# Classical and Learned MR to Pseudo-CT Mappings for Accurate Transcranial Ultrasound Simulation

Maria Miscouridou<sup>®</sup>, José A. Pineda-Pardo, Charlotte J. Stagg, Bradley E. Treeby<sup>®</sup>, and Antonio Stanziola<sup>®</sup>

Abstract — Model-based treatment planning for transcranial ultrasound therapy typically involves mapping the acoustic properties of the skull from an X-ray computed tomography (CT) image of the head. Here, three methods for generating pseudo-CT (pCT) images from magnetic resonance (MR) images were compared as an alternative to CT. A convolutional neural network (U-Net) was trained on paired MR-CT images to generate pCT T images from either T1-weighted or zero-echo time (ZTE) MR images (denoted tCT and zCT, respectively). A direct mapping from ZTE to pCT was also implemented (denoted cCT). When comparing the pCT and ground-truth CT images for the test set, the mean absolute error was 133, 83, and 145 Hounsfield units (HU) across the whole head, and 398, 222, and 336 HU within the skull for the tCT, zCT, and cCT images, respectively. Ultrasound simulations were also performed using the generated pCT images and compared to simulations based on CT. An annular array transducer was used targeting the visual or motor cortex. The mean differences in the simulated focal pressure, focal position, and focal volume were 9.9%, 1.5 mm, and 15.1% for simulations based on the tCT images; 5.7%, 0.6 mm, and 5.7% for the zCT; and 6.7%, 0.9 mm, and 12.1% for the cCT. The improved results for images mapped from ZTE highlight the advantage of using imaging sequences, which improves the contrast of the skull bone. Overall, these results demonstrate that acoustic simulations based on MR images can give comparable accuracy to those based on CT.

Index Terms—Acoustic simulation, computed tomography (CT), convolutional neural network, deep learning,

Manuscript received 30 June 2022; accepted 6 August 2022. Date of publication 15 August 2022; date of current version 27 September 2022. This work was supported in part by the Engineering and Physical Sciences Research Council (EPSRC), U.K., under Grant EP/S026371/1 and in part by the UK Research and Innovation (UKRI) Centre for Doctoral Training in Al-enabled Healthcare Systems, under Grant EP/S021612/1. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) license to any Author Accepted Manuscript version arising. (*Corresponding author: Maria Miscouridou.*)

This work involved human subjects in its research. Approval of all ethical and experimental procedures and protocols was granted by the HM Hospitales Ethics Committee for Clinical Research.

Maria Miscouridou, Bradley E. Treeby, and Antonio Stanziola are with the Department of Medical Physics and Biomedical Engineering, University College London, London WC1E 6BT, U.K. (e-mail: maria.miscouridou.16@ucl.ac.uk).

José A. Pineda-Pardo is with HM CINAC, Fundación HM Hospitales de Madrid, Hospital HM Puerta del Sur. CEU-San Pablo University, 28938 Madrid, Spain.

Charlotte J. Stagg is with the Wellcome Centre for Integrative Neuroimaging, FMRIB, Nuffield Department of Clinical Neuroscience, University of Oxford, Oxford OX3 9DU, U.K.

Digital Object Identifier 10.1109/TUFFC.2022.3198522

MRI, pseudo-CT (pCT), transcranial ultrasound stimulation (TUS).

# I. INTRODUCTION

RANSCRANIAL ultrasound therapy is a class of non-I invasive techniques that leverage ultrasound energy to modify the structure or function of the brain, for example, to ablate brain tissue [1], modulate brain activity [2], or deliver therapeutic agents through the blood-brain barrier [3]. The major challenge is the delivery of ultrasound through the intact skull bone, which can significantly aberrate and attenuate the transmitted ultrasound waves. To counter the effect of the skull, computer simulations are often used to predict the intracranial pressure field [4] or to adjust phase delays to ensure a coherent focus [5]. Conventionally, these simulations are based on acoustic material property maps extracted from X-ray computed tomography (CT) images [6]. For clinical treatments in a hospital environment, obtaining pretreatment CT images does not necessarily pose significant challenges. However, for transcranial ultrasound stimulation (TUS), which is being widely explored as a neuroscientific tool in addition to its potential clinical applications [7], obtaining CT images for healthy volunteer studies can be more problematic. In the current work, the use of pseudo-CT (pCT) images for computer simulations of transcranial ultrasound propagation in a TUS setting is investigated. Three different methods of pCT generation are compared: 1) using a deep neural network based on T1-weighted (T1w) magnetic resonance (MR) images; 2) using a deep neural network based on zero-echo time (ZTE) MR images; and 3) directly mapping from ZTE MR images using classical image processing techniques following [8].

The image-to-image translation (I2IT) of MR to pCT images of the brain and skull has been widely explored in the imaging literature, particularly in the context of positron emission tomography (PET)-MR, where the pCT images are used for PET attenuation correction [9] and radiotherapy planning [10]. In many cases, deep learning has been shown to outperform classical techniques [11]. A variety of models and techniques have been used, including generative neural networks (GANs) [12], [13], supervised learning [14], contrastive learning [15], [16], and denoising diffusion probabilistic models [17]. Often, the network architecture accounts for features at multiple scales using a U-Net [11], [18]. The inputs

1525-8955 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information. to the neural network can be the full 3-D image volumes [19], 2-D slices along one or multiple planes [20], or small 3-D patches [21].

The performance of the trained network is heavily influenced by the quality of the training dataset and choice of loss function. Ideally, the loss function should be directly related to the final performance of interest: in our case, the acoustic properties of the generated fields from the pCT (see Section II-E). However, in the absence of an efficient differentiable acoustic simulator, the corresponding training regime is often inefficient and losses are therefore defined in image space. Simple voxel-based metrics, such as mean absolute error (MAE) and mean squared error (MSE), are often used when registered image pairs are available [22].

Unpaired images can also be used for successfully training I2IT models, often by relying on some form of cycleconsistency loss [23], [24], [25] implemented using a discriminator neural network [26]. While theoretically effective and often capable of training complex models, such as ones that disentangle geometrical features and imaging modality [27], training GANs against a discriminator can be challenging and often unstable [28].

Several previous works have investigated the use of learned pCTs for transcranial ultrasound simulation. In the context of high-intensity focused ultrasound (HIFU) ablation, Su *et al.* [29] used a 2-D U-Net trained on transverse slices from a dataset of 41 paired dual-echo ultrashort echo time (UTE) MR images and segmented CT images (only the segmented skull bone was used for training). Within the skull, the pCT images generated from the test set had an MAE of  $105 \pm 21$  HU and showed a good correlation with CT in terms of skull thickness and skull density ratio. Coupled acoustic-thermal simulations were performed for a 1024 multielement array and a deep brain target. Differences in the simulated acoustic field were not reported, but differences in the predicted peak temperature were less than 2 °C.

In a related study, Liu *et al.* [30] used a 3-D conditional generative adversarial network (cGAN) trained on patches from a dataset of 86 paired T1w and segmented CT skull images. Within the skull, the pCT images generated from the test set had an MAE of  $191 \pm 22$  HU and similarly showed a good correlation with CT in terms of skull thickness and skull density ratio. Acoustic simulations for a 1024 multielement array and deep brain target showed an average  $23 \pm 6.5\%$  difference in the simulated intracranial spatial-peak pressure when using pCT versus CT and  $0.35 \pm 0.40$  mm difference in the focal position.

In the context of TUS, Koh *et al.* [31] used a 3-D cGAN trained on 3-D patches from a dataset of 33 paired T1w and CT images. The generated pCT images had an MAE of  $86 \pm 9$  HU within the head and  $280 \pm 24$  HU within the skull. Acoustic simulations were performed using a singleelement bowl transducer driven at 200 kHz targeted at three brain regions (motor cortex, visual cortex, and dorsal anterior cingulate cortex). Across all targets for the test set, the mean difference in the simulated intracranial peak pressure was  $3.11 \pm 2.79\%$  and the mean difference in the focal position was  $0.86 \pm 0.52$  mm. However, aberrations to the ultrasound waves are known to be significantly reduced at low transmit frequencies [32], [33], and thus, the performance of learned pCTs at higher frequencies more commonly used for TUS remains an open question.

Several studies have also explored directly mapping the skull acoustic properties from T1w [34], UTE [35], [36], [37], or ZTE [38] images. Wintermark *et al.* [34] generated virtual CTs from three different MR sequences and used a Bayesian segmentation strategy using a skull mask (obtained by CT thresholding) as a prior. Linear regression was used to estimate skull thickness and density, with the mappings from T1w images performing best. Acoustic simulations using calculated phase correction values from virtual CT showed good agreement with those from real CT. Phantom HIFU experiments were also performed where the difference in maximum temperature was 1 °C when using MR-based correction compared to CT-based correction.

In a similar study, Miller *et al.* [35] investigated using UTE images instead of CT to apply aberration corrections for HIFU treatments. Three *ex vivo* skull phantoms were imaged by UTE and CT, and the UTE was segmented into a binary skull mask. Experimental transcranial sonications were performed on each skull using aberration corrections derived from MR, CT, and using no corrections. The measured temperature rises with aberration corrections were 45% higher than noncorrected sonications, while there was no significant difference between the results from MR and CT-calculated corrections. UTE has also been shown to produce images with bone contrast highly correlated with that in CT images [36].

Guo *et al.* [37] also investigated the use of UTE images using a series of linear mappings and thresholding operations to derive pCT images. A linear regression model comparing skull density from UTE and CT showed that they were highly correlated. Acoustic properties of the skull derived from the UTE and CT images had less than 5% error and were used to run acoustic and thermal simulations. The temperature rise was 1 °C higher in CT-based simulations compared to UTE simulations and the focal location was usually within 1 pixel (1.33 mm).

Finally, Caballero *et al.* [38] extracted the bone from ZTE and CT images with a series of thresholding and morphological operators and used the bone maps to extract skull measures, such as skull thickness and skull density ratio. It was demonstrated with linear regression that the skull measures derived from the two modalities highly correlate to each other and also correlate with treatment efficiency.

The recent studies outlined above clearly demonstrate the feasibility of using pCT images for treatment planning in transcranial ultrasound therapy. Here, the previous results are extended in two ways. First, the ability to generate pCT images from T1w and ZTE images is directly compared using a unique dataset with high-resolution multimodality imaging data. While T1w images are widely available, they suffer from poor skull contrast, particularly compared to specialized bone imaging sequences that enable visualization of the bony anatomy [8]. Second, acoustic simulations are performed in the context of TUS using a commonly used transducer geometry

and a 500 kHz transmit frequency for both motor [39], [40] and visual [41] targets.

# II. METHODS

# A. Multimodality Dataset

The dataset used for the study consisted of paired highresolution CT and MR images. Subjects had previously been scheduled for transcranial MR-guided focused ultrasound surgery (tcMRgFUS) thalamotomy. The study protocol was approved by the HM Hospitales Ethics Committee for Clinical Research and all participants provided written consent forms. Each subject had a CT and T1w (fast-spoiled gradient echo) MR image, and a subset of them also had a ZTE MR image, giving a total of 171 paired CT-T1w datasets and 90 paired CT-ZTE datasets. The CT images were reconstructed using a bone-edge enhancement filter (FC30) and had a slice thickness of 1 mm and an in-plane resolution of approximately 0.45 mm  $\times$  0.45 mm. The in-plane resolution varied slightly between subjects due to the selected field-of-view (the number of pixels was fixed at 512  $\times$  512).

The MR images were acquired using a 3T GE Discovery 750 with an isotropic voxel size of 1 mm. Image acquisition parameters are described in detail in [38].

The images were processed as follows. First, a bias-field correction was performed on the MR images using FAST [42] to reduce image intensity nonuniformities resulting from transmit RF inhomogeneities and receive coil sensitivities. The CT images were then registered to the corresponding MR images using FLIRT [43], [44], using affine transformation with 12 degrees of freedom and a mutual-information cost function. As part of the registration step, CT images were resampled to match the resolution (1-mm isotropic) and fieldof-view of the MR images. After registration, all volumes were padded to a cube with an edge length of 256 voxels. Following Han et al. [11], the MR image intensities were then processed using midway histogram equalization [45]. The reference histogram was computed using images from the training set only (see Section II-B) and was computed separately for the ZTE and T1w images. The images were also normalized using the mean and standard deviation of the corresponding training set.

To remove image artifacts (such as dental implants, equipment, and headphones) and the bones outside the neurocranium (which are not relevant to TUS), a binary mask was manually annotated on the MNI152\_T1\_1mm template brain [46], [47]. The Montreal Neurological Institute (MNI) mask excluded the nasal and oral cavities, as well as everything below the foramen magnum. The MNI mask was then mapped to subject space by registering the MR images to the MNI template using FLIRT. The registered MNI mask was combined with a head mask (obtained by thresholding and filling the MR) to give a subject-specific mask that was used during training as described in the following. An example is given in Fig. 1. Note that the generation of the subject-specific mask does not require a ground-truth CT, so it can be applied to any new subject data during inference.



Fig. 1. Input to the neural network is one or multiple adjacent transverse MR slices, acquired either using ZTE or T1w sequences. The output of the neural network is multiplied by a head mask and compared to the registered CT inside the mask.

## B. Network Architecture and Training

For mapping MR images to pCT, a five-level U-Net [48] was implemented using Pytorch [49] similar to the architecture reported by Han [11]. Each level of the encoder consisted of either two (levels 1 and 2) or three (levels 3 to 5) convolutional layers with a rectified linear unit (ReLU) activation function followed by a batch normalization layer. The convolutions used zero padding, a  $3 \times 3$  kernel, and a stride of 1. A maxpooling layer was used between each level of the encoder, and skip connections between the encoder and the decoder were used for the first four levels. The decoder was implemented as a mirrored version of the encoder with convolutional transpose (unpooling) layers [50] instead of max-pooling layers. Both pooling and unpooling layers used a  $2 \times 2$  kernel size and stride 2, followed by a dropout layer. The dropout probability  $p_D$  was treated as a tunable hyperparameter, which was optimized using the validation set. The best models for T1w and ZTE inputs were obtained with  $p_D = 0.1$  and  $p_D = 0$ , respectively.

The input to the network was an  $n \times 256 \times 256$  block of n consecutive 2-D transverse MR slices (see the top right and bottom left of Fig. 1 for an example of one slice). Using a small stack of images gives some 3-D structural information to the network, as suggested in the discussion section of [11]. The network output was always a single 2-D transverse slice corresponding to the middle input slice, and 3-D pCT volumes were reconstructed slice by slice. Input stack sizes of n = 1, 3, 5, 7, 9, 11, and 15 were used in preliminary testing. A stack size of 11, i.e., five additional image slices on either side of the primary input slice, gave the lowest MAE on the validation set. Notably, the use of multiple input slices significantly reduced the occurrence of skull discontinuities

2899

between slices in the 3-D pCT images, particularly for the T1w inputs (this is discussed further in Section III-A). A stack size of 11 was thus used in all subsequent training.

The input was mapped to 64 channels in the first convolution. The number of channels was doubled at each layer of the encoder and halved in each layer of the decoder, except for the deep-most layer. A final  $1 \times 1$  convolutional layer was implemented to map each 64-component feature vector from the previous layer to image voxels in Hounsfield units (HU).

The network was trained separately for T1w and ZTE inputs. The ZTE dataset consisted of 90 subjects, with 62 subjects used for the training set  $\Omega_T$ , 14 for the validation set  $\Omega_V$ , and 14 for the testing set  $\Omega_E$ . Subjects were randomly assigned to each set. The T1w dataset consisted of 171 subjects. The same 14 subjects were used for the test set, with 26 subjects used for the validation set, and 131 for the training set. Data augmentation was also performed using random affine transformations with bilinear interpolation, with rotations in the range  $\pm 10^{\circ}$ , translations between  $\pm 5\%$  of the image size, and shears parallel to the *x*-axis of  $\pm 2.5^{\circ}$ .

The network  $f_{\theta}$  was trained by minimizing the average  $\ell_1$  norm of the masked error over the training set

$$\mathcal{L}(\theta) = \frac{1}{|\Omega_T|} \sum_{i \in \Omega_T} \frac{1}{N_i} \|m_i (y_i - f_\theta(x_i))\|_1 \tag{1}$$

where  $x_i$  is the stack of one or more adjacent input MR image slices,  $y_i$  is the corresponding slice from the ground-truth CT,  $m_i$  is the corresponding mask (see Fig. 1), and  $N_i$  is the number of pixels in the *i*th image. The loss function was evaluated stochastically for each optimization step, by drawing a random mini-batch of size 32 from the training set. The Adam optimizer [51] was used along with reduction on plateau scheduling with patience = 5, factor = 0.2 and a learning rate reducing from  $10^{-4}$  to  $10^{-6}$ . The networks were trained for 300 epochs on an NVIDIA Tesla P40 GPU, with 24-GB RAM.

## C. Classical ZTE Mapping

For comparison with the learned pCT mappings, a direct conversion of the ZTE images to pCT was also implemented following [8] and [52]. In this case, the bias-field correction was applied to the ZTE images using the N4ITK method within 3-D Slicer (V4.11.20210226). Voxel intensities for each image were individually normalized based on the soft-tissue peak in the image histogram to give a soft-tissue intensity of 1. A skull mask was generated by thresholding the voxel intensities between 0.2 and 0.75 (bone/air and bone/soft tissue), taking the largest connected component, and then filling the mask using morphological operations. Within the skull mask, ZTE values were mapped directly to CT HU using the linear relationship CT = -2085 ZTE + 2329. This relationship was calculated by taking the first principal component of the density plot of ZTE values versus CT values within the skull. Outside the skull mask, air and soft tissue were assigned values of -1000 and 42 HU, respectively.

# D. Evaluation of the pCT Images

The three different methods for generating pCT images were evaluated by comparing the generated 3-D pCT volumes for the 14 subjects in the test set against the corresponding groundtruth CTs. Image intensities were compared using MAE and the root-mean-squared error (RMSE). Both metrics were evaluated across the whole head (comparing voxels within the subject-specific head mask) and in the skull only (using a skull mask derived by thresholding the ground-truth CTs and combining with the head mask to exclude bones outside the neurocranium). For convenience, the different generated pCTs are referred to as tCT (for the learned pCT mapped from a T1w image), zCT (for the learned pCT mapped from a ZTE image), and cCT (for the pCT directly converted from a ZTE image using classical image processing techniques as explained in Section II-C).

# E. Ultrasound Simulations Using k-Wave

Acoustic simulations were performed using the open-source k-Wave toolbox [53], [54]. CT and pCT image pairs were converted to medium property maps as follows. First, the images were resampled to the simulation resolution (0.5 mm) using linear interpolation and then cropped. The images were then segmented into skull, skin, brain, and background regions using intensity-based thresholding along with morphological operations. Within the soft tissue, reference values were assigned for the acoustic properties. Within the skull, the sound speed and density were mapped directly from the image values in HU. The density was calculated using the conversion curve from [55] (using the hounsfield2density function in k-Wave). The sound speed c within the skull was then calculated from the density values  $\rho$  using a linear relationship of  $c = 1.33\rho + 167$  [56]. A constant value of attenuation within the skull was used. This general approach to mapping the acoustic properties from CT has been widely used in the literature and generally compares well with experimental and clinical measurements [4], [57], [58]. Note that, other mappings from CT images to acoustic properties are also possible (e.g., [59]). However, as the same mapping is used for all image sets, this choice does not strongly influence the simulated results.

The simulations were based on the NeuroFUS CTX-500 four-element annular array transducer (Sonic Concepts, Bothell, WA, USA). The transducer was modeled using a staircase-free formulation [60] using nominal values for the radius of curvature (63.2 mm) and element aperture diameters (32.8, 46, 55.9, and 64 mm). Simulations were run at six points per wavelength (PPW) in water and 60 points per period (PPP), which was sufficient to reproduce the relevant benchmark results (PH1-BM7-SC1) reported in [61] with less than 0.2% difference in the maximum pressure and no difference in the focal position. The transducer was driven using a continuous sinusoidal driving signal at 500 kHz until a steady state was reached.

Simulations were performed for the 14 skulls that formed the test set. For each skull, four transducer positions were used targeting the occipital pole of the primary visual cortex and the hand knob of the primary motor cortex in both hemispheres (giving a total of 56 comparisons for each pCT image type). These positions are two common targets for TUS studies. The variation in the skull bone thickness for the V1 targets

TABLE I MAE AND RMSE FOR THE 3-D PCT IMAGES GENERATED USING THE 14 SUBJECTS IN THE TEST SET AGAINST THE GROUND-TRUTH CT IMAGES

mask		MAE (HU)	RMSE (HU)
head	tCT zCT cCT	$ \begin{array}{r} 133 \pm 46 \\ 83 \pm 26 \\ 145 \pm 35 \end{array} $	$\begin{array}{c} 288 \pm 83 \\ 186 \pm 55 \\ 350 \pm 62 \end{array}$
skull	tCT zCT cCT	$\begin{array}{c} 398 \pm 116 \\ \textbf{222} \pm 78 \\ 336 \pm 80 \end{array}$	$528 \pm 144$ $305 \pm 102$ $465 \pm 101$

is typically greater due to the internal and external occipital protuberance. The target positions were first identified on the MNI152\_T1\_1mm MR imaging template brain [46], [47]. Approximate positions of the targets on the individual skulls were then calculated by registering the MR images with the MNI template and using the transformation matrix to map the target positions back to the individual skulls. For comparison between the CT and pCT image pairs, all other simulation parameters were kept identical except the image used to derive the acoustic property maps.

The calculated acoustic pressure fields for each CT/pCT pair were compared using the focal metrics outlined in [61] using the code available from [62]. Briefly, the magnitude and position of the spatial peak pressure within the brain were compared, along with the -6 dB focal volume.

#### **III. RESULTS**

# A. pCT Images

Results for the MAE and RMSE for the generated 3-D pCT images against the ground-truth CT images are given in Table I. The zCT images generated using the learned mapping from ZTE have significantly lower error values across both metrics and both regions (complete head and skull only) compared to the tCT and cCT images. This is not surprising given: 1) the ZTE image visually contains significantly more information about the morphology of the skull bone compared to the T1w image and 2) the network has significantly more power to learn an adaptive mapping to pCT compared to the fixed mapping used for the cCT. In the skull, the cCT images slightly outperform the tCT images. This is possible because the additional bone information available to the cCT (but not the tCT) outweighs the additional predictive power available to the tCT (but not the cCT). Overall, the error values are similar to those discussed in Section I.

Fig. 2 shows the correlation between the CT and pCT values within the skull mask (as identified from the ground-truth CT). There is a good correlation observed for all pCT images, with the lowest spread for the zCT matching the results given in Table I. The good fit observed for the cCT values demonstrates that a linear mapping from ZTE is a reasonable choice. The robustness of these correlations to variations in the specific MR sequence parameters (along with processing parameters such as the debiasing settings) still needs to be explored further.

Examples of the generated pCT images and corresponding error maps against CT for one subject from the test set



Fig. 2. Density plot showing the correlation between the pCT and groundtruth CT images for the test set. The white lines show y = x.

are shown in Fig. 3. The central sagittal slice through all 14 subjects along with individual MAE values is shown in Fig. 4. Overall, there is a good quantitative agreement between the pCT and ground-truth CT images. The biggest differences occur at the brain-skull and skull-skin boundaries, consistent with previous studies [11]. This is primarily due to the imperfect registration between the MR and CT images. There are two contributing factors. First, the skin-air interfaces are usually in slightly different places physically, e.g., due to the mobility of the skin and differences in the subject positioning during the MR and CT scans. Second, there are differences in the rigid brain-skull boundaries due to geometric distortions in the images that are not corrected by the affine transformation used in the registration step. This is evident in the difference plots for the cCT images, which uses a simple thresholding of the ZTE image to obtain the skull boundaries. The same misalignments are also apparent in the training/validation images. Improving the registration step, for example, by learning the registration as part of training the network [63] or by iterating between training and reregistration of the generated pCT images, will likely improve the predictive capabilities of the learned mappings and will be the subject of future work.

As discussed in Section II-B, the input to the network was 11 consecutive 2-D slices in the transverse direction. For the tCT images, this significantly reduced the occurrence of discontinuities in the skull between slices in the out-of-plane (e.g., sagittal) direction. The discontinuities occur because of the difficulty in consistently identifying the outer skull boundary in the 2-D T1w image slices, which is improved by providing the network with local 3-D information. However, even with 11 slices, visible discontinuities were still observed for a small number of subjects in the test set (tCT 4, 6, and



Fig. 3. Example of pCT and error maps for a single subject in the test set. The pCTs are registered to the ground-truth CT before comparison.

13 in Fig. 4). The equivalent network trained with a single input slice had similar MAE and RMSE values to those given in Table I, while the occurrence of local discontinuities in the skull was much worse. Thus, this feature does not seem to be captured by the error metrics used during training. This motivates using 3-D networks in the future, following related work [19], which has shown that training on the full 3-D skull volumes can produce better results, albeit at a higher computational cost. The same discontinuities were not observed for the zCT images, even when using a single slice as input to the network.

Although not formally investigated in this study, it is interesting to note that many of the subjects had small calcifications within the deep structures of the brain visible on the CT images (e.g., due to calcification of the choroid plexus). Some, but not all, of these were visible on the zCT images in the test set, and none were visible on the tCT images. Further work is needed to quantify the ability of the learned mappings to correctly reconstruct calcifications.

# B. Acoustic Simulations

Results for the acoustic simulations are summarized in Table II and Fig. 5. Subjects 4 and 6 from the tCT dataset were excluded from the simulation evaluation as they displayed discontinuities on the top of the skull, which made them unsuitable for acoustic simulations. Similarly, cCT from subject 10 was excluded as the predicted skull was not continuous due to the thresholding value not being suitable for this subject. Across all pCT images and target locations, the mean differences in the simulated focal pressure, focal position, and focal volume were  $7.3 \pm 5.7\%$ ,  $0.99 \pm 0.79$  mm, and  $11 \pm 12\%$  (to two significant figures). The best results were obtained when using the zCT images for motor targets, where the equivalent differences were 3.7%, 0.5 mm, and 3.9%. These values compare well with the differences observed in experimental repeatability [64] and numerical intercomparison [61] studies. For reference, the focal volume simulated in water is 3.9 mm wide and 24 mm long.

Between the different pCT images, the simulations based on the zCT images generally had the lowest average errors,

Target	СТ	focal pressure (%)		focal position (mm)		focal volume (%)	
		mean	std	mean	std	mean	std
V1	tCT	13.1	9.0	1.6	1.1	17.2	15.8
	zCT	7.6	4.9	<b>0.7</b>	0.4	<b>7.4</b>	7.1
	cCT	<b>5.7</b>	4.0	1.1	0.6	16.8	17.6
M1	tCT	6.7	3.5	1.4	0.8	13.0	13.3
	zCT	<b>3.7</b>	2.7	<b>0.5</b>	0.4	<b>3.9</b>	2.8
	cCT	7.8	3.8	0.7	0.5	7.4	5.7
All	tCT	9.9	7.5	1.5	1.0	15.1	14.6
	zCT	<b>5.7</b>	4.4	<b>0.6</b>	0.4	<b>5.7</b>	5.6
	cCT	6.7	4.0	0.9	0.6	12.1	13.8

TABLE II

ACOUSTIC METRICS ROUNDED TO ONE DECIMAL PLACE

as well as the smallest variation (see the interquartile range shown using the blue boxes in Fig. 5). The simulations based on both the zCT and cCT images consistently outperformed the tCT images, demonstrating that skull-specific imaging can improve the accuracy of the predictions. Interestingly, the simulations based on the learned mappings (tCT and zCT) generally overestimated the focal pressure, while simulations based on the direct mapping (cCT) generally underestimated the focal pressure. This may be related to the sharpness of the skull boundaries in the cCT images compared to the learned images resulting in a stronger reflection coefficient.

Between the two targets, the errors for the motor cortex were lower than the visual cortex, except for the focal pressure metric for the cCT images. This is expected, as the shape of the skull is generally more variable close to the visual cortex due to the internal and external occipital protuberance, which can result in stronger aberrations to the acoustic field. Examples from one test subject of the generated sound speed maps and acoustic field distributions are given in Fig. 6.

### IV. SUMMARY AND DISCUSSION

Three different approaches for generating pCT images from MR images were investigated in the context of treatment planning simulations for TUS. A convolutional neural network



Fig. 4. Middle sagittal slice for the learned pCT images for the 14 test subjects. The MAE value reported is calculated for the volume inside the head mask. A small number of the learned images mapped from T1w MR images have discontinuous skull boundaries.



Fig. 5. Differences in the focal pressure, focal position, and focal volume for acoustic simulations using different pCT images against simulations using a ground-truth CT. Results are divided into two sets, targeting the visual cortex (V1) and motor cortex (M1).

(U-Net) was trained to generate pCT images using paired MR-CT data with either T1w or ZTE MR images as input.

A direct mapping from ZTE to pCT was also implemented based on [8]. For the image-based metrics, the learned mapping from ZTE gave the lowest errors, with MAE values of 83 and 222 HU in the whole head and skull-only, respectively. The significant improvement compared to the mapping from T1w images demonstrates the advantage of using skull-specific MR imaging sequences.

Acoustic simulations were performed using the generated pCT images for an annular array transducer geometry operating at 500 kHz. The transducer was targeting either the left or right motor cortices or the left or right visual cortices. The choice of transducer geometry and brain targets was motivated by devices and targets used in the rapidly growing literature on human TUS.

Simulations using the pCT images based on ZTE showed close agreement with ground-truth simulations based on CT, with mean errors in the focal pressure and focal volume less than 7% and 9%, respectively, and mean errors in the focal position less than 0.8 mm (see Table II). Errors in simulations using the learned pCT images mapped from T1w images were higher but still may be acceptable depending on the accuracy required. These results demonstrate that acoustic simulations based on mapping pCT images from MR can give comparable results to simulations based on ground-truth CT.

For context, in the motor cortex, the smallest relevant target area for stimulation could reasonably be considered the cortical representation of a single hand muscle. A recent transcranial magnetic stimulation (TMS) study suggests that



Fig. 6. Sound speed maps and simulated acoustic field for a subject from the test set for targets in (a) visual cortex (V1) and (b) motor cortex (M1). The difference plots show the difference against the ground-truth generated using the real CT images. The V1 and M1 results show sagittal and coronal slices through the spatial peak pressure, respectively.

the functional area of the first dorsal interosseous (FDI), a muscle commonly targeted in motor studies, is approximately 26 mm<sup>2</sup> [65]. In the primary visual cortex, the mean distance between approximate centers of TMS-distinct visual regions (V1 and V2d) is 11 mm [66], suggesting a similar necessary resolution to M1. The errors in the focal position using the pCT images are significantly less, which should allow for accurate targeting of the intended structures.

Interestingly, there was not a strong correlation between the pixel-based image errors (MAE and RMSE) and the errors in the predicted acoustic field. This highlights the importance of running acoustic simulations when developing I2IT methods for TUS and potentially calls for the use of acoustic-based loss functions. The use of MR images for treatment planning is particularly important for neuroscience studies in healthy populations, where obtaining ethical approval to acquire CT images is usually problematic. Even in a clinical setting, there may be benefits to using MR-based acquisitions when the therapy is performed under MR guidance (e.g., multimodal image registration errors are reduced [67]). These results may also be of interest in other areas where pCTs are used, for example, PET-MR and radiotherapy planning.

Note that, in a clinical setting, the regulatory approval of learned mappings remains an important open question. However, we demonstrate that the learned pCTs from T1w and ZTE can be used for acoustic simulations that have comparable accuracy to those based on CTs.

#### REFERENCES

- W. J. Elias *et al.*, "A randomized trial of focused ultrasound thalamotomy for essential tremor," *New England J. Med.*, vol. 375, no. 8, pp. 730–739, 2016.
- [2] W. Legon *et al.*, "Transcranial focused ultrasound modulates the activity of primary somatosensory cortex in humans," *Nat. Neurosci.*, vol. 17, pp. 322–329, Feb. 2014.
- [3] C. Gasca-Salas *et al.*, "Blood-brain barrier opening with focused ultrasound in Parkinson's disease dementia," *Nature Commun.*, vol. 12, no. 1, pp. 1–7, 2021.
- [4] T. Deffieux and E. E. Konofagou, "Numerical study of a simple transcranial focused ultrasound system applied to blood-brain barrier opening," *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 57, no. 12, pp. 2637–2653, Dec. 2010.
- [5] J. Sun and K. Hynynen, "Focusing of therapeutic ultrasound through a human skull: A numerical study," J. Acoust. Soc. Amer., vol. 104, no. 3, pp. 1705–1715, Sep. 1998.
- [6] J.-F. Aubry, M. Tanter, M. Pernot, J.-L. Thomas, and M. Fink, "Experimental demonstration of noninvasive transskull adaptive focusing based on prior computed tomography scans," *J. Acoust. Soc. Amer.*, vol. 113, no. 1, pp. 84–93, Jan. 2003.
- [7] G. Darmani *et al.*, "Non-invasive transcranial ultrasound stimulation for neuromodulation," *Clin. Neurophysiol.*, vol. 135, pp. 51–73, 2022.
- [8] F. Wiesinger *et al.*, "Zero TE-based pseudo-CT image conversion in the head and its application in PET/MR attenuation correction and MRguided radiation therapy planning," *Magn. Reson. Med.*, vol. 80, no. 4, pp. 1440–1451, Oct. 2018.
- [9] N. Burgos *et al.*, "Attenuation correction synthesis for hybrid PET-MR scanners: Application to brain studies," *IEEE Trans. Med. Imag.*, vol. 33, no. 12, pp. 2332–2341, Dec. 2014.
- [10] M. Boulanger *et al.*, "Deep learning methods to generate synthetic CT from MRI in radiotherapy: A literature review," *Phys. Medica*, vol. 89, pp. 265–281, Sep. 2021.
- [11] X. Han, "MR-based synthetic CT generation using a deep convolutional neural network method," *Med. Phys.*, vol. 44, no. 4, pp. 1408–1419, 2017.
- [12] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2223–2232.
- [13] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1125–1134.
- [14] P. Zhang, B. Zhang, D. Chen, L. Yuan, and F. Wen, "Cross-domain correspondence learning for exemplar-based image translation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.* (CVPR), Jun. 2020, pp. 5143–5153.
- [15] T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu, "Contrastive learning for unpaired image-to-image translation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 319–345.
- [16] W. Jiangtao, W. Xinhong, J. Xiao, Y. Bing, Z. Lei, and Y. Yidong, "MRI to CT synthesis using contrastive learning," in *Proc. IEEE Int. Conf. Med. Imag. Phys. Eng. (ICMIPE)*, Nov. 2021, pp. 1–5.
- [17] H. Sasaki, C. G. Willcocks, and T. P. Breckon, "UNIT-DDPM: UNpaired image translation with denoising diffusion probabilistic models," 2021, arXiv:2104.05358.
- [18] L. Presotto *et al.*, "Evaluation of a 2D UNet-based attenuation correction methodology for PET/MR brain studies," *J. Digit. Imag.*, vol. 35, no. 3, pp. 432–445, 2022.
- [19] J. Fu *et al.*, "Deep learning approaches using 2D and 3D convolutional neural networks for generating male pelvic synthetic computed tomography from magnetic resonance imaging," *Med. Phys.*, vol. 46, no. 9, pp. 3788–3798, Sep. 2019, doi: 10.1002/mp.13672.
- [20] M. F. Spadea *et al.*, "Deep convolution neural network (DCNN) multiplane approach to synthetic CT generation from MR images— Application in brain proton therapy," *Int. J. Radiat. Oncol.\*Biol.\*Phys.*, vol. 105, no. 3, pp. 495–503, Nov. 2019.
- [21] S. Roy, J. A. Butman, and D. L. Pham, "Synthesizing CT from ultrashort echo-time MR images via convolutional neural networks," in *Proc. Int. Workshop Simulation Synth. Med. Imag.* Cham, Switzerland: Springer, 2017, pp. 24–32.

- [22] P. M. Martinez-Girones *et al.*, "Franken-CT: Head and neck MRbased pseudo-CT synthesis using diverse anatomical overlapping MR-CT scans," *Appl. Sci.*, vol. 11, no. 8, p. 3508, Apr. 2021.
- [23] A. Jabbarpour, S. R. Mahdavi, A. V. Sadr, G. Esmaili, I. Shiri, and H. Zaidi, "Unsupervised pseudo CT generation using heterogenous multicentric CT/MR images and CycleGAN: Dosimetric assessment for 3D conformal radiotherapy," *Comput. Biol. Med.*, vol. 143, Apr. 2022, Art. no. 105277.
- [24] Y. Liu et al., "CT synthesis from MRI using multi-cycle GAN for headand-neck radiation therapy," Computerized Med. Imag. Graph., vol. 91, Jul. 2021, Art. no. 101953.
- [25] H. Arabi, G. Zeng, G. Zheng, and H. Zaidi, "Novel adversarial semantic structure deep learning for MRI-guided attenuation correction in brain PET/MRI," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 46, no. 13, pp. 2746–2759, Dec. 2019.
- [26] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Process.*, vol. 35, no. 1, pp. 53–65, Jan. 2017.
- [27] R. Wang and G. Zheng, "Disentangled representation learning for deep MR to CT synthesis using unpaired data," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2021, pp. 274–278.
- [28] M. Wiatrak, S. V. Albrecht, and A. Nystrom, "Stabilizing generative adversarial networks: A survey," 2019, arXiv:1910.00927.
- [29] P. Su *et al.*, "Transcranial MR imaging–guided focused ultrasound interventions using deep learning synthesized CT," *Amer. J. Neuroradiol.*, vol. 41, no. 10, pp. 1841–1848, Oct. 2020.
- [30] H. Liu, M. K. Sigona, T. J. Manuel, L. M. Chen, C. F. Caskey, and B. M. Dawant, "Synthetic CT skull generation for transcranial MR imaging-guided focused ultrasound interventions with conditional adversarial networks," *Proc. SPIE*, vol. 12034, pp. 135–143, Apr. 2022.
- [31] H. Koh, T. Y. Park, Y. A. Chung, J.-H. Lee, and H. Kim, "Acoustic simulation for transcranial focused ultrasound using GAN-based synthetic CT," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 1, pp. 161–171, Jan. 2022.
- [32] K. Hynynen and F. A. Jolesz, "Demonstration of potential noninvasive ultrasound brain therapy through an intact skull," *Ultrasound Med. Biol.*, vol. 24, no. 2, pp. 275–283, Feb. 1998.
- [33] L. A. Gimeno, E. Martin, O. Wright, and B. E. Treeby, "Experimental assessment of skull aberration and transmission loss at 270 kHz for focused ultrasound stimulation of the primary visual cortex," in *Proc. IEEE Int. Ultrason. Symp. (IUS)*, Oct. 2019, pp. 556–559.
- [34] M. Wintermark *et al.*, "T1-weighted MRI as a substitute to CT for refocusing planning in MR-guided focused ultrasound," *Phys. Med. Biol.*, vol. 59, no. 13, p. 3599, 2014.
- [35] G. W. Miller, M. Eames, J. Snell, and J.-F. Aubry, "Ultrashort echo-time MRI versus CT for skull aberration correction in MR-guided transcranial focused ultrasound: *In vitro* comparison on human calvaria," *Med. Phys.*, vol. 42, no. 5, pp. 2223–2233, Apr. 2015.
- [36] E. M. Johnson, U. Vyas, P. Ghanouni, K. B. Pauly, and J. M. Pauly, "Improved cortical bone specificity in UTE MR imaging," *Magn. Reson. Med.*, vol. 77, no. 2, pp. 684–695, Feb. 2017.
- [37] S. Guo et al., "Feasibility of ultrashort echo time images using full-wave acoustic and thermal modeling for transcranial MRI-guided focused ultrasound (tcMRgFUS) planning," *Phys. Med. Biol.*, vol. 64, no. 9, Apr. 2019, Art. no. 095008.
- [38] J. Caballero-Insaurriaga *et al.*, "Zero TE MRI applications to transcranial MR-guided focused ultrasound: Patient screening and treatment efficiency estimation," *J. Magn. Reson. Imag.*, vol. 50, no. 5, pp. 1583–1592, Nov. 2019.
- [39] W. Legon, P. Bansal, R. Tyshynsky, L. Ai, and J. K. Mueller, "Transcranial focused ultrasound neuromodulation of the human primary motor cortex," *Sci. Rep.*, vol. 8, no. 1, pp. 1–14, Dec. 2018.
- [40] K. Zeng *et al.*, "Induction of human motor cortex plasticity by theta burst transcranial ultrasound stimulation," *Ann. Neurol.*, vol. 91, no. 2, pp. 238–252, Feb. 2022.
- [41] W. Lee *et al.*, "Transcranial focused ultrasound stimulation of human primary visual cortex," *Sci. Rep.*, vol. 6, no. 1, pp. 1–12, Sep. 2016.
- [42] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectationmaximization algorithm," *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 45–57, Jan. 2001.
- [43] M. Jenkinson and S. Smith, "A global optimisation method for robust affine registration of brain images," *Med. Image Anal.*, vol. 5, no. 2, pp. 143–156, Jun. 2001.

- [44] M. Jenkinson, P. Bannister, M. Brady, and S. Smith, "Improved optimization for the robust and accurate linear registration and motion correction of brain images," *NeuroImage*, vol. 17, no. 2, pp. 825–841, Oct. 2002.
- [45] T. Guillemot and J. Delon, "Implementation of the midway image equalization," *Image Process. Line*, vol. 6, pp. 114–129, Jun. 2016.
- [46] V. S. Fonov, A. C. Evans, R. C. McKinstry, C. Almli, and D. Collins, "Unbiased nonlinear average age-appropriate brain templates from birth to adulthood," *NeuroImage*, vol. 47, p. S102, Jan. 2009.
- [47] V. Fonov, A. C. Evans, K. Botteron, C. R. Almli, R. C. McKinstry, and D. L. Collins, "Unbiased average age-appropriate atlases for pediatric studies," *NeuroImage*, vol. 54, no. 1, pp. 313–327, Jan. 2011.
- [48] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241, doi: 10.1007/978-3-319-24574-4\_28.
- [49] A. Paszke et al., "PyTorch: An imperative style, high-performance deep learning library," in Proc. Adv. Neural Inf. Process. Syst. Red Hook, NY, USA: Curran Associates, 2019, pp. 8024–8035. [Online]. Available: http://papers.neurips.cc/paper/9015-pytorch-an-imperativestyle-high-performance-deep-learning-library.pdf
- [50] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2528–2535.
- [51] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, arXiv:1412.6980.
- [52] F. Wiesinger et al., "Zero TE MR bone imaging in the head," Magn. Reson. Med., vol. 75, no. 1, pp. 107–114, 2016.
- [53] B. E. Treeby and B. T. Cox, "k-wave: MATLAB toolbox for the simulation and reconstruction of photoacoustic wave fields," *J. Biomed. Opt.*, vol. 15, no. 2, 2010, Art. no. 021314.
- [54] B. E. Treeby, J. Jaros, A. P. Rendell, and B. T. Cox, "Modeling nonlinear ultrasound propagation in heterogeneous media with power law absorption using a k-space pseudospectral method," *J. Acoust. Soc. Amer.*, vol. 131, no. 6, pp. 4324–4336, 2012.
- [55] U. Schneider, E. Pedroni, and A. Lomax, "The calibration of CT Hounsfield units for radiotherapy treatment planning," *Phys. Med. Biol.*, vol. 41, no. 1, pp. 111–124, Jan. 1996.

- [56] F. Marquet *et al.*, "Non-invasive transcranial ultrasound therapy based on a 3D CT scan: Protocol validation and *in vitro* results," *Phys. Med. Biol.*, vol. 54, no. 9, p. 2597, 2009.
- [57] N. McDannold, P. J. White, and R. Cosgrove, "Predicting bone marrow damage in the skull after clinical transcranial MRI-guided focused ultrasound with acoustic and thermal simulations," *IEEE Trans. Med. Imag.*, vol. 39, no. 10, pp. 3231–3239, Oct. 2020.
  [58] D. Chauvet *et al.*, "Targeting accuracy of transcranial magnetic
- [58] D. Chauvet *et al.*, "Targeting accuracy of transcranial magnetic resonance–guided high-intensity focused ultrasound brain therapy: A fresh cadaver model," *J. Neurosurg.*, vol. 118, no. 5, pp. 1046–1052, May 2013.
- [59] S. Pichardo, V. W. Sin, and K. Hynynen, "Multi-frequency characterization of the speed of sound and attenuation coefficient for longitudinal transmission of freshly excised human skulls," *Phys. Med. Biol.*, vol. 56, no. 1, pp. 219–250, Jan. 2011.
- [60] E. S. Wise, B. T. Cox, J. Jaros, and B. E. Treeby, "Representing arbitrary acoustic source and sensor distributions in Fourier collocation methods," *J. Acoust. Soc. Amer.*, vol. 146, no. 1, pp. 278–288, Jul. 2019.
- [61] J.-F. Aubry *et al.*, "Benchmark problems for transcranial ultrasound simulation: Intercomparison of compressional wave models," 2022, *arXiv*:2202.04552.
- [62] B. Treeby. (2022). Benchmark Problems for Transcranial Ultrasound Simulation: Intercomparison Library. (V1.0) [Code] GitHub. [Online]. Available: https://github.com/ucl-bug/transcranial-ultrasoundbenchmarks
- [63] L. Kong *et al.*, "Breaking the dilemma of medical image-to-image translation," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 1964–1978.
- [64] E. Martin and B. Treeby, "Investigation of the repeatability and reproducibility of hydrophone measurements of medical ultrasound fields," *J. Acoust. Soc. Amer.*, vol. 145, no. 3, pp. 1270–1282, 2019.
- [65] J. Reijonen, M. Pitkänen, E. Kallioniemi, A. Mohammadi, R. J. Ilmoniemi, and P. Julkunen, "Spatial extent of cortical motor hotspot in navigated transcranial magnetic stimulation," *J. Neurosci. Methods*, vol. 346, Dec. 2020, Art. no. 108893.
- [66] N. Salminen-Vaparanta, V. Noreika, A. Revonsuo, M. Koivisto, and S. Vanni, "Is selective primary visual cortex stimulation achievable with TMS?" *Hum. Brain Mapping*, vol. 33, no. 3, pp. 652–665, Mar. 2012.
- TMS?" *Hum. Brain Mapping*, vol. 33, no. 3, pp. 652–665, Mar. 2012.
  [67] J. M. Edmund and T. Nyholm, "A review of substitute CT generation for MRI-only radiation therapy," *Radiat. Oncol.*, vol. 12, no. 1, pp. 1–15, Dec. 2017.