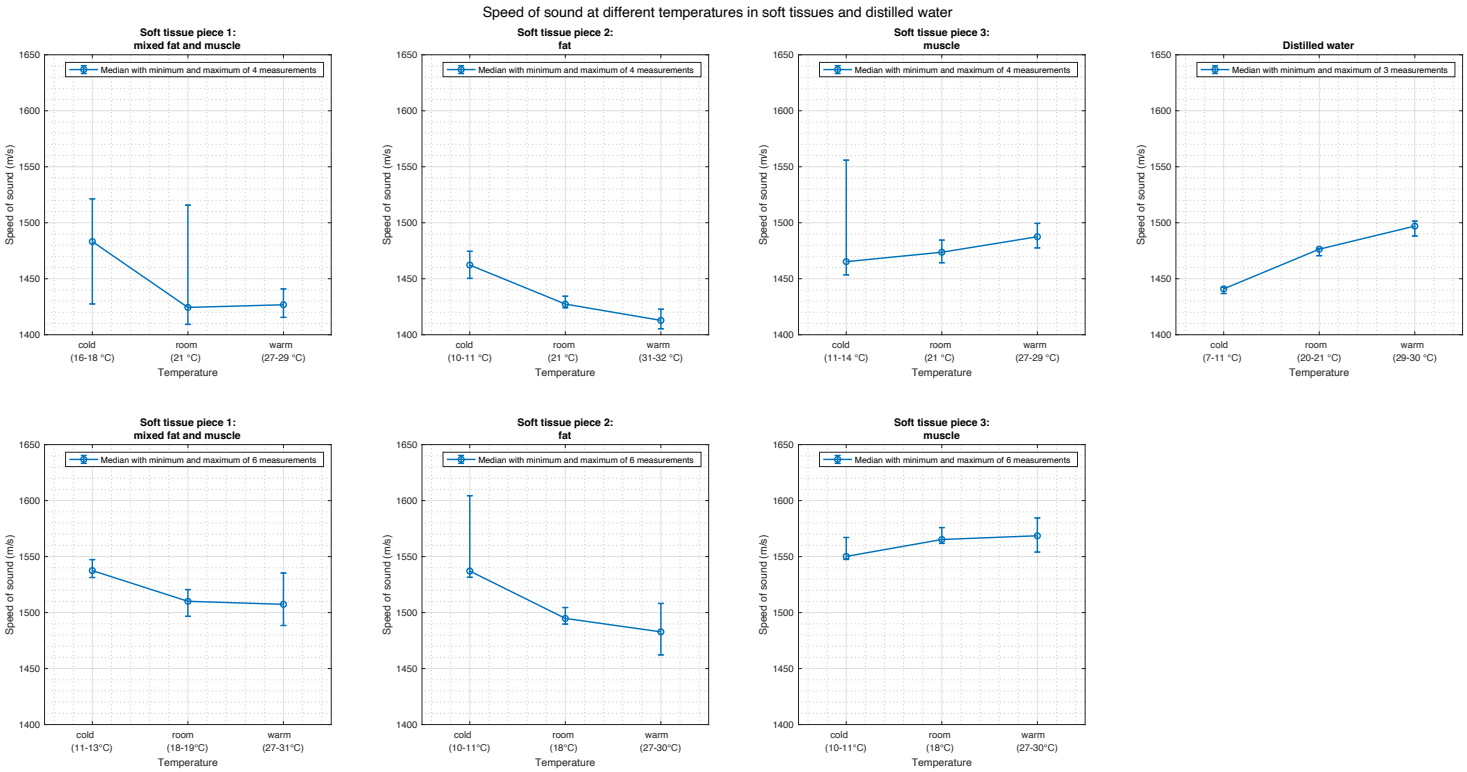



Figure A.9 Speed of sound versus temperature results per soft tissue piece and distilled water (columns) for experiment 1 (top row) and experiment 2 (bottom row), showing median (blue circle), and minimum and maximum values (blue error bars).

Discussion

The aim of this preliminary study was to verify what the speed of sound is in different soft tissues from a human cadaver leg. In general, the measured speed of sound (range 1405 – 1584 m/s) during the two experiments was lower than the established value for speed of sound in human soft tissues *in vivo* (1540 m/s).[18] Strikingly, the speed of sound during the second experiment was higher than during the first experiment. It is presumed that this is due to a change of properties of the tissue, caused by heating and freezing of the soft tissues. These processes cause dehydration, the tissues become more rigid, and the compressibility (B) decreases. Since $B = \frac{1}{K}$ with $K$ the bulk modulus, the equation for speed of sound ($v$) becomes: $v = \sqrt{\frac{K}{\rho}} = \sqrt{\frac{1}{\rho * B}}$ with ρ the density. Hence, the speed of sound increases with decreasing compressibility.

For most cases, the speed of sound decreased after compression of the tissue. This is potentially due to an increased density of the tissue after compression. Seven measurement showed increased speed of sound after compression, which is presumably due to a different measuring spot before and after compression, as a result of slight movement of the tissue during temperature measurement (poking with the thermometer pin). The positive correlation between the speed of sound and temperature for muscle tissue, and the negative correlation for the mixed and fat tissue pieces, were observed and consistent in both experiments. These results are in agreement with the temperature-dependence of excised human muscle,[29] and fat tissues.[28,29]

The main limitation of this preliminary study is that the cadaveric tissue used for this experiment deviates from the tissue of the cadaveric leg that is used for the A-mode US skin-to-bone depth estimation experiment. Tissues used for this preliminary experiment are potentially fresher, especially for experiment 1. Hence, the results might not directly correspond to the speed of sound in the soft tissues of the other cadaveric leg (that was thawed and frozen multiple times). Another limitation was that the tissue composition of the three soft tissue pieces could not be defined other than "mixed", "mostly fat" and "mostly muscle". This is an insurmountable problem, as a suitable method for determining the exact layers of soft tissue between the transducer's tip and the bottom of the glass container is lacking.

*In conclusion*, the results from this preliminary experiment show that the speed of sound in human cadaver tissue is generally lower than 1540 m/s. With regard to the future experiment to validate the skin-to-bone depth measurements by the A-mode US system, the results of soft tissue piece 1, mixed fat and muscle, at room temperature is of greatest interest. The median value ranges from 1424 m/s (experiment 1) to 1510 m/s (experiment 2). Hence, it is recommended to vary the speed of sound in the validation experiment from 1540 m/s to 1424 m/s.

# Appendix B Inter-method variability of the ground truth

Introduction

Determining the correct ground truth distance from the CT scan data is essential for adequate validation of the US depth estimation. However, defining the ground truth is not a straightforward task. The aim of this sub-study was to determine the inter-method variability of the ground truth distance, in order to verify to what extent the different methods affect the ground truth value. Therefore, a total of three different methods were elaborated and analyzed.

Methods

The bone surface depth ground truth is assumed to be the distance between the tip of the transducer that touches the skin, and the corresponding bone surface. Therefore, a hypothetical US beam line is defined as a virtual extension of the z-axis (long axis) of the US transducer towards the bone surface (Figure B.1). Four different methods were used to determine the ground truth: (1) manually using 3D Multiplanar Reconstruction (MPR), (2) fiducial based, and (3) transducer based, line method, and (4) transducer based, tube method.

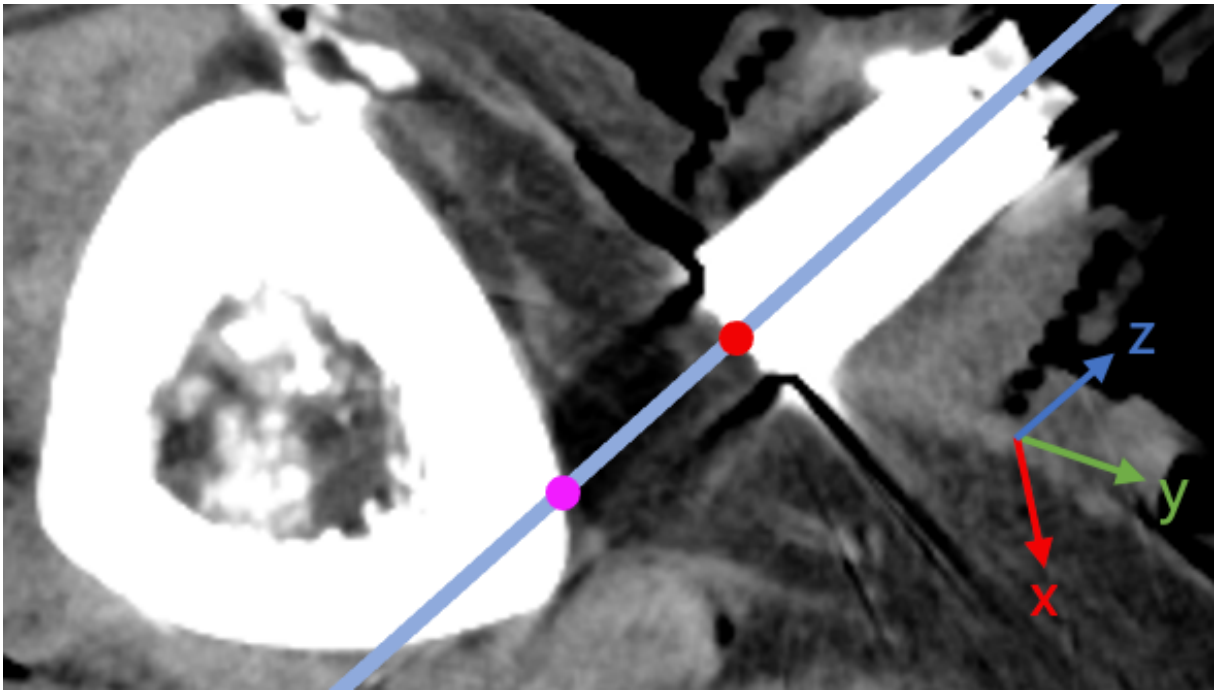



Figure B.1 Hypothetical beam line (light blue line) through the A-mode ultrasound transducer and the tibia. Bone surface depth ground truth is assumed to be the distance between the transducer's tip (red dot) and the corresponding tibia surface (magenta dot).

Firstly, for the method using MPR, the 3D MPR tool of RadiAnt DICOM Viewer[30] was used to manually adjust the three axes such that the reconstructed plane displays the longitudinal cross-section of a transducer. Thereby, the distance from the tip of each transducer to the corresponding bone surface could be measured manually using the measuring tool in the software (Figure B.2). This method also allows for manually measuring the depth of different layers of soft tissue if the quality of the CT scan image is adequate.

Secondly, for the fiducial based method (Figure B.3), the fiducials were used to determine the 3D pose (and thus the direction of the z-axis and hypothetical beam line) of the transducers. A

sphere was fit to each segmented fiducial, and the best fitting plane through the centers of the four fiducials was defined for every transducer. Next, the negative normal vector of the plane (directing towards the tibia), through the center of the plane, was defined as the z-axis and hypothetical beam line of the transducer. Subsequently, the intersection points of the hypothetical beam line and the tibia's 3D mesh were detected using the geom3d library in Matlab R2022b,[31] and the nearest point of intersection with respect to the transducer's tip is of interest (since the intersection further away represents the opposite site of the tibia). As the exact length of each polyamide screw, screwhead and transducer was previously measured (Appendix D.2), the offset is known and the distance between the US transducer's tip and bone surface could be calculated by:

$$s_{tt} = \sqrt{(x_{cp} - x_{ts})^2 + (y_{cp} - y_{ts})^2 + (z_{cp} - z_{ts})^2} - s_{offset}$$

where $s_{tt}$ is the distance from the transducer's tip to the tibia surface, $(x_{cp}, y_{cp}, z_{cp})$ the center point of the plane through the four fiducials, and $(x_{ts}, y_{ts}, z_{ts})$ the point on the tibia surface corresponding to the intersection of the hypothetical beam line and the tibia, and $s_{offset}$ the length of the polyamide screw and US transducer, minus the screwhead. An additional offset, due to a deviation of the axial measuring point of the transducer (i.e. the point in the transducer where the signal is registered is not exactly at the tip of the transducer), is neglected in the analysis based on the results of a preliminary experiment with the current setup and distilled water (Appendix A.1).

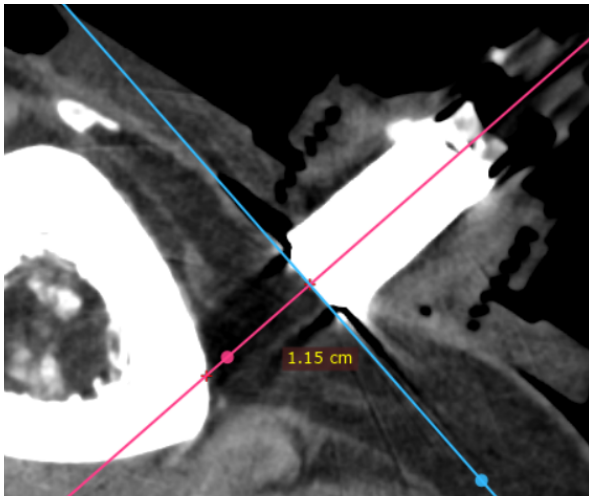



Figure B.2 Illustrative screenshot of the Multiplanar Reconstruction (MPR) method for determining the ground truth distance, using RadiAnt (CT2, transducer 26). The longitudinal (pink line) and transverse (blue line) axis are manually set in the center of the A-mode ultrasound transducer and at the tip of the transducer, respectively. The distance between the transducer's tip (blue line) and the tibia is manually measured between two selected points (pink crosses) and equals 11.5 mm.

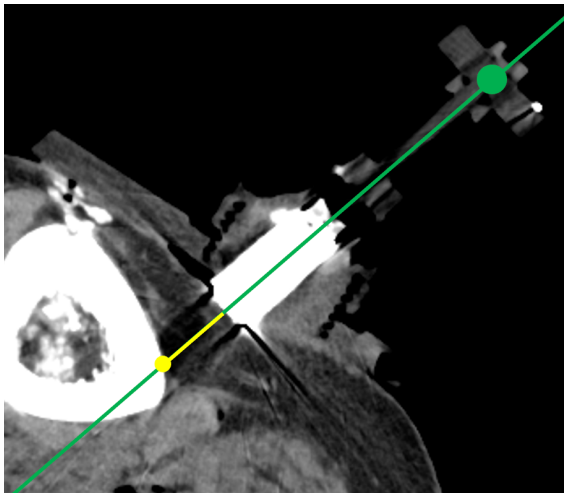


Figure B.3 Schematic 2D illustration of the fiducial based method. A best fit plane is determined trough the four fiducials on the platform. The normal vector of the plane is assumed to be the extension of the longitudinal axis (z-axis) of the transducer (green line). The center point (green dot) is defined as the intersection of the line and the plane through the fiducials. The ground truth distance (yellow line) is calculated as the distance between the intersection of the line with (1) the transducer's tip and (2) the tibia (yellow dot).

Thirdly, for the transducer based line method (Figure B.4), the central z-axis of the transducers was determined in 3D space using Principal Component Analysis (PCA) on the 3D meshes of each segmented transducer. The intersection of (1) the line that matches the first principal

component, and (2) the transducer's 3D mesh was detected using the geom3d library in Matlab R2022b[31] again, and was defined as the center point of the transducer's tip. Next, the nearest point of intersection with the 3D mesh of the tibia is defined as a point on the tibia surface $(x_{ts}, y_{ts}, z_{ts})$. Hence, for each origin point $(x_{op}, y_{op}, z_{op})$ on the transducer's tip, the distance to the underlying tibia surface was calculated by:

$$s_{tt} = \sqrt{(x_{op} - x_{ts})^2 + (y_{op} - y_{ts})^2 + (z_{op} - z_{ts})^2}$$

where $s_{tt}$ is the distance between the origin point on the transducer's tip and the underlying tibia, $(x_{op}, y_{op}, z_{op})$ the origin point at the tip of the US transducer, and $(x_{ts}, y_{ts}, z_{ts})$ the intersection of the hypothetical US beam line and the tibia. A potential additional offset due to a deviation of the axial measuring point of the transducer is neglected again (Appendix A.1).

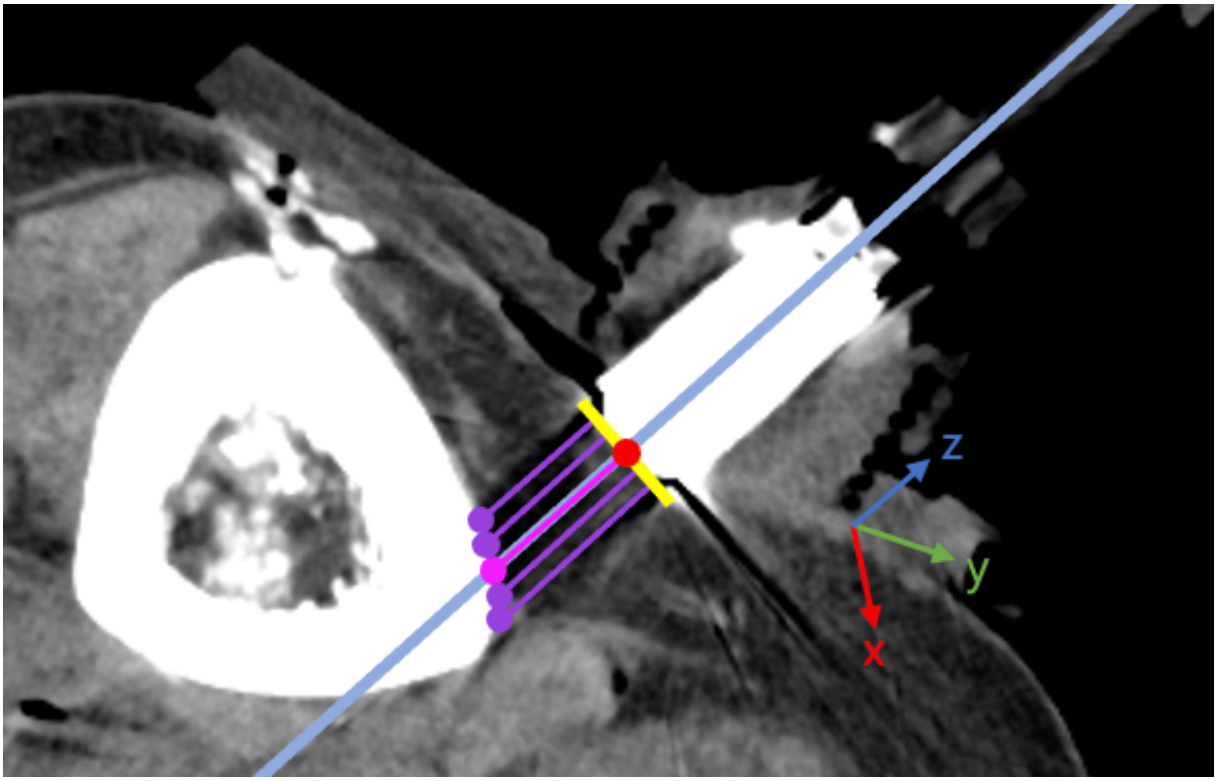



Figure B.4 Schematic 2D illustration of the line and tube method for determining the ground truth for one transducer. The center point of the transducer's tip (red dot) was defined as the intersection of the central z-axis of the transducer (blue line) and transducer's tip. For the line method, the distance from the center point to the intersection with the tibia surface (magenta dot) was defined using the central hypothetical US beam (magenta line). For the tube method, around the center point of the transducer's tip (red dot), a circle (here in 2D; yellow line) is reconstructed as starting point of a set of hypothetical US beam lines (purple lines) parallel to the central z-axis. The ground truth depth for the tube method is defined as the mean of the distances from the transducer's tip (yellow line) to the intersection points with the tibia (purple dots).

Finally, for the transducer based tube method (Figure B.4), the line method is extended. Therefore, around the center point at the tip of the transducer, 112 additional points were defined within a circle (radius 3 mm) in the x/y-plane of the transducer, resulting in a total of 113 origin points for the hypothetical US beam lines (Figure B.5). For each coordinate, the distance to the bone surface is determined parallel to hypothetical beam line, using the same equation as for the line method. In the end, a total of 113 distances per transducer were determined, of which the mean value was selected as bone surface depth ground truth.

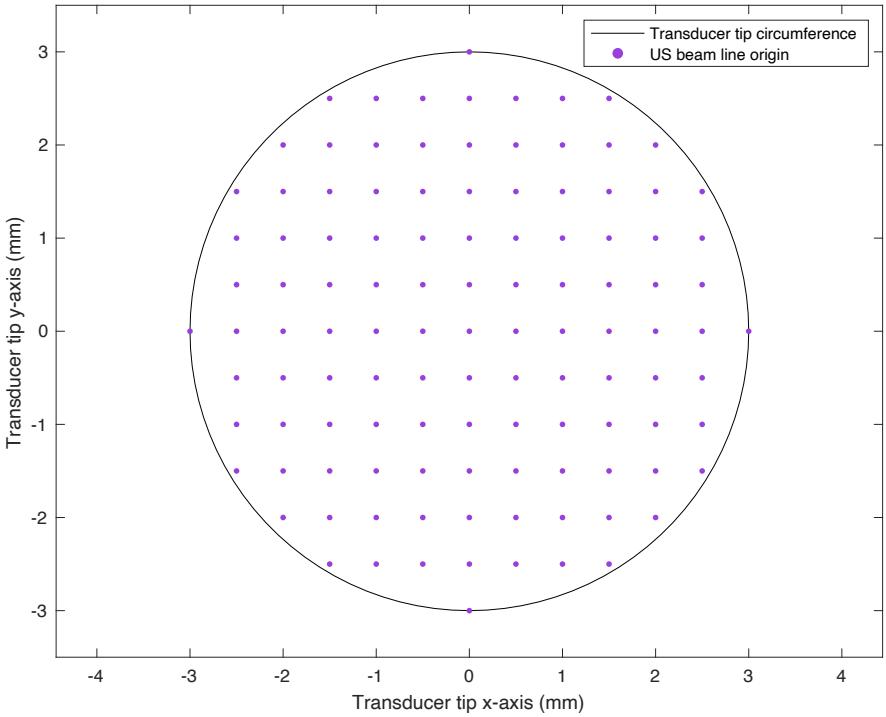



Figure B.5 Distribution of the 113 origin points (magenta dots) on the US transducer's tip in 2D (x/y-plane). Hypothetical US beam lines are defined parallel to the z-axis, from the origin points on the US transducer's tip towards the underlying tibia surface. For each transducer, the mean of 113 distances is used as bone surface depth ground truth.

Results

The data of the two CT scans were used in the analysis, with nine transducers per CT scan and four different methods for determining the ground truth. Of the results, the data for transducer 20 was excluded since the transducer did not face the tibia. For CT1 transducer 21, the ground truth could only be determined using the tube method, since the MPR and line method did not have an intersection with the tibia as well. In addition, all data from the fiducial based method was excluded in the analysis, since the determined center point at the tip of the transducer was clearly off-center during 3D visualization. Hence, for the remaining 15 transducers the ground truth could be determined and analyzed using three methods (MPR, line method and tube method), yielding a total of 45 measurements.

Ground truth values and differences between methods for every transducer are displayed in Table B.1 and Figure B.6. The smallest mean absolute difference in ground truth was identified for the line versus MPR method (0.1573±0.1109 mm), followed by tube versus line method (0.2689±0.3956 mm), and the largest mean absolute difference was found for the tube versus MPR method (0.3040±0.3936 mm).

Table B.1 Ground truth values, obtained using three different methods: tube method (GT_tube), line method (GT_line) and Multiplanar Reconstruction (MPR, GT_MPR). Direct comparison of absolute differences in ground truth values between the three methods. N/A = not applicable; SD = standard deviation.

| Transducer | GT_tube (mm) | GT_line (mm) | GT_MPR (mm) | Tube vs line (abs diff, mm) | Tube vs MPR (abs diff, mm) | Line vs MPR (abs diff, mm) |
|---|---|---|---|---|---|---|
| 20, CT1 | N/A | N/A | N/A | N/A | N/A | N/A |

| | | | | | | |
|---|---|---|---|---|---|---|
| 21, CT1 | 17.3156 | N/A | N/A | N/A | N/A | N/A |
| 24, CT1 | 9.4835 | 9.3270 | 9.36 | 0.1565 | 0.1235 | 0.0330 |
| 25, CT1 | 4.8964 | 4.8260 | 4.74 | 0.0704 | 0.1564 | 0.0860 |
| 26, CT1 | 6.0112 | 5.9439 | 6.02 | 0.0673 | 0.0088 | 0.0761 |
| 27, CT1 | 5.4924 | 5.5214 | 5.48 | 0.0290 | 0.0124 | 0.0414 |
| 28, CT1 | 15.0566 | 15.1283 | 14.8 | 0.0717 | 0.2566 | 0.3283 |
| 29, CT1 | 3.3755 | 3.4026 | 3.5 | 0.0271 | 0.1245 | 0.0974 |
| 30, CT1 | 3.1193 | 2.4872 | 2.56 | 0.6321 | 0.5593 | 0.0728 |
| 20, CT2 | N/A | N/A | N/A | N/A | N/A | N/A |
| 21, CT2 | 16.3828 | 16.5585 | 16.4 | 0.1756 | 0.0172 | 0.1585 |
| 24, CT2 | 12.0774 | 13.0258 | 13.4 | 0.9484 | 1.3226 | 0.3742 |
| 25, CT2 | 5.8581 | 5.7478 | 5.69 | 0.1103 | 0.1681 | 0.0578 |
| 26, CT2 | 11.6736 | 11.5916 | 11.5 | 0.0820 | 0.1736 | 0.0916 |
| 27, CT2 | 5.6386 | 5.6554 | 5.37 | 0.0168 | 0.2686 | 0.2854 |
| 28, CT2 | 16.2134 | 16.0796 | 16.2 | 0.1338 | 0.0134 | 0.1204 |
| 29, CT2 | 3.9389 | 4.0319 | 3.75 | 0.0930 | 0.1889 | 0.2819 |
| 30, CT2 | 5.3259 | 3.9059 | 4.16 | 1.4199 | 1.1659 | 0.2541 |
| | | | Mean | 0.2689 | 0.3040 | 0.1573 |
| | | | SD | 0.3956 | 0.3936 | 0.1109 |

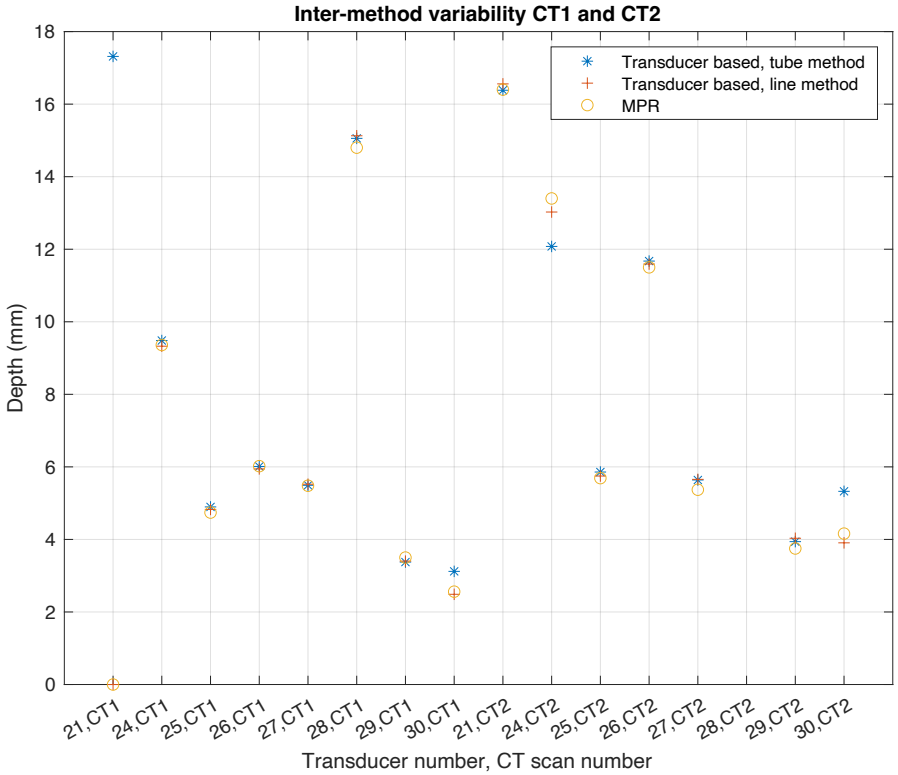



Figure B.6 Ground truth distance (bone surface depth from transducer tip to tibia surface) from two CT scan data sets, with nine transducers in each scan. MPR = Multiplanar Reconstruction.

Discussion

Three different methods were analyzed and compared to find the inter-method variability. Overall, the line method best corresponds to both the MPR and the tube method, where the mean difference between line and MPR method was the smallest. This can be explained by the fact that both methods rely on one distance, measured from the center of the transducer's tip to the tibia surface. The difference is that the line method involves the 3D meshes of the transducers and the tibia, whereas the MPR method is a fully manual method. Apparently, the mean difference between these methods is small (0.1573±0.1109 mm) and within the size of a voxel (0.20 * 0.20 * 0.25 mm). With regard to the tube versus line method, the mean absolute difference is almost twice as large with an even larger standard deviation (0.2689±0.3956 mm). An explanation for these results can be that the tube method yields a mean of 113 distances. Thereby, the tube method includes distances that surround the central hypothetical beam line, and could therefore deviate from the result of the line method, especially when the transducer faces the tibia at a non-perpendicular angle. Three outliers (CT1 transducer 30, CT2 transducer 24, and CT2 transducer 30) are identified. Visual inspection revealed that these elevated differences are potentially due to a curved tibia surface and/or metal artefacts in the CT scan image for all outliers. The largest mean absolute difference was found for the tube versus MPR method (0.3040±0.3936 mm), which can be attributed to a combination of the factors that contribute to both the line versus tube, and the line versus MPR differences. For this comparison, two outliers were identified (CT 2 transducer 24, and CT2 transducer 30), due to the same reasons as the outliers of the tube versus line method comparison.

A fourth method, the fiducial based method, was not included in the analysis because the determined center point at the tip of the transducer was clearly off-center when the 3D meshes with the defined points and lines were visually checked by two observers. This deviation is due to several factors. First of all, the fiducials are manually glued to the platform, and therefore, the four fiducials are not exactly positioned in one plane. The plane that is fit in Matlab could therefore be not exactly perpendicular to the true long axis of the transducer. Secondly, since the segmentations of the fiducials were not exactly spherical, a sphere was fit to find the center of the fiducials. This could also induce inaccuracy in the 3D pose of the fitted plane and thus the direction of the reconstructed line towards the tibia surface. Thirdly, the plastic bolt that connects the platform to the ball-joint casing, and presses the transducer to the skin, might not be exactly in line with the long axis of the transducer since the screw thread in the ball-joint casing was also tapped manually.

*In conclusion,* the fiducial based method for determining the bone surface ground truth depth is not adequate due to several inaccuracies. The mean absolute difference in ground truth depth between the three methods that were analyzed ranges from 0.1573 to 0.3040 mm, which corresponds to the size of approximately 0.5 to 1.5 voxel. Even though the line method corresponds best to the other two methods (tube method and MPR) and could therefore be considered a reliable choice for determining the ground truth depth, the tube method is favored as superior method based on the logical reasoning that taking the mean of several distances within the diameter of the transducer's tip is most representative of the A-mode ultrasound working principle for depth estimation.

# Appendix C Supplementary tables and figures

## C.1 Supplementary tables

Table C.1 Ground truth depth values for the tube method, using the segmentations of two independent observers (DC: Dennis Christie; MD: Maxime Devos), and the absolute difference between the ground truth depth values with mean and standard deviation (SD) values.

| Transducder | DC s_gt (tube mean, mm) | MD s_gt (tube mean, mm) | Absolute difference (tube mean, mm) |
|---|---|---|---|
| 20, CT1 | | | |
| 21, CT1 | 17.4226 | 17.3156 | 0.1070 |
| 24, CT1 | 9.3972 | 9.4835 | 0.0863 |
| 25, CT1 | 4.8718 | 4.8964 | 0.0246 |
| 26, CT1 | 5.975 | 6.0112 | 0.0362 |
| 27, CT1 | 5.2833 | 5.4924 | 0.2091 |
| 28, CT1 | 14.8368 | 15.0566 | 0.2198 |
| 29, CT1 | 3.2971 | 3.3755 | 0.0784 |
| 30, CT1 | 3.2736 | 3.1193 | 0.1543 |
| | | **Mean (excl. #30)** | **0.1088** |
| | | **SD** | **0.0719** |

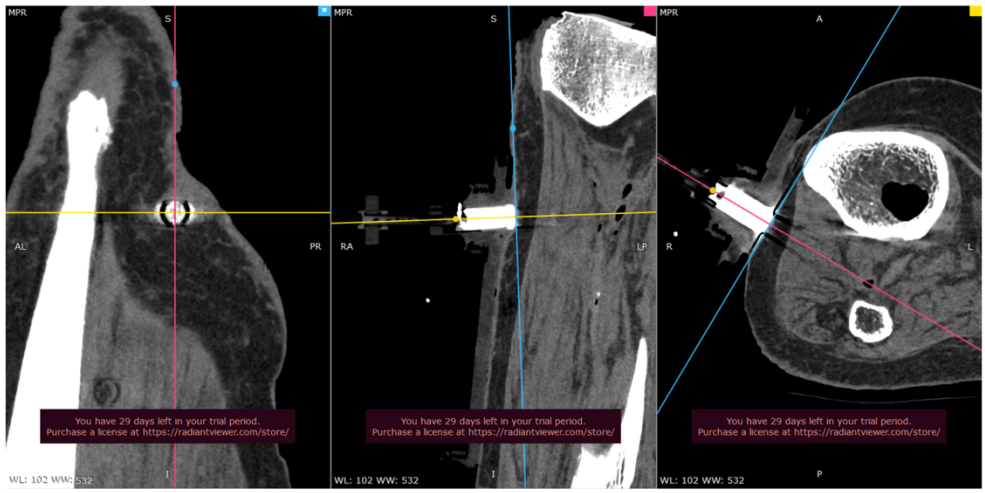
Table C.2 Results for the US estimated depths (using speed of sound 1540 m/s (green) and 1424 m/s (yellow)), ground truth depths (using the tube method, and segmentations from MD (purple)), differences, and absolute differences per transducer in both CT scans. Overall mean and standard deviation (SD) for the absolute differences are listed at the bottom.

| Trans-ducer # | US estimated depth (mm), using 1540 m/s | US estimated depth (mm), using 1424 m/s | CT ground truth depth (mm) | Difference $US_{1540}$ and $CT_{gt}$ (mm) | Absolute dif-ference $US_{1540}$ and $CT_{gt}$ (mm) | Difference US and $CT_{gt}$ (mm) | Absolute dif-ference $US_{1540}$ and $CT_{gt}$ (mm) |
|---|---|---|---|---|---|---|---|
| 21, CT1 | 18.4184 | 17.03104 | 17.3156 | 1.1028 | 1.1028 | -0.2846 | 0.2846 |
| 24, CT1 | 10.3642 | 9.58352 | 9.4835 | 0.8807 | 0.8807 | 0.1000 | 0.1000 |
| 25, CT1 | 5.9136 | 5.46816 | 4.8964 | 1.0172 | 1.0172 | 0.5717 | 0.5717 |
| 26, CT1 | 6.8684 | 6.35104 | 6.0112 | 0.8572 | 0.8572 | 0.3398 | 0.3398 |
| 27, CT1 | 6.6682 | 6.16592 | 5.4924 | 1.1758 | 1.1758 | 0.6735 | 0.6735 |
| 28, CT1 | 15.6464 | 14.46784 | 15.0566 | 0.5898 | 0.5898 | -0.5887 | 0.5887 |
| 29, CT1 | 4.1888 | 3.87328 | 3.3755 | 0.8133 | 0.8133 | 0.4978 | 0.4978 |
| 21, CT2 | 17.8948 | 3.78784 | 16.3828 | 1.5120 | 1.5120 | 0.1641 | 0.1641 |
| 24, CT2 | 12.7204 | 32.93712 | 12.9976 | -0.2772 | 0.2772 | -1.2354 | 1.2354 |
| 25, CT2 | 6.1446 | 16.54688 | 5.8581 | 0.2865 | 0.2865 | -0.1763 | 0.1763 |

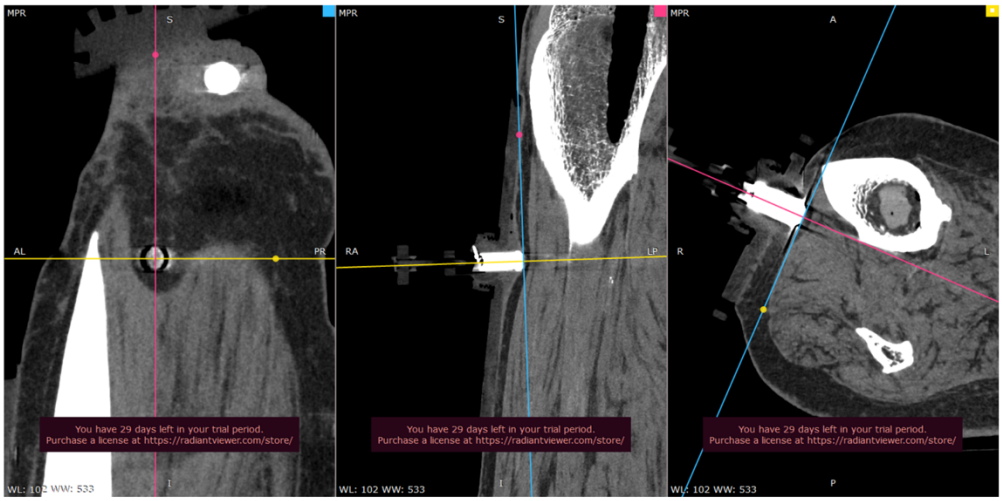
| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 26, CT2 | 13.1362 | 11.76224 | 11.6736 | 1.4626 | 1.4626 | 0.4731 | 0.4731 |
| 27, CT2 | 6.6836 | 5.68176 | 5.6386 | 1.0450 | 1.0450 | 0.5416 | 0.5416 |
| 28, CT2 | 14.3990 | 12.14672 | 16.2134 | -1.8144 | 1.8144 | -2.8990 | 2.8990 |
| 29, CT2 | 3.9732 | 6.18016 | 3.9389 | 0.0343 | 0.0343 | -0.2650 | 0.2650 |
| | | | | **Mean** | **0.9192** | **Mean** | **0.6293** |
| | | | | **SD** | **0.4850** | **SD** | **0.6872** |

C.2 Supplementary figures

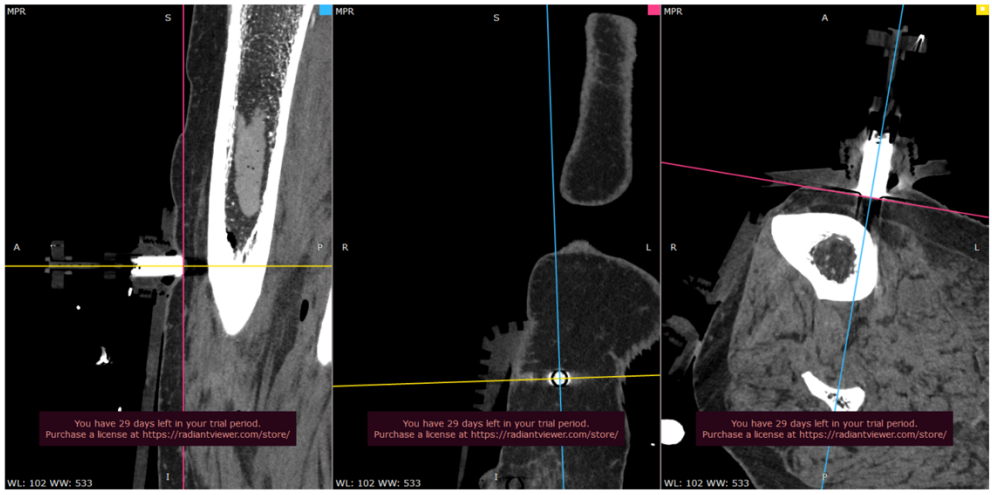
## CT1 transducer 20



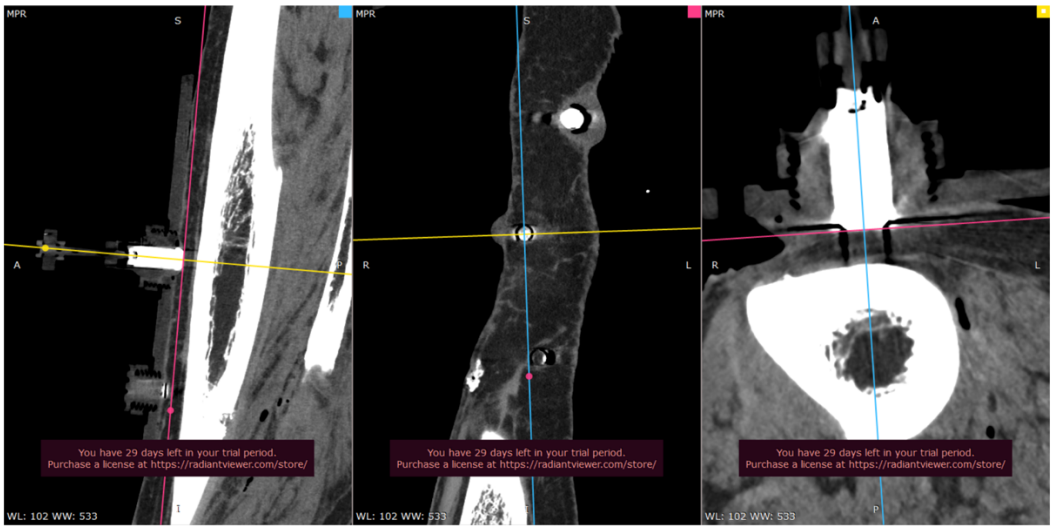

## CT1 transducer 21



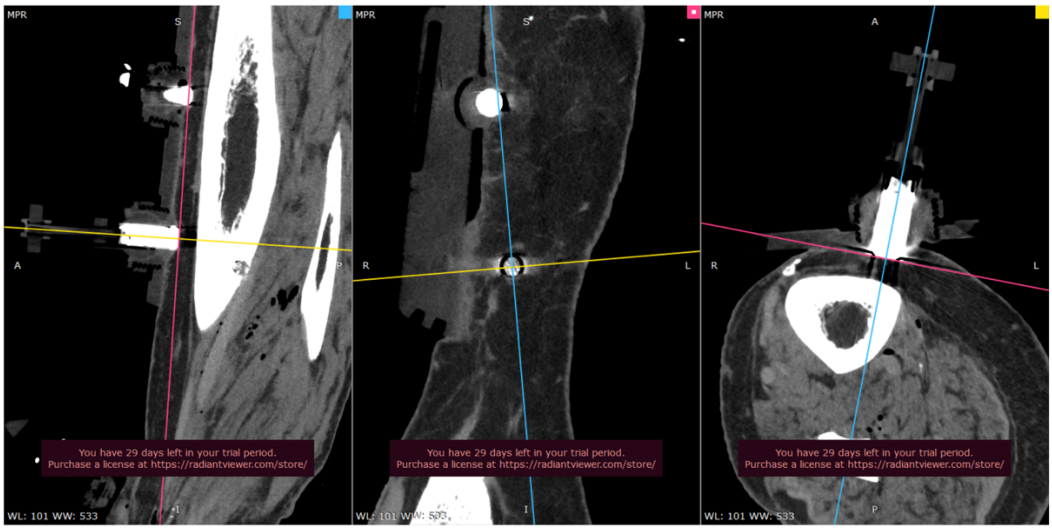

## CT1 transducer 24



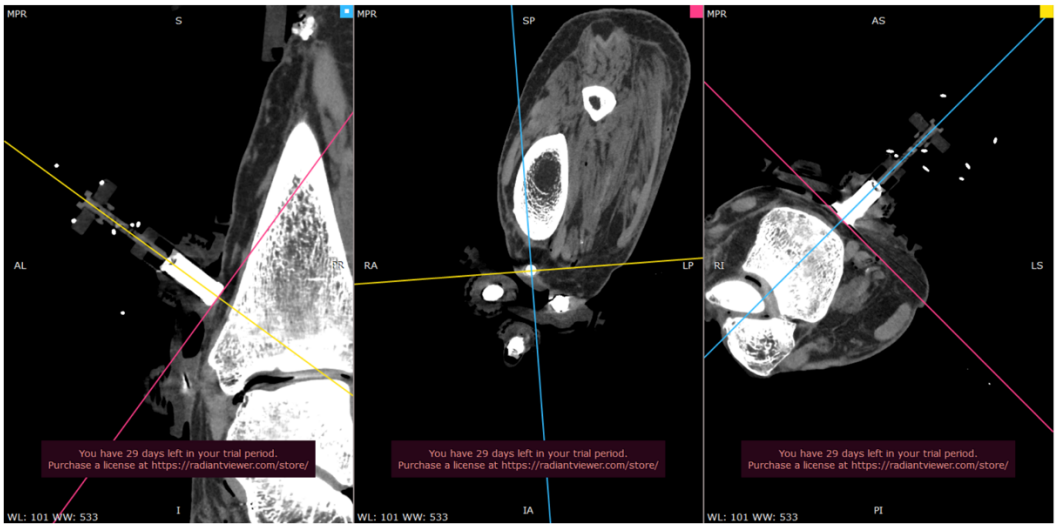
# CT1 transducer 25



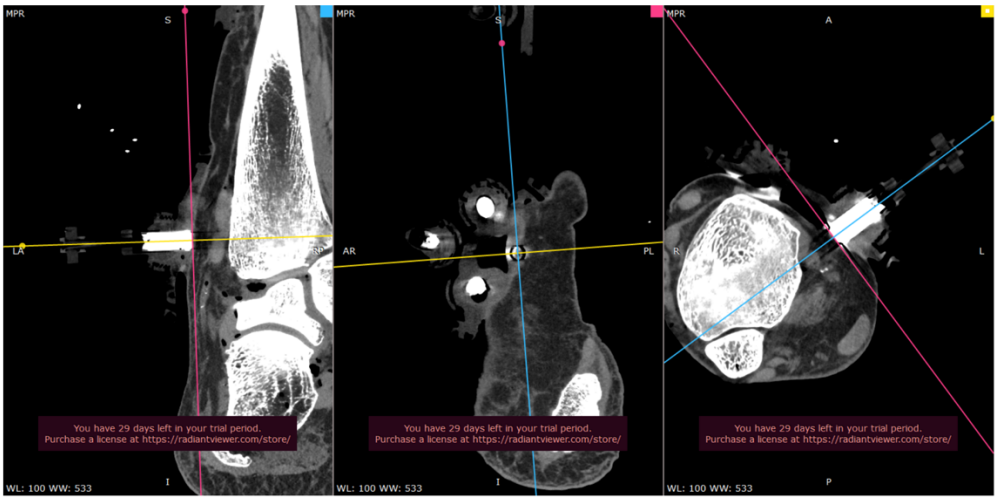

# CT1 transducer 26



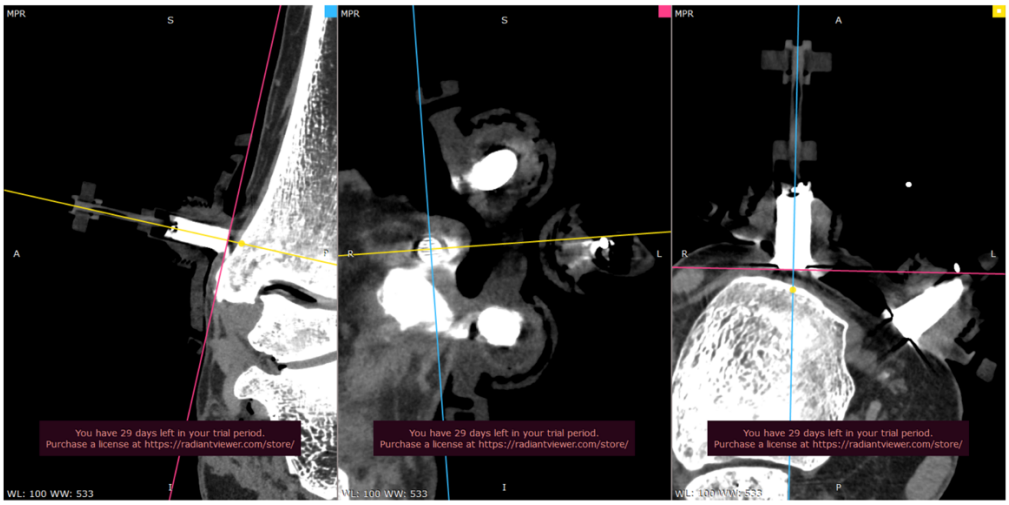

# CT1 transducer 27



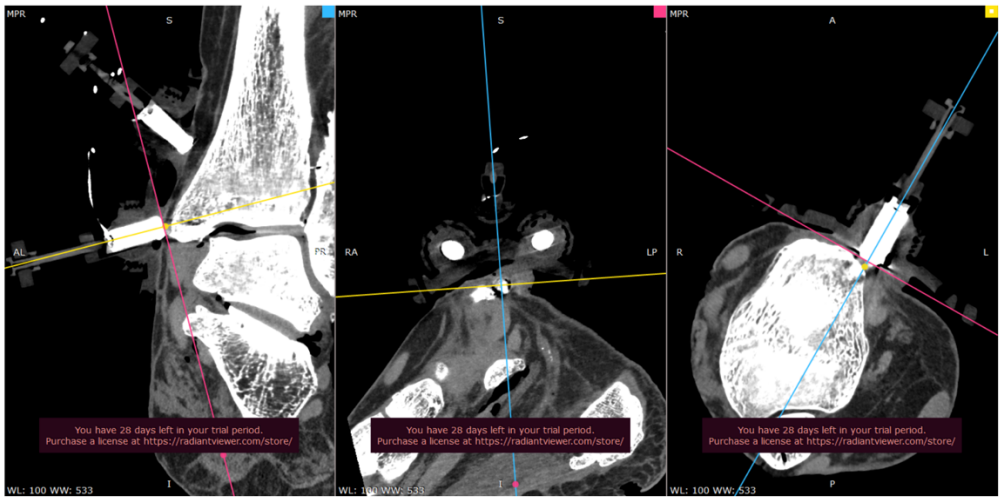
CT1 transducer 28



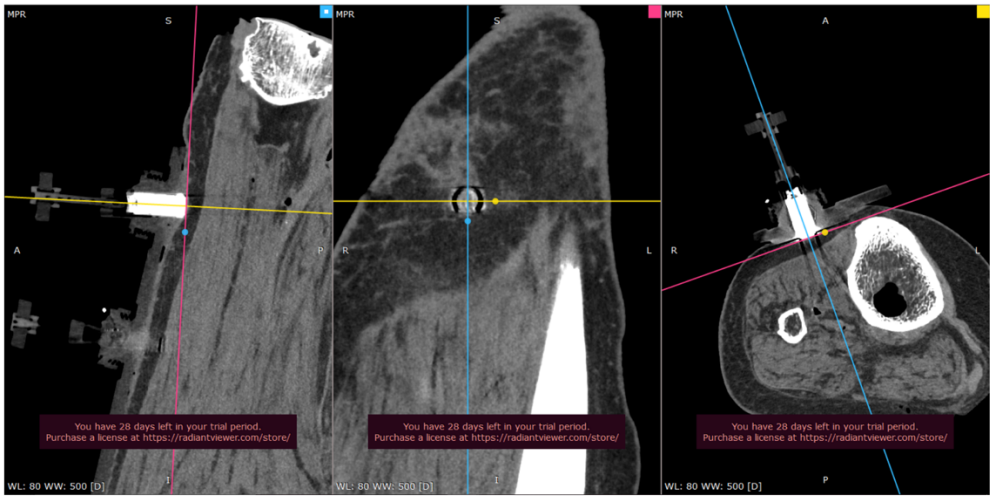

CT1 transducer 29



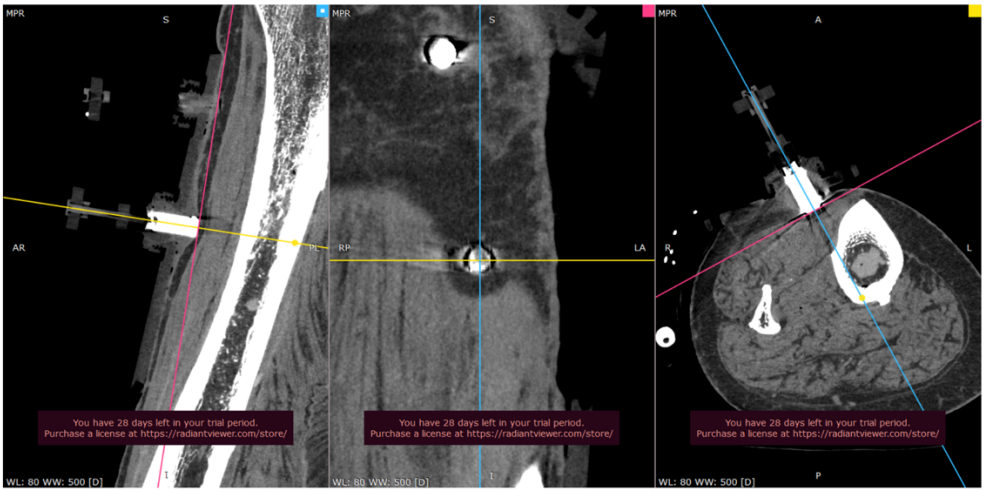

CT1 transducer 30



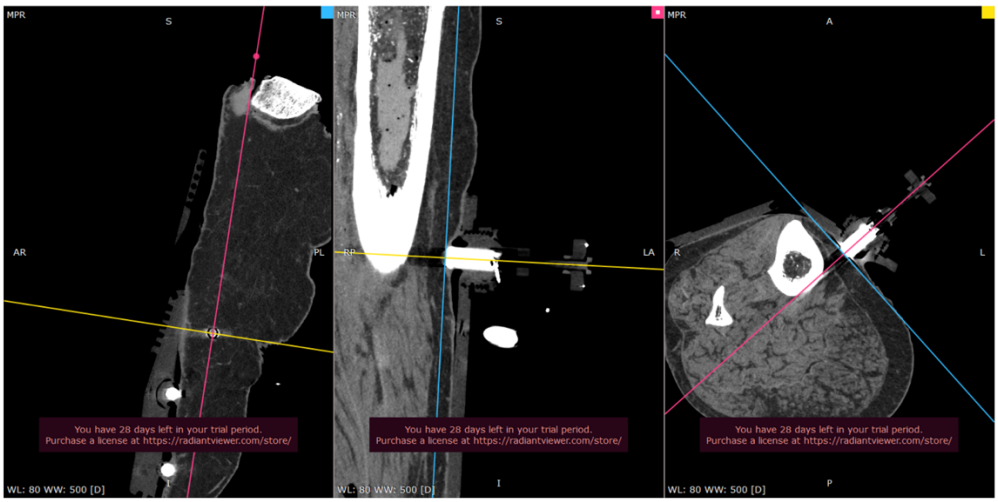
# CT2 transducer 20



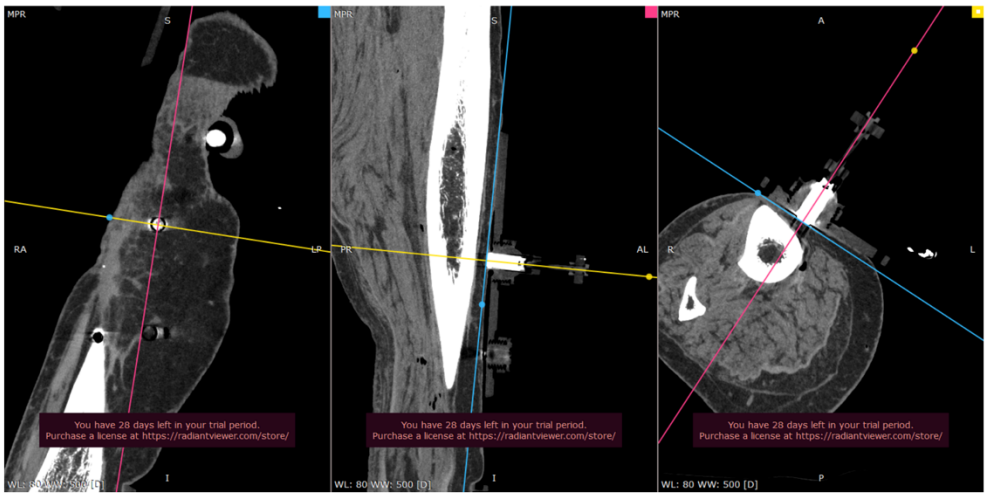

# CT2 transducer 21



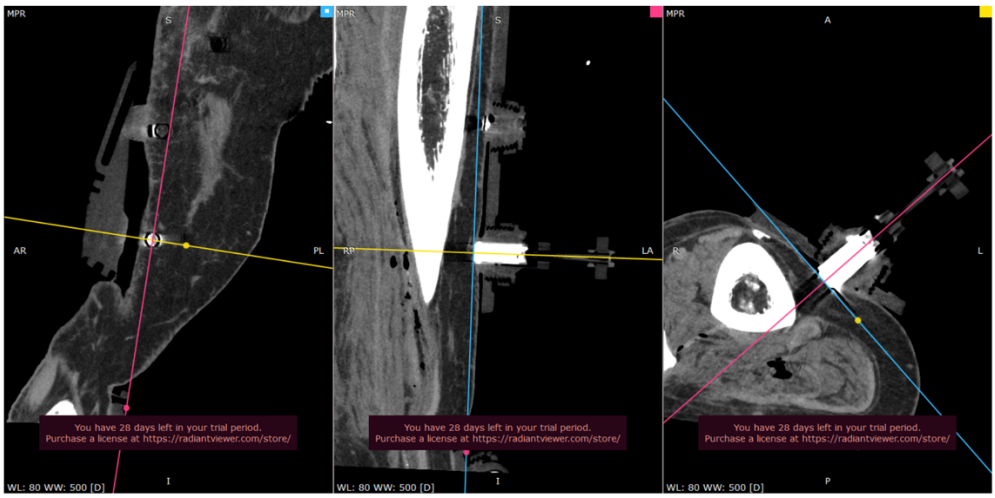

# CT2 transducer 24



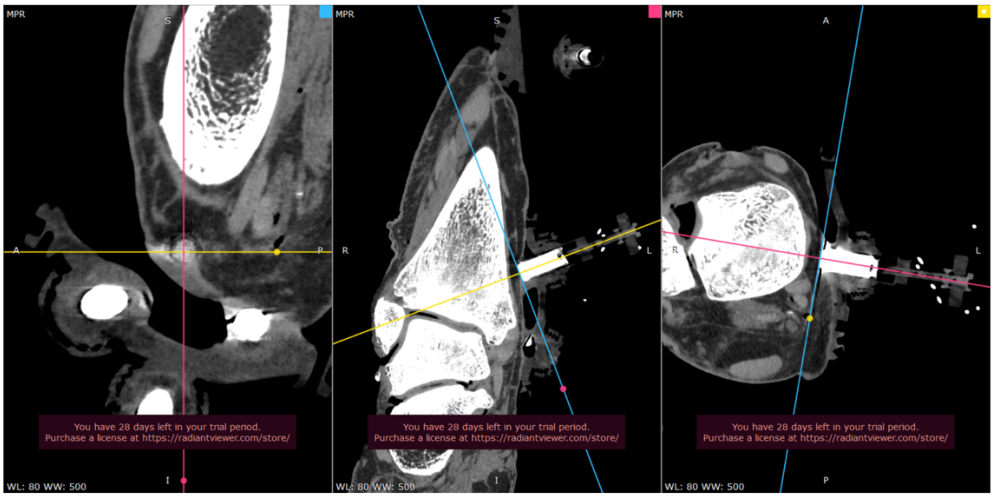
CT2 transducer 25



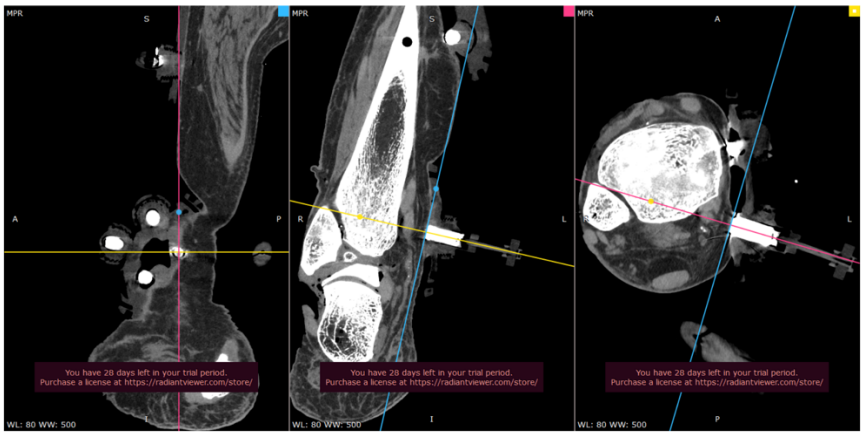

CT2 transducer 26



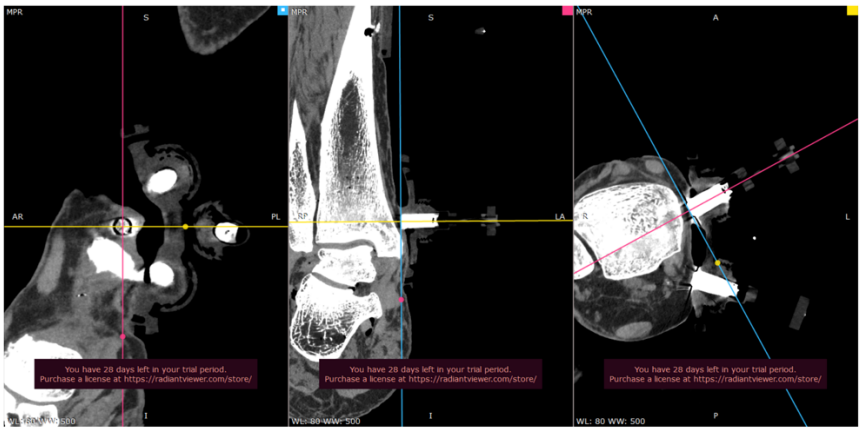

CT2 transducer 27



CT2 transducer 28




CT2 transducer 29




CT2 transducer 30

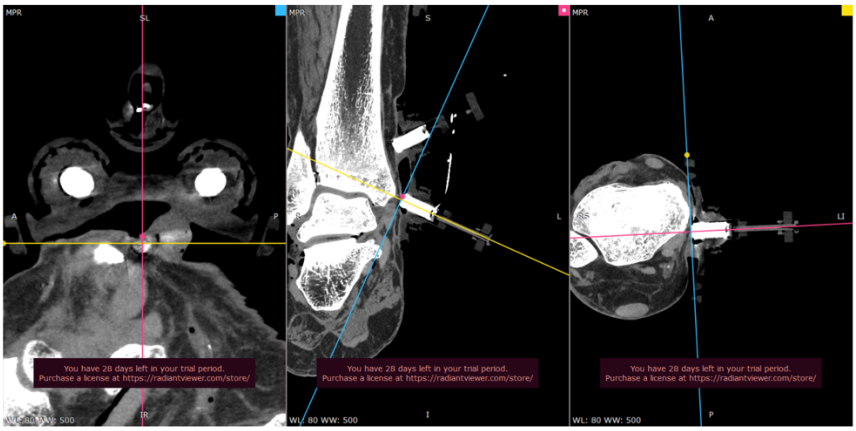
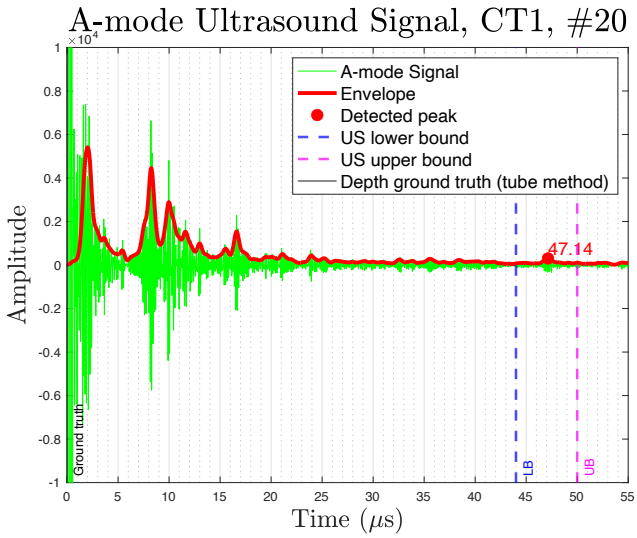
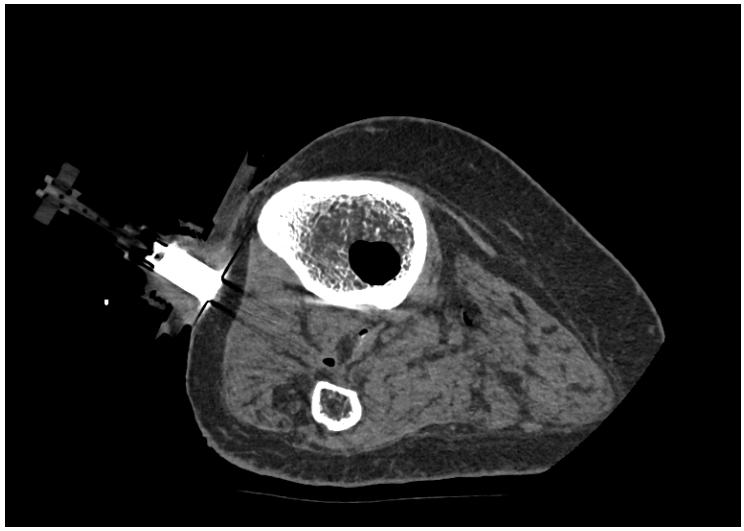
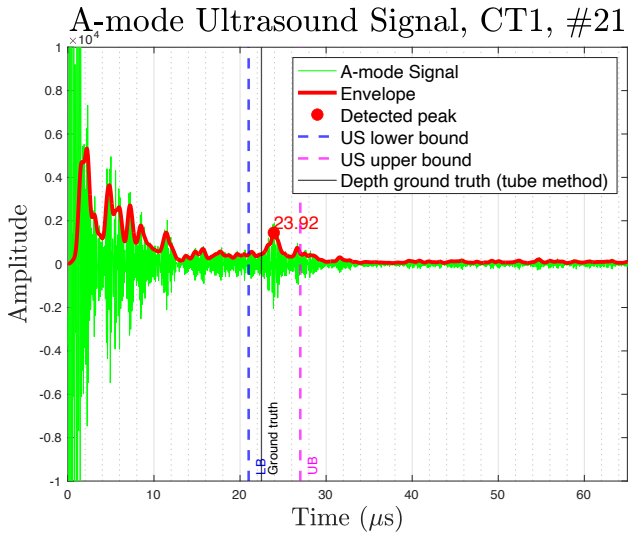
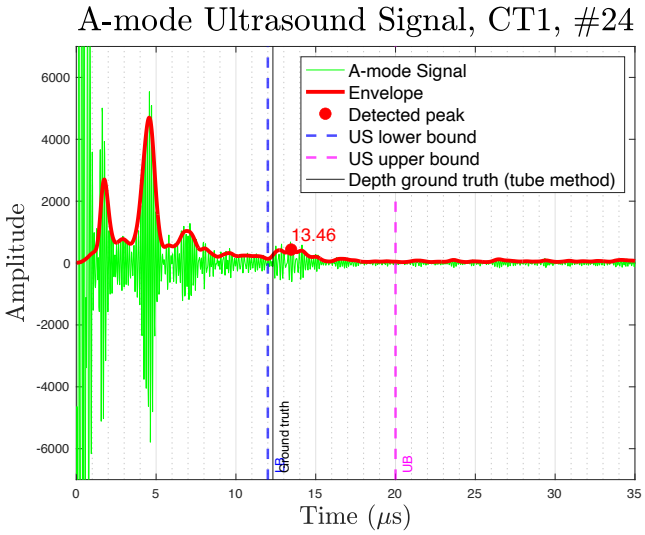


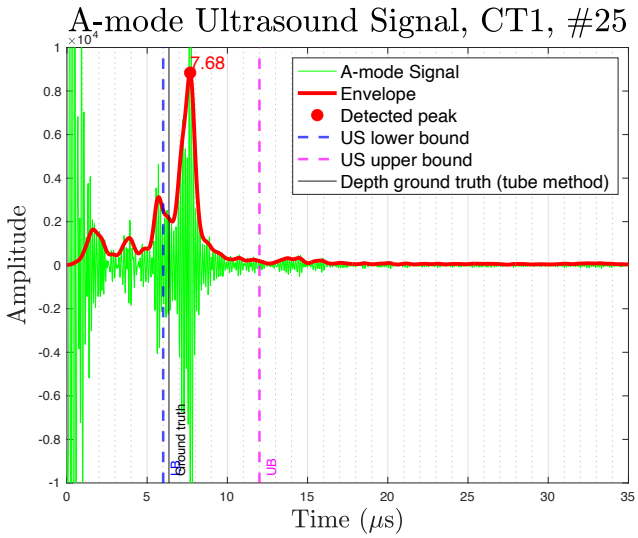

Figure C.1 Screen captures of all transducers, showing the three planes after Multiplanar Reconstruction (MPR) using RadiAnt software.

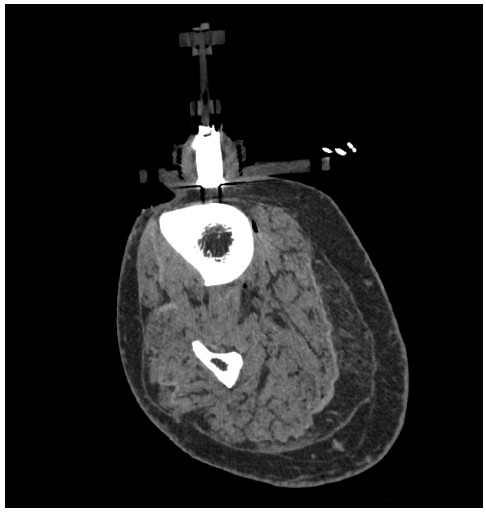

A-mode Ultrasound Signal, CT1, #20
A-mode Signal
Envelope
Detected peak
US lower bound
US upper bound
Depth ground truth (tube method)
47.14
Ground truth
LB
UB
Amplitude
Time (μs)

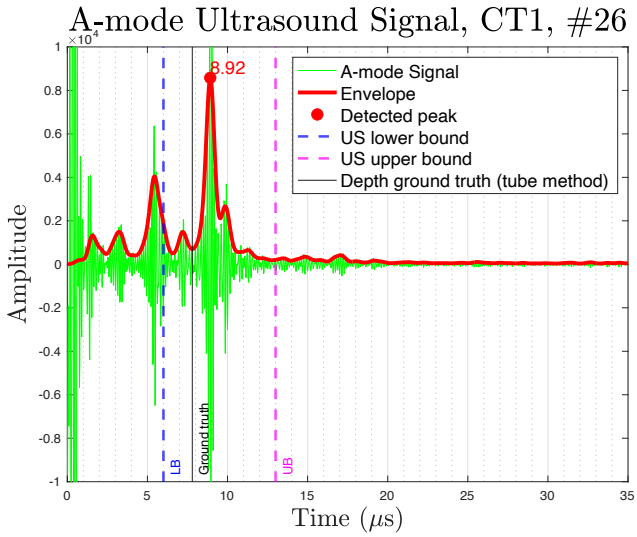


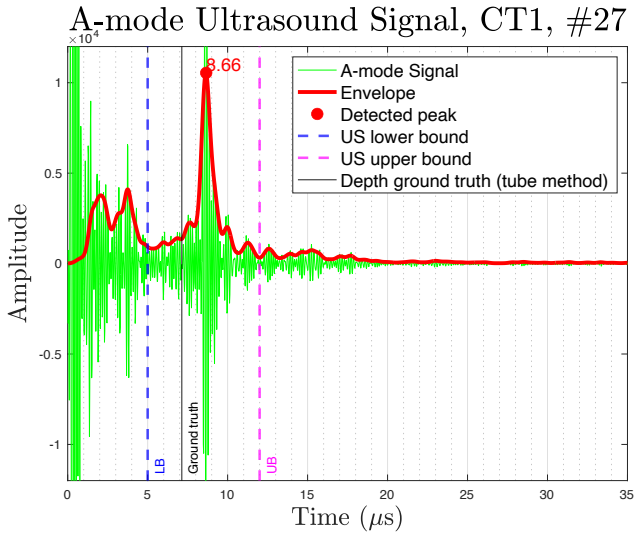

A-mode Ultrasound Signal, CT1, #21
A-mode Signal
Envelope
Detected peak
US lower bound
US upper bound
Depth ground truth (tube method)
23.92
LB
Ground truth
UB
Amplitude
Time (μs)

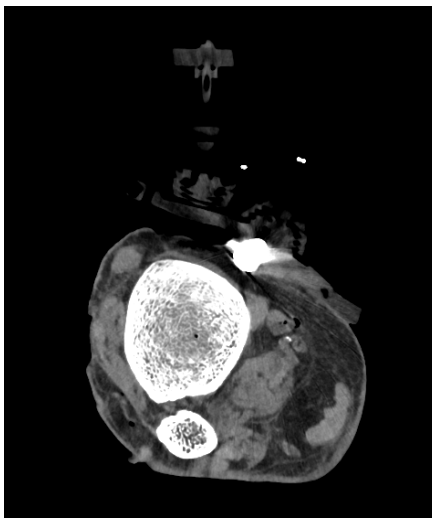

A-mode Ultrasound Signal, CT1, #24
A-mode Signal
Envelope
Detected peak
US lower bound
US upper bound
Depth ground truth (tube method)
13.46
LB
Ground truth
UB
Amplitude
Time (μs)

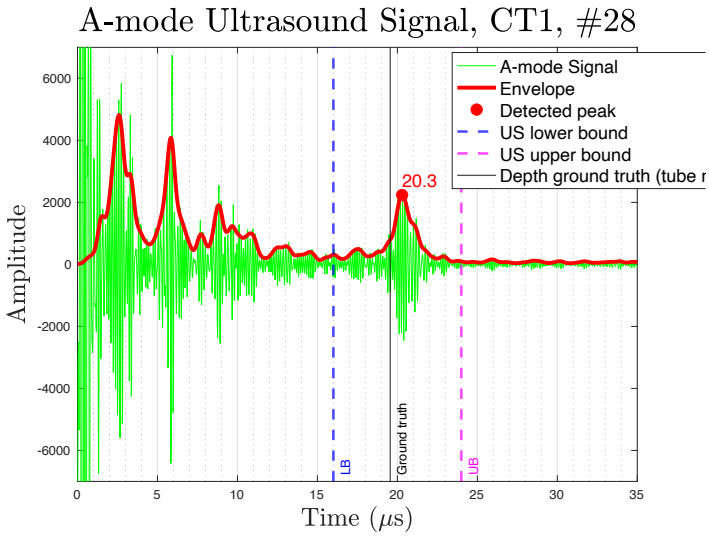

A-mode Ultrasound Signal, CT1, #25
7.68
A-mode Signal
Envelope
Detected peak
US lower bound
US upper bound
Depth ground truth (tube method)
Amplitude
Time ($\mu$s)
LB
Ground truth
UB

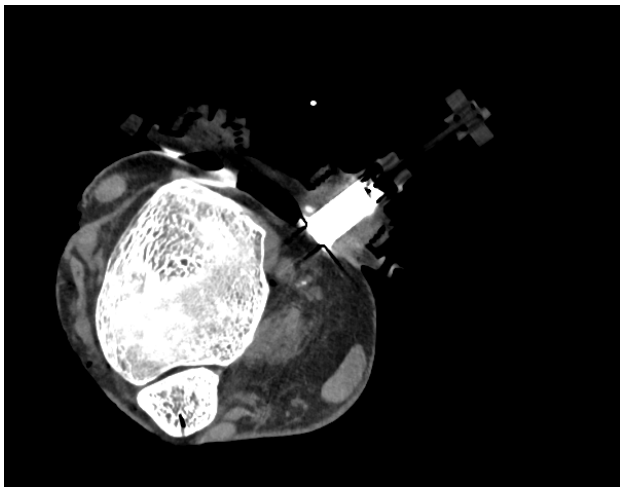


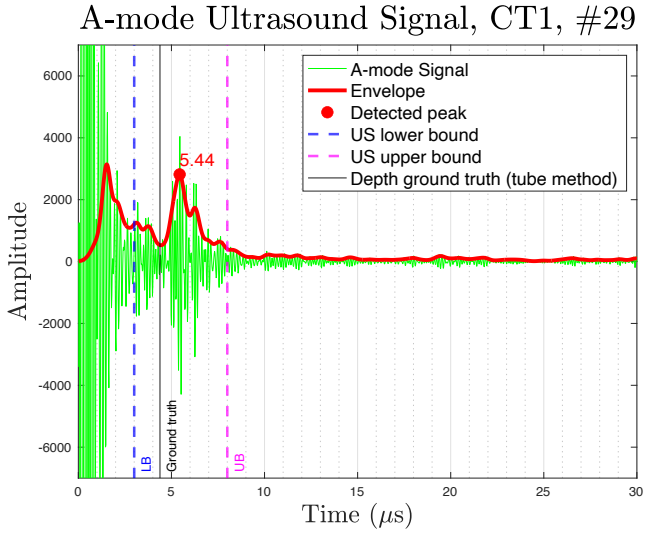

A-mode Ultrasound Signal, CT1, #26
8.92
A-mode Signal
Envelope
Detected peak
US lower bound
US upper bound
Depth ground truth (tube method)
Amplitude
Time ($\mu$s)
LB
Ground truth
UB

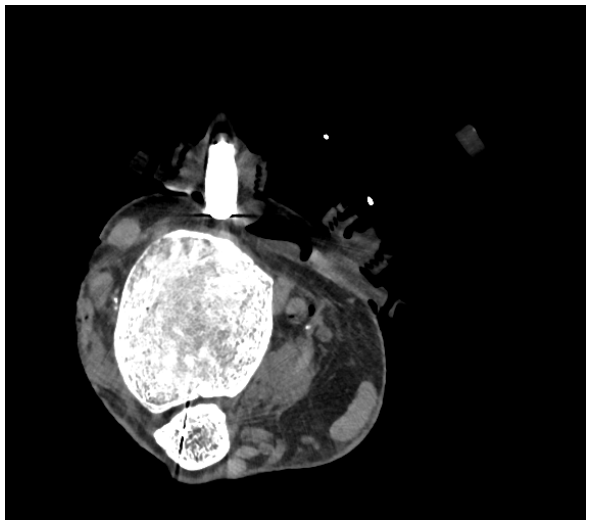

A-mode Ultrasound Signal, CT1, #27
8.66
A-mode Signal
Envelope
Detected peak
US lower bound
US upper bound
Depth ground truth (tube method)
Amplitude
Time ($\mu$s)
LB
Ground truth
UB

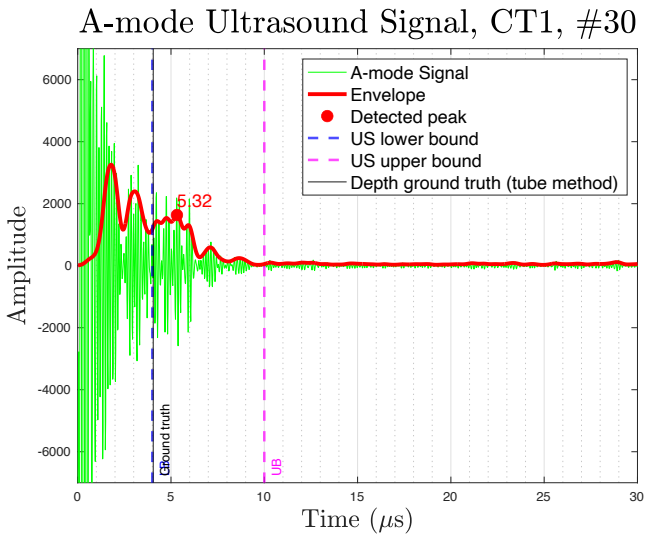


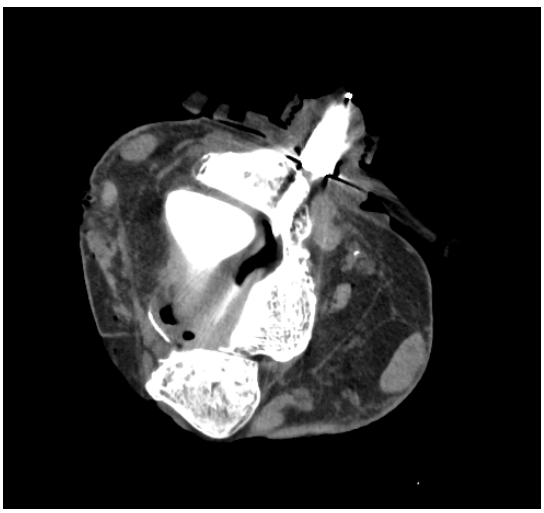

A-mode Ultrasound Signal, CT1, #28
A-mode Signal
Envelope
Detected peak
US lower bound
US upper bound
Depth ground truth (tube
20.3
LB
Ground truth
UB
Amplitude
Time ($\mu$s)

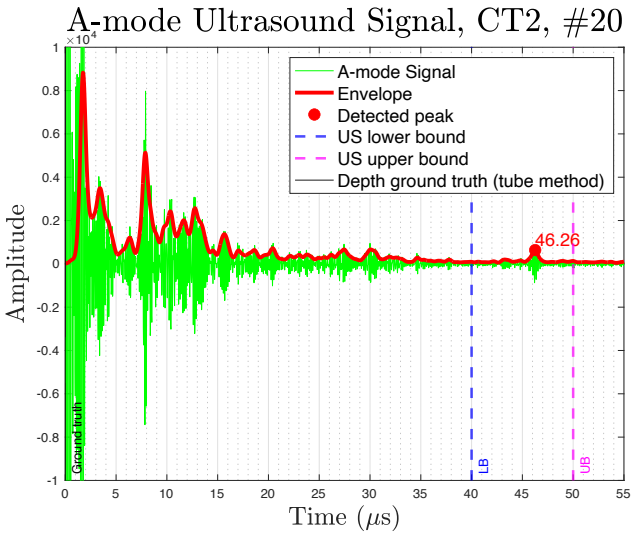


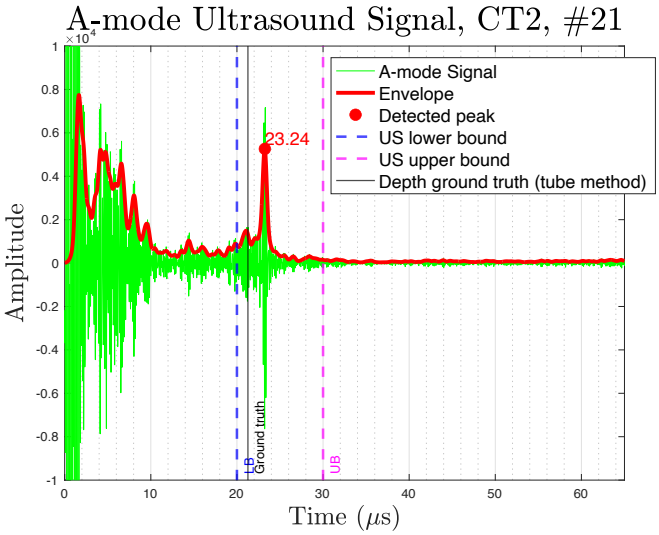

A-mode Ultrasound Signal, CT1, #29
A-mode Signal
Envelope
Detected peak
US lower bound
US upper bound
Depth ground truth (tube method)
5.44
LB
Ground truth
UB
Amplitude
Time ($\mu$s)

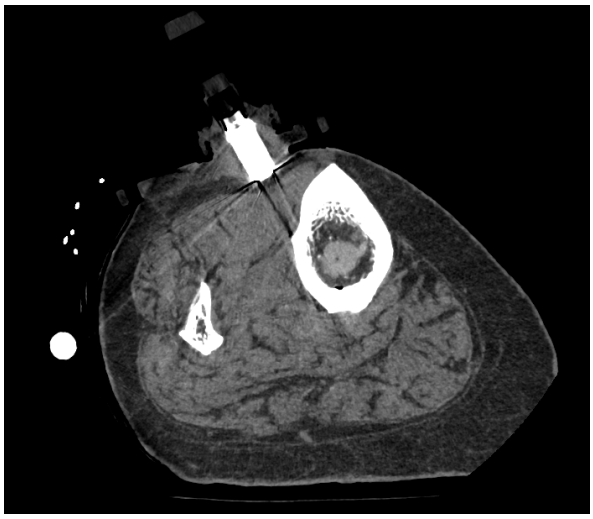


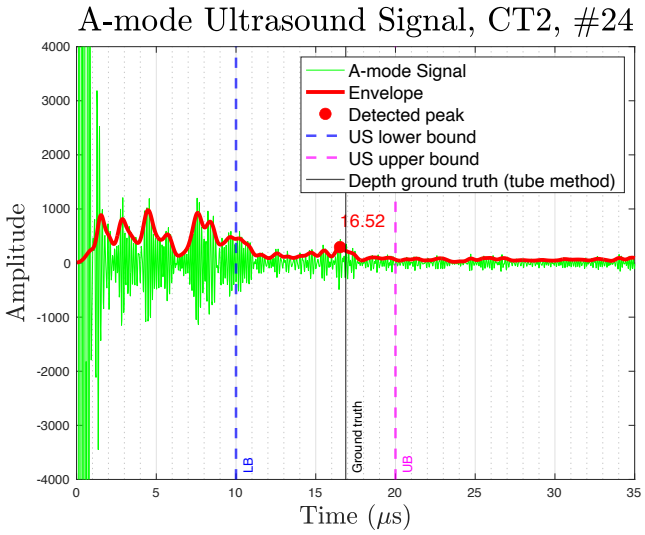

A-mode Ultrasound Signal, CT1, #30
A-mode Signal
Envelope
Detected peak
US lower bound
US upper bound
Depth ground truth (tube method)
5.32
Ground truth
UB
Amplitude
Time ($\mu$s)

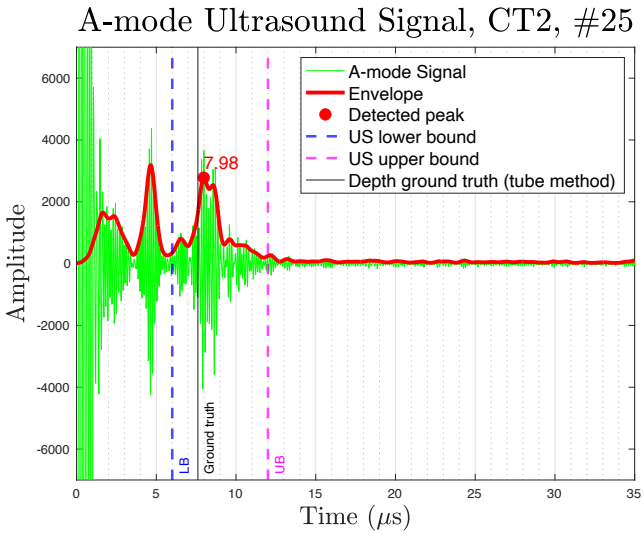


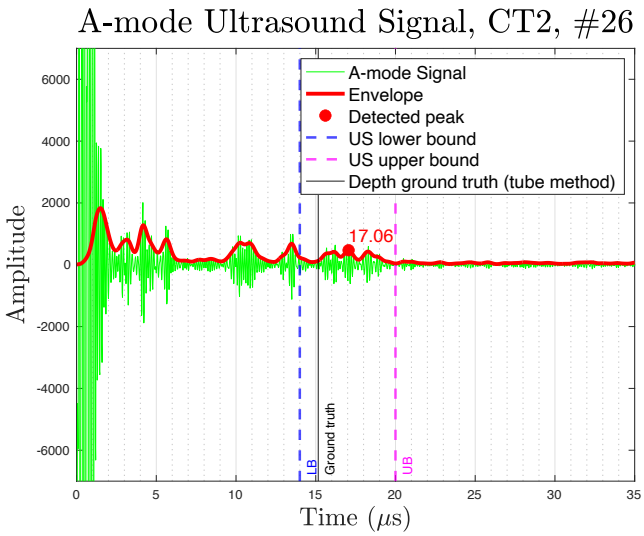

A-mode Ultrasound Signal, CT2, #20
A-mode Signal
Envelope
Detected peak
US lower bound
US upper bound
Depth ground truth (tube method)
46.26
Ground truth
LB
UB
Amplitude
Time (μs)

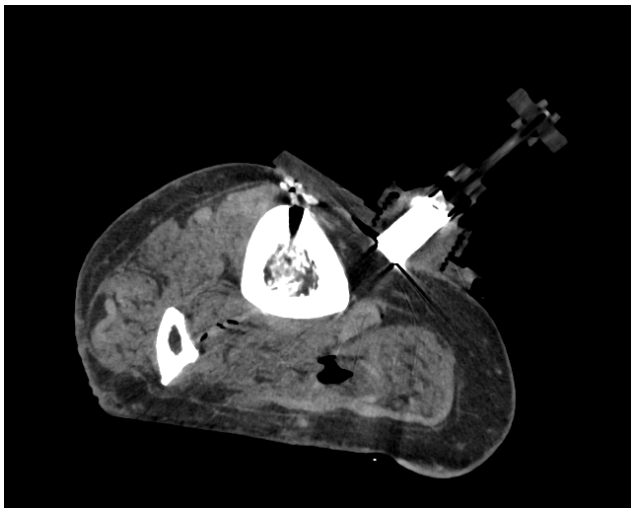

A-mode Ultrasound Signal, CT2, #21
A-mode Signal
Envelope
Detected peak
US lower bound
US upper bound
Depth ground truth (tube method)
23.24
LB
Ground truth
UB
Amplitude
Time (μs)

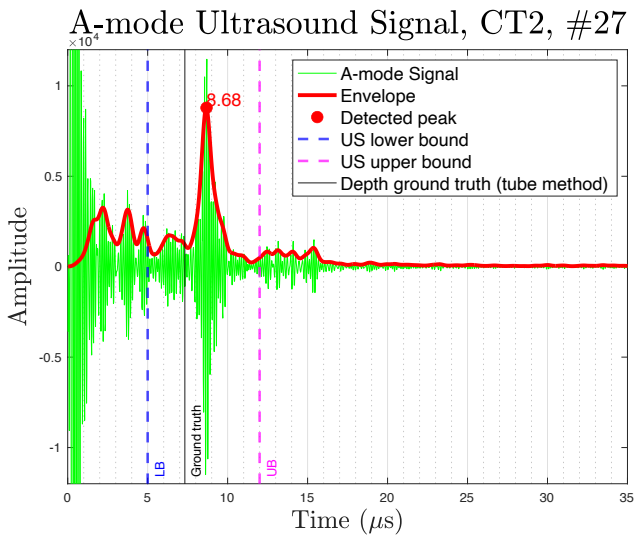


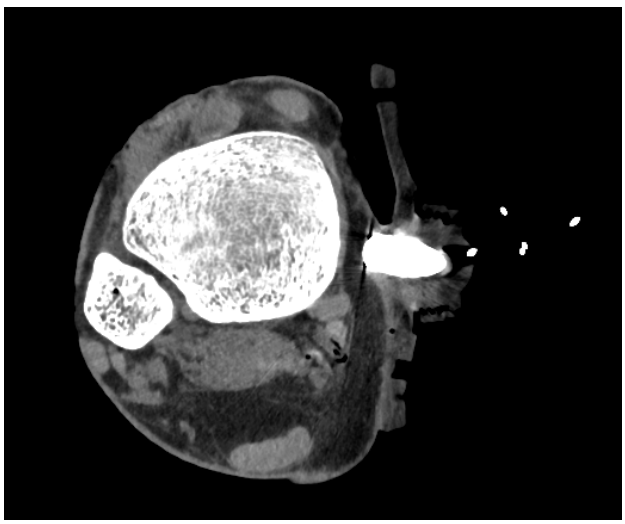

A-mode Ultrasound Signal, CT2, #24
A-mode Signal
Envelope
Detected peak
US lower bound
US upper bound
Depth ground truth (tube method)
16.52
LB
Ground truth
UB
Amplitude
Time (μs)

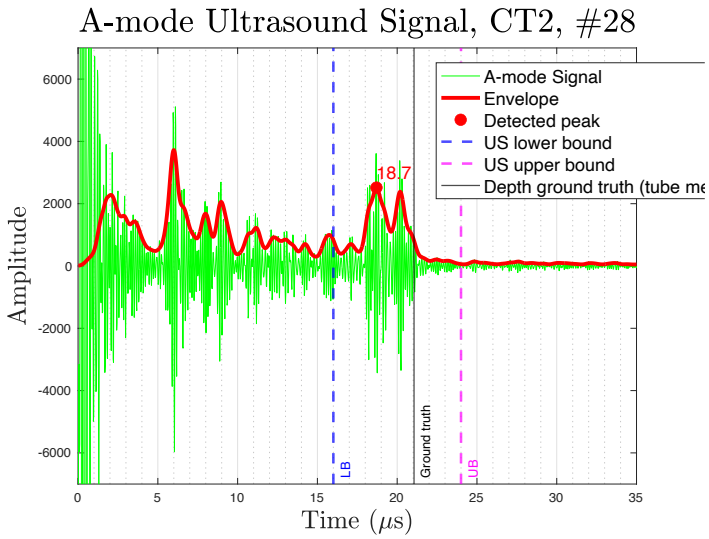

A-mode Ultrasound Signal, CT2, #25
A-mode Signal
Envelope
Detected peak
US lower bound
US upper bound
Depth ground truth (tube method)
7.98
LB
Ground truth
UB
Amplitude
Time ($\mu$s)

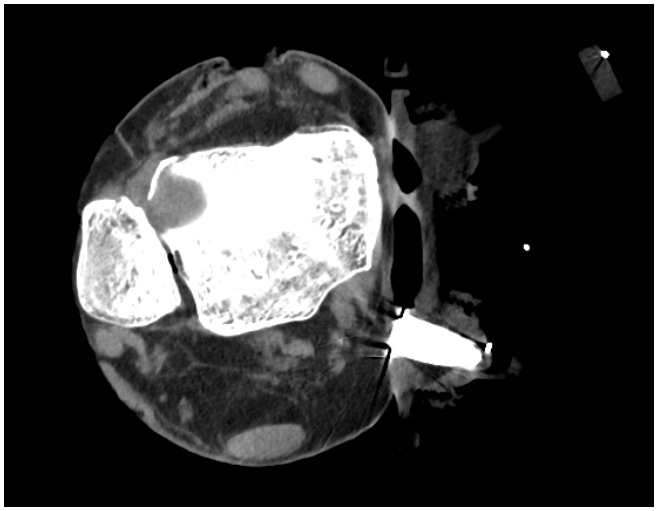

A-mode Ultrasound Signal, CT2, #26
A-mode Signal
Envelope
Detected peak
US lower bound
US upper bound
Depth ground truth (tube method)
17.06
LB
Ground truth
UB
Amplitude
Time ($\mu$s)

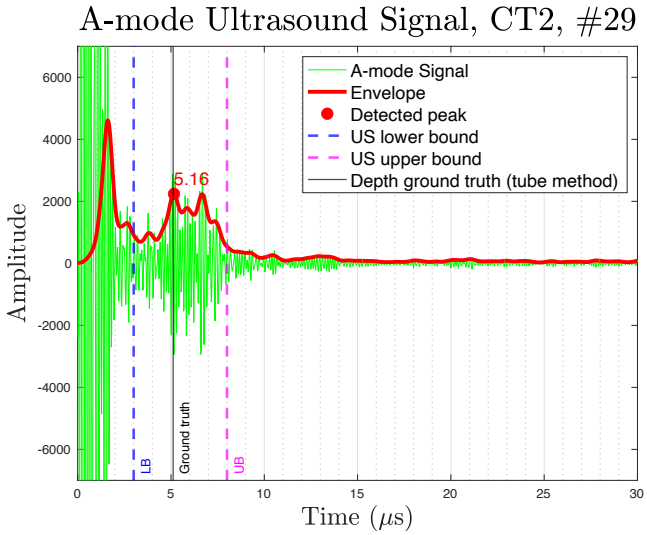


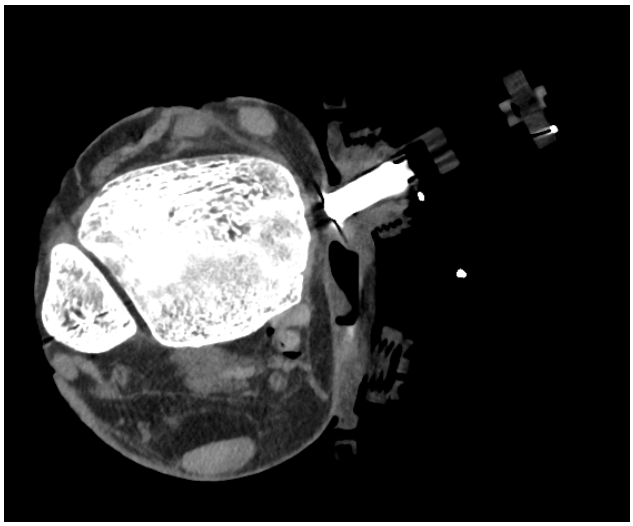

A-mode Ultrasound Signal, CT2, #27
A-mode Signal
Envelope
Detected peak
US lower bound
US upper bound
Depth ground truth (tube method)
8.68
LB
Ground truth
UB
Amplitude
Time ($\mu$s)

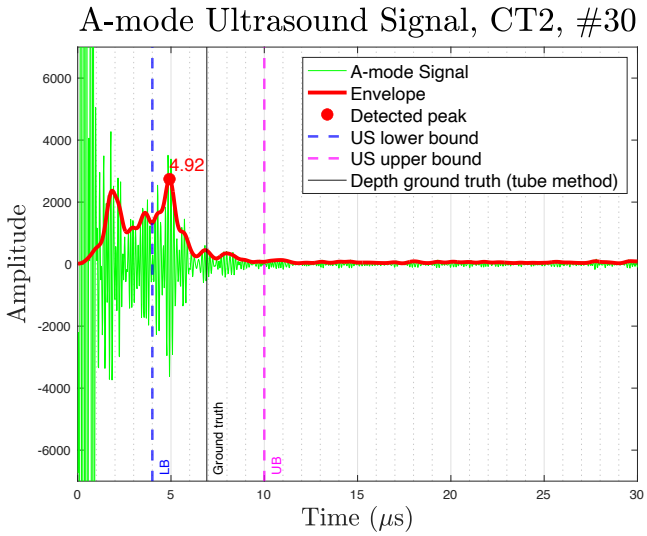








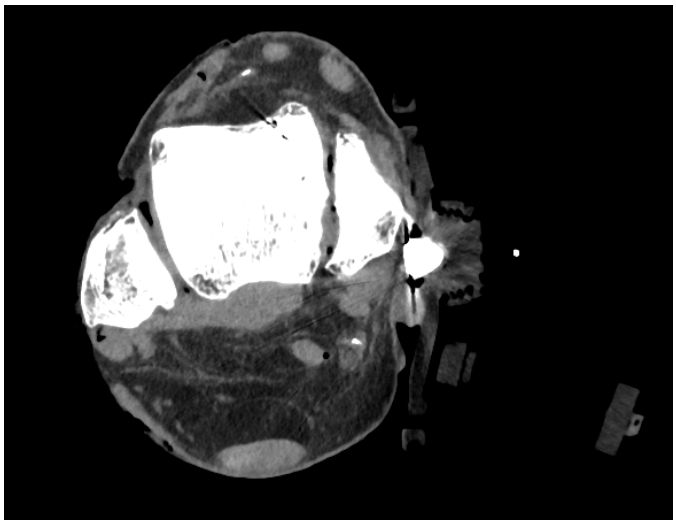
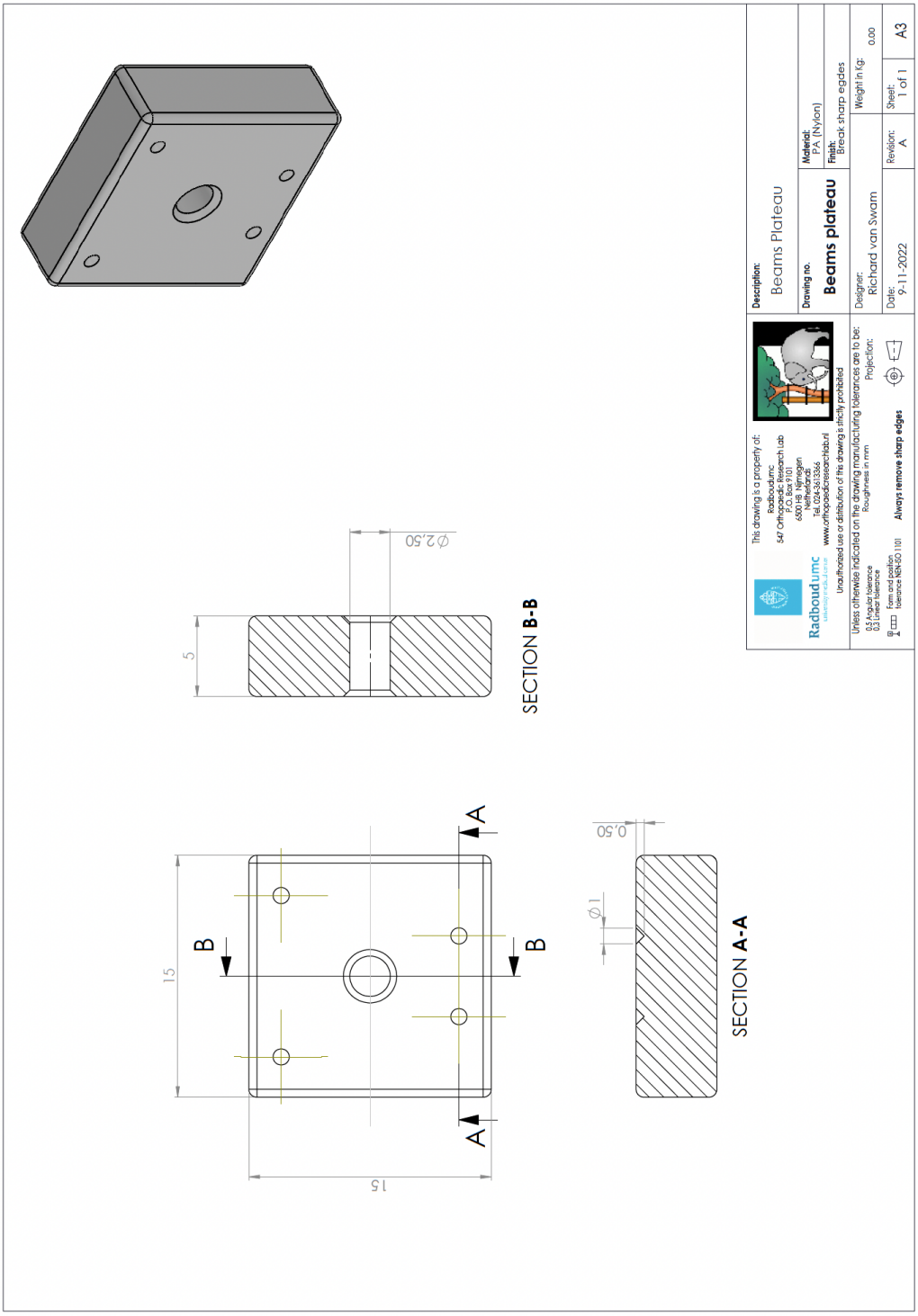



Figure C.2 The ultrasound raw signal, signal envelope, US lower and upper bounds, and CT ground truths for all transducers (left) and the corresponding axial CT slice (right).

# Appendix D Supplementary details of the setup

## D.1 Fiducial marker platform design



## D.2 Polyamide screw specifications

The length of all polyamide screws, screwheads and transducers was manually measured using a caliper (Mitutoyo Europe GmbH, Neuss, Germany) with an accuracy of 0.05 mm.

Table D.1 Polyamide screw and A-mode US transducer length specifications, manually measured using a caliper with accuracy 0.05 mm.

| Transducer # | Length bolt (mm) | Head (mm) | Body (mm) | Transducer Height (mm) | Total offset (mm) |
|---|---|---|---|---|---|
| **1** | 33.10 | 2.55 | 30.55 | 20.00 | 50.55 |
| **2** | 33.10 | 2.60 | 30.50 | 20.00 | 50.50 |
| **3** | 33.05 | 2.55 | 30.50 | 20.00 | 50.50 |
| **4** | 33.10 | 2.55 | 30.55 | 20.00 | 50.55 |
| **5** | 33.10 | 2.55 | 30.55 | 20.00 | 50.55 |
| **6** | 33.10 | 2.60 | 30.50 | 20.00 | 50.50 |
| **7** | 33.10 | 2.55 | 30.55 | 20.00 | 50.55 |
| **8** | 33.05 | 2.55 | 30.50 | 20.00 | 50.50 |
| **9** | 33.10 | 2.55 | 30.55 | 20.00 | 50.55 |

# Appendix E Ultrasound signal analysis

Introduction

In order to further investigate the characteristics of the US signal, the perpendicularity of each US transducer is compared to (1) the mean absolute difference between US estimated depth and CT ground truth depth, (2) the height of the amplitude peak in the US signal, and (3) whether there was a clear bone peak in the signal.

Methods

The perpendicularity of each US transducer with respect to the tibia surface is visually analyzed in all planes using the Multiplanar Reconstruction (MPR) results of the CT images (Appendix B and Figure C.1), and each transducer is subsequently categorized into three clusters (1: not perpendicular, 2: perpendicular in one plane, and 3: perpendicular in all planes). Also, for each transducer, the previously determined absolute differences (using speed of sound 1540 m/s) are collected, as well as the height of the US amplitude peaks. Furthermore, the peak in the US signal of each transducer is analyzed and scored by one observer (MD) as follows; 1: no clear peak, 2: questionable peak, and 3: clear high (bone) peak. Per cluster of perpendicularity, the mean ± SD values for the absolute differences, the US amplitude peak height, and scoring for clear bone peak are calculated. In the end, the means are compared between the categories in order to evaluate a potential trend that is related to the transducers' perpendicularity with respect to the bone surface.

Results

The US signal was visually analyzed to evaluate a potential relation between the US transducers' perpendicularity to the bone surface, and (1) the difference between US estimated depth and CT ground truth, (2) the height of the amplitude peak in the US signal, and (3) whether there was a clear bone peak in the signal. Of the 14 transducers, six were categorized as not perpendicular, six as perpendicular in one plane, and the remaining two as perpendicular in all planes (Table E.1). Mean absolute difference between US estimated depth and CT ground truth was similar for all three clusters of perpendicularity: 0.9174 ± 0.6241 mm versus 0.9150 ± 0.3965 m versus 0.9372 ± 0.0800 mm for non-perpendicular transducers, transducers perpendicular in one plane, and transducers perpendicular in all planes, respectively (Figure E.1).

Table E.1 Scoring of perpendicularity of each transducer with respect to the corresponding tibia surface (1 = not perpendicular, 2 = perpendicular in one plane, 3 = perpendicular in all planes). The absolute difference between ultrasound (US) estimated depth and CT ground truth depth, the height of US amplitude peak, and whether or not there was a clear (bone) peak in the signal (1 = no clear peak, 2 = questionable peak ,3 = clear high peak).

| Transducer | Perpendicular to tibia? | Absolute difference US est depth vs CT gt depth (mm) | Height of US amplitude peak | Clear peak? |
|---|---|---|---|---|
| 20, CT1 | N/A | N/A | 303.86 | N/A |
| 21, CT1 | 1 | 1.1028 | 1448.20 | 2 |
| 24, CT1 | 2 | 0.8807 | 4699.78 | 1 |
| 25, CT1 | 3 | 1.0172 | 8841.72 | 3 |
| 26, CT1 | 3 | 0.8572 | 8577.64 | 3 |
| 27, CT1 | 2 | 1.1758 | 10544.50 | 3 |

| | | | | |
|---|---|---|---|---|
| 28, CT1 | 2 | 0.5898 | 2236.52 | 3 |
| 29, CT1 | 1 | 0.8133 | 2817.44 | 3 |
| 30, CT1 | 1 | N/A | 1634.94 | N/A |
| 20, CT2 | N/A | N/A | 635.32 | N/A |
| 21, CT2 | 2 | 1.5120 | 5261.72 | 3 |
| 24, CT2 | 1 | 0.2772 | 294.54 | 1 |
| 25, CT2 | 2 | 0.2865 | 2781.24 | 2 |
| 26, CT2 | 1 | 1.4626 | 470.38 | 1 |
| 27, CT2 | 2 | 1.0450 | 8788.30 | 3 |
| 28, CT2 | 1 | 1.8144 | 2516.56 | 2 |
| 29, CT2 | 1 | 0.0343 | 2240.21 | 2 |
| 30, CT2 | 2 | N/A | 2743.31 | N/A |

The height of the US peak amplitude ranged from 294.54 to 10544.50, and mean US peak amplitude was highest for the transducers that were perpendicular in all planes (8709.68 ± 132.04), followed by the cluster of transducers that were perpendicular in one plane (5718.68 ± 3019.71), and the cluster of transducers that were not perpendicular (1631.22 ± 977.21), respectively (Figure E.2). For a total of seven transducers, a clear (bone) peak could be distinguished (score: 3), whereas four transducers had a questionable peak (score: 2) in the signal, and three transducers had no clear peak (score: 1). The mean score for clear peaks increases per cluster of perpendicularity, with mean ± SD of 1.83± 0.69 for non-perpendicular transducers, 2.50 ± 0.76 for transducers perpendicular in one plane, and 3.00 ± 0.00 for the transducers perpendicular in all planes (Figure E.3).

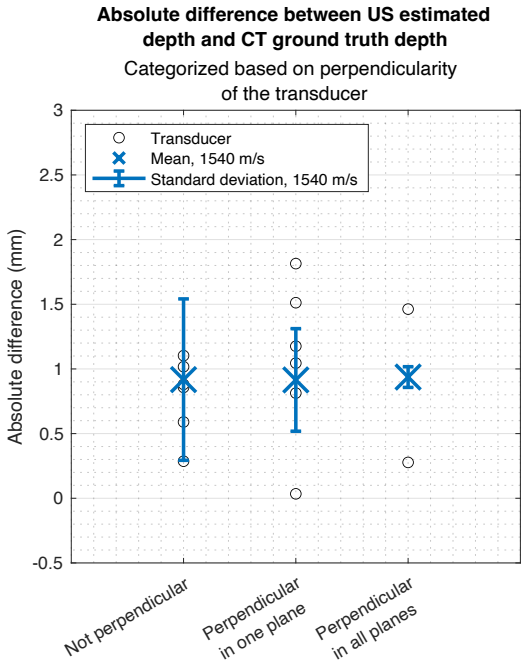



Figure E.1 Absolute differences between US estimated depth and CT ground truth depth for each transducer (black circles), categorized by the orientation of the transducer with respect to the tibia surface. The means and standard deviations are shown per category (blue crosses and error bars).

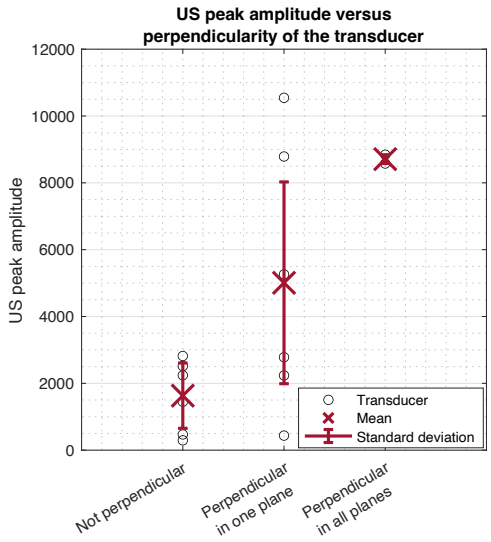



Figure E.2 US peak amplitude height for each transducer (black circles), categorized by the orientation of the transducer with respect to the tibia surface. The means and standard deviations are shown per category (burgundy crosses and error bars).

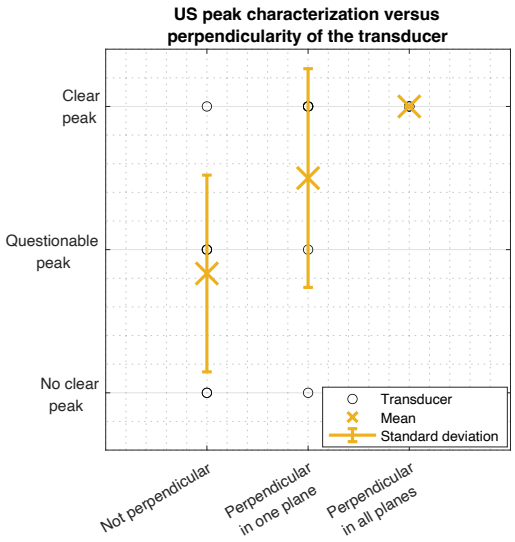



Figure E.3 US peak characterization for each transducer (black circles), categorized by the orientation of the transducer with respect to the tibia surface. The US signal per transducer was scored with 1 point in case of no clear peak, 2 points for a questionable peak, and 3 points for a clear, high peak. Means and standard deviations of these scores were calculated per category of perpendicularity (yellow crosses and error bars).

Discussion

The mean absolute difference did not show a clear correlation with the perpendicularity of the US transducer, whereas higher US peak amplitudes and clearer peaks in the US signal did show a potential positive relationship with increased perpendicularity of the transducer. This can be explained by the fact that a fully perpendicular transducer more likely received more US signal and therefore a higher and more clear US amplitude peak for the bone is shown. However, it must be noted that the sample size for this analysis was very small, with a total of 14 transducers. Furthermore, the amount of transducers per cluster is not divided equally, since only two of the 14 transducers could be classified as perpendicular in all planes, whereas the other clusters of perpendicularity (not perpendicular and perpendicular in one plane) both included six transducers. Besides, it would be better if an additional observer independently defined the level of perpendicularity for each transducer.