



## Head-and-Neck MRI-only radiotherapy treatment planning From acquisition in treatment position to pseudo-CT generation

A Largent, L Marage, I Gicquiau, J-C Nunes, N Reynaert, J Castelli, E Chajon, O Acosta, G Gambarota, R de Crevoisier, et al.

### ► To cite this version:

A Largent, L Marage, I Gicquiau, J-C Nunes, N Reynaert, et al.. Head-and-Neck MRI-only radiotherapy treatment planning From acquisition in treatment position to pseudo-CT generation. *Cancer/Radiothérapie*, 2020, 24 (4), pp.288-297. 10.1016/j.canrad.2020.01.008 . hal-02532848

**HAL Id: hal-02532848**

**<https://univ-rennes.hal.science/hal-02532848>**

Submitted on 22 Aug 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# **Head-and-Neck MRI-only radiotherapy treatment planning: from acquisition in treatment position to pseudo-CT generation**

## **Planification de radiothérapie externe pour la sphère ORL à partir d'imagerie par résonance magnétique : de l'acquisition en position de traitement à la génération de pseudo- scanographie**

Axel Largent<sup>1</sup>, Louis Marage<sup>1</sup>, Ivan Gicquiau<sup>1</sup>, Jean-Claude Nunes<sup>1</sup>, Nick Reynaert<sup>2</sup>, Joël Castelli<sup>1</sup>, Enrique Chajon<sup>3</sup>, Oscar Acosta<sup>1</sup>, Giulio Gambarota<sup>1</sup>, Renaud de Crevoisier<sup>1</sup> and Hervé Saint-Jalmes<sup>1</sup>.

<sup>1</sup>Univ Rennes, CLCC Eugène Marquis, Inserm, LTSI – UMR 1099, F-35000 Rennes, France

<sup>2</sup>Medical physics Department, Institut Jules Bordet, Brussels, Belgium

<sup>3</sup>CLCC Eugène Marquis, F-35000 Rennes, France

Corresponding author: [axel.largent@hotmail.fr](mailto:axel.largent@hotmail.fr)

# Abstract

**Purpose:** In context of head-and-neck radiotherapy, this study aims to compare MR image quality according to diagnostic (DIAG) and radiotherapy (RT) setups; and to optimize an MRI-protocol (including 3D T<sub>1</sub> and T<sub>2</sub>-weighted sequences) for dose-planning (based on pseudo-CT generation).

**Materials and methods:** To compare DIAG and RT setups, Signal-to-Noise-Ratio (SNR) and percentage-image-uniformity (PIU) were computed on T<sub>1</sub> images of phantoms and volunteers. Influence of the sample conductivity on SNR was quantified using home-made phantoms. To obtain reliable T<sub>1</sub> and T<sub>2</sub> images for RT-planning, an experimental design was performed on volunteers by using SNR, Contrast-to-Noise-Ratio (CNR) and Mean-Opinion-Score (MOS). Further, pseudo-CTs were generated from 8 patients T<sub>2</sub> images with a state-of-art deep-learning method. These pseudo-CTs were evaluated by mean-absolute-error (MAE) and mean-error (ME).

**Results:** SNR was higher for DIAG-setup compared to RT-setup (SNR-ratio = 1.3). A clear influence of the conductivity on SNR was observed. PIU was higher for DIAG-setup (38.8%) compared to RT-setup (33.5%). Regarding the protocol optimization, SNR, CNR, and MOS were 20.6, 6.16, and 3.91 for the optimal T<sub>1</sub> sequence. For the optimal T<sub>2</sub> sequence, SNR, CNR and MOS were 25.6, 44.46 and 4.0. In the whole head-and-neck area, the mean MAE and ME of the pseudo-CTs were 82.8 and -3.9 HU.

**Conclusion:** We quantified the image quality decrease induces by using an RT-setup for head-and-neck radiotherapy. To compensate this decrease, an MRI protocol was optimized by using an experimental design. This protocol of 15 minutes provides accurate images which could be used for MRI-dose-planning in clinical practice.

**Keywords:** radiotherapy treatment planning; MRI-only workflow; protocol optimization; pseudo-CT;

# Résumé

**Objectif :** Cette étude compare la qualité d'IRM obtenues à partir de systèmes d'acquisition dédiés au diagnostic (DIAG) et dédiés à la radiothérapie de la sphère ORL (RT) ; puis à optimiser un protocole d'IRM permettant le calcul de dose utilisant des pseudo-dvianographie qui a été validé sur huit patients.

**Matériels et méthodes :** Pour comparer les systèmes d'acquisition, le rapport-signal-sur-bruit et le pourcentage-d'uniformité-de-l'image (PIU) ont été calculés sur des images pondérées en  $T_1$  de fantômes et de volontaires. Afin d'obtenir des images adéquates pour la planification, un plan d'expérience a été réalisé. Par la suite, pour huit patients, des pseudo-scanographies ont été générées à partir d'une méthode d'apprentissage profond et évalués en utilisant l'erreur-moyenne-absolue (MAE) et l'erreur-moyenne (ME).

**Résultats :** Le rapport-signal-sur-bruit était supérieur de 30% pour le système-DIAG comparé au système-RT. Le PIU était supérieur pour le système-DIAG (39%) comparé au système RT (33%). Le protocole optimisé comprenant deux séquences 3D dure 15 minutes. Avec ce protocole, les moyennes des MAE et ME des pseudo-scanographies générés sur toute la sphère ORL des patients, étaient de 82.8 et -3.9 UH (unités Hounsfield).

**Conclusion :** La perte en qualité d'images induite par l'utilisation d'un système-RT pour la radiothérapie ORL a été quantifiée. Pour compenser cette perte, un protocole d'IRM a été optimisé en utilisant un plan d'expérience. Ce protocole de 15 minutes permet de générer des pseudo-scanographies pour la planification de radiothérapie.

**Mots clés :** planification de radiothérapie externe à partir d'IRM ; optimisation de protocole ; pseudo-CT.

# Introduction

In radiotherapy, X-ray imaging (CT-scan and CBCT) is the main imaging modality for treatment planning and delivery. CT-scan provides tissue electron density information required for dose calculation. CBCT allows fast imaging for patient positioning and tracking/gating of the tumor. However, X-ray imaging has a poor soft tissue contrast and is ionizing. MRI currently used for diagnosis has a better soft tissue contrast and is non-ionizing. MRI could improve patient positioning, delineations, and dose calculation. To take full advantage of this imaging, an MRI-only workflow has been proposed. This new workflow lead to new open research areas as MRI acquisition in treatment position and MRI dose calculation.

MR image acquisition in treatment position is the first step of an MRI-only radiotherapy workflow. This acquisition ensures reproducibility of patient positioning, which is required to deliver accurately the dose in the desired localization. Head-and-neck is one of the most challenging sites for this acquisition. The main difficulty is that the radiotherapy treatment (RT) equipment, such as radiation therapy flat couch, head-and-shoulder mask, head board, do not fit within the standard diagnostic coil system. To address this issue, specific receiving coils for RT planning (RT coil system) were designed by the manufacturers <sup>1-3</sup>. However, these setups are composed of flexible coils poorly fitting the patient anatomy, which decreased the image quality. In a limited number of studies <sup>4, 5</sup>, the image quality achieved by a RT coil system was quantified, and compared with a diagnostic coil system. In Liney et al. <sup>4</sup>, a commissioning of a 1.5 T MRI equipped with a RT head-and-neck coils was conducted. Volunteer and phantom 2D scans were used to determine the image quality. However, the coverage of the RT coils was too small to acquire correctly images of the lower neck region, limiting its use to some specific tumor sites. In another study, Wong et al. <sup>5</sup> assessed the image quality of a 1.5 T MRI with a RT setting composed of two flexible coils and a customized bi-lateral coil holder. The image quality assessment was conducted by following the standard ACR MRI phantom test <sup>6</sup>. Nevertheless, these tests were not performed in clinical conditions on volunteers as we propose in our study.

MRI dose calculation is the main challenge of an MRI-only workflow. As there is no direct mapping from MR intensities to electron density, this research area remains an outstanding problem. To address this issue, several pseudo-CT generation methods from MRI were proposed. These methods can be roughly divided in bulk density <sup>7</sup>, atlas-based <sup>8, 9</sup>, and learning-based methods (patch-based <sup>10, 11</sup> and deep learning methods (DLM) <sup>12-16</sup>). The DLMs appeared to be the most efficient and accurate methods in the literature. These methods aim to model the relations between HU values of the CTs and the MR intensities by training a neuronal network. Once the optimal network parameters are estimated, the model can be finally applied to a test patient MRI to generate

its corresponding pseudo-CT.

The aim of this study was firstly to compare the image quality provided by a diagnostic (DIAG) and RT coil systems and to optimize an MRI protocol for head-and-neck RT planning with volunteers in treatment position and the RT setup; secondly, to evaluate the accuracy of pseudo-CTs obtained from our optimized MRIs. These pseudo-CTs were generated by a state-of-the-art DLM<sup>15, 16</sup> which was compared favorably with other methods in the literature.

## **Materials and Methods**

All experiments were conducted according to the procedures approved by the local Institutional Review Board. Written informed consent was obtained from all participants.

### **Image acquisitions**

#### *Comparison and optimization studies*

MRI of the manufacturer phantom (GE cylindrical phantom), a salted home-made phantom and three healthy volunteers in head-and-neck treatment position, was performed at 1.5 T (Optima MR 450W, GE Healthcare, Milwaukee, USA). For the DIAG-MRI acquisitions, the standard coil system (32 channels head-and-neck units) was used. For the RT-MRI acquisitions, the specific coil system (16 channels GE MR Radiation-Oncology suite) with RT equipment (five points head-and-shoulder mask, radiotherapy flat couch and head board) was used. Figure 1 shows the DIAG and RT coil systems. A RT protocol was designed with isotropic 3D T<sub>1</sub> gradient echo and 3D T<sub>2</sub> fast spin echo (CUBE) sequences. The T<sub>1</sub> and T<sub>2</sub> sequences were applied with three and four distinct parameter sets (Table 1). The acquisitions were performed in sagittal orientation and reconstructed in axial orientation to minimize acquisition time and avoid aliasing, except for the RT T<sub>2</sub> sequences. To allow reliable estimation of noise and contrast ratio with parallel imaging acceleration<sup>17</sup>, all images were acquired twice. This double acquisition was repeated five times on manufacturer phantom and each volunteer.

#### *Pseudo-CT generation*

Eight patients treated for head-and-neck cancer were considered. These patients received a CT scan

and MRI in treatment position. CT scans were acquired with a Philips large-bore scanner (120 Kv, 2 mm slice thickness). For MRI, 3D T<sub>2</sub>-weighted sequences were performed using our optimized protocol (TE = 100.3 ms, TR = 3000 ms, flip angle = 90°, FOV = 410 mm, voxel sizes = 1.6 mm<sup>3</sup>). To correct MRI non-uniformity, the images were preprocessed by using N4 bias-field correction and histogram matching as in Dowling et al. <sup>8</sup>. Each CT was registered to its corresponding MRI with a rigid registration, followed by a non-rigid registration.

## **Image quality metrics**

To evaluate the MR image quality Signal-to-Noise Ratio (SNR), Contrast-to-Noise Ratio (CNR), Percentage Image Uniformity (PIU) and Mean Opinion Score (MOS) were considered. These metrics were described in the paragraphs below.

### ***Signal-to-Noise Ratio***

The Signal-to-Noise Ratio (SNR) is a metric used to estimate the quality of the acquired signal related to the noise. A simple method to determine SNR is to compute, on a single image, the ratio between the mean of the signal in the region of interest (ROI) and the standard deviation in the background. However, new MRI reconstruction methods modify the noise distribution in the images, and the standard computation method of the SNR is not reliable. In this study, SNR was computed with a more appropriate method using double acquisition <sup>17, 18</sup>. The signal was estimated by computing the mean image of the double acquisition. The noise was estimated by computing the difference image of the double acquisition. ROIs were placed inside the resulting images and the SNR was computed with the following formula:

$$SNR = \frac{\mu}{\sigma} \sqrt{2} \quad (1)$$

where  $\mu$  is the signal measured on the average image of the double acquisition,  $\sigma$  is the standard deviation measured on the difference image,  $\sqrt{2}$  is the correction factor taking into account the double acquisition.

### ***Contrast-to-Noise Ratio***

CNR aims to quantify the contrast between two given tissues. As for SNR, a double acquisition was performed for CNR computation. This metric was computed, inside two ROIs ( $ROI_1$ ,  $ROI_2$ ) placed

in two distinct tissues, with the following formula:

$$CNR = \frac{|I_{mean}(\mu(ROI_1)) - I_{mean}(\mu(ROI_2))|}{\frac{1}{2}(I_{diff}(\sigma(ROI_1)) + I_{diff}(\sigma(ROI_2)))} \sqrt{2} \quad (2)$$

where  $I_{mean}$  is the mean image of the double acquisition and  $I_{diff}$  the difference image.

### ***Percentage Image Uniformity***

PIU was used to evaluate this image uniformity<sup>19</sup>. PIU was computed with the following formula:

$$PIU = 1 - \left( \frac{\mu(ROI_{max}) - \mu(ROI_{min})}{\mu(ROI_{max}) + \mu(ROI_{min})} \right) \quad (3)$$

where  $ROI_{max}$  ( $ROI_{min}$ ) is the region with the highest signal (the region with the lowest signal) inside the image. In our study, a 3D extension of this metric was implemented by using the Insight-Toolkit library<sup>20</sup> (ITK).

### ***Mean Opinion Score***

MOS is a measure used in psychometry, which aims to quantify the opinion of a group of individuals. This metric consists of averaging the scores (appreciations) given by several experts.

## **Comparison of the DIAG and RT coil systems**

To compare the image quality achieved by the DIAG and RT systems, SNR (Eq. 1) and PIU (Eq. 3) were estimated on 3D T<sub>1</sub>-weighted images of phantoms and volunteers (TE: 4.2 ms; TR: 7.0 ms; Flip angle: 15°; bandwidth: 195.3 Hz/px, FOV: 380 x 380 x 414 mm<sup>3</sup>). Moreover, PIU was used to compare PURE and N4<sup>21</sup> non-uniformity correction algorithms.

### **Signal-to-Noise ratio study**

For the manufacturer phantom images, five circular ROIs were placed in the central slice. For the volunteer images, five circular ROIs were placed in the organs-at-risk (brainstem, up to corpus callosum white matter, cerebellum, tongue and scalene muscle). SNR measurements were performed inside the ROIs and averaged for each coil system. The SNR ratios between the DIAG and RT coil systems were computed to compare image quality.

The SNR (Eq. 1) is impacted by different noise sources in the MR scanning process. These sources



are mainly the receiving coil resistance and losses originating from human body tissues. These losses are proportional to the conductivity of tissues<sup>22</sup>. To simulate the noise source, independently from the coil resistance, a home-made uniform phantom mimicking the head-and-neck anatomy and containing a saline solution was imaged. Three salt concentrations were used (0, 3.2g/L and 5 g/L). For the unsalted solution, the noise originating from the coil is evaluated, at 3.2 g/L an average conductivity of human body tissues is considered<sup>22</sup>, and 5g/L corresponds to a maximum barely present in human body tissues.

### ***Percentage Image Uniformity study***

Manual delineations of the manufacturer phantom were performed, separating the VOI from the background. Then, PIU (Eq. 3) was computed inside the VOI before and after application of the PURE (phased-array uniformity enhancement)<sup>23</sup> and N4 image non-uniformity correction algorithms<sup>21, 24</sup>. PURE is a tool developed by an MRI manufacturer used in clinical imaging and N4 is an algorithm originated from computer vision used in research. The efficiency of both non-uniformity correction algorithms is compared.

## **Optimization of the T<sub>1</sub> and T<sub>2</sub> sequences for the RT coil system**

To obtain reliable 3D T<sub>1</sub> and T<sub>2</sub> MR images for head-and-neck radiotherapy planning, an experimental design was performed. Three T<sub>1</sub> and four T<sub>2</sub> parameter sets were defined, and SNR (Eq. 1), CNR (Eq. 2) and MOS were used as criteria to select the optimal parameter set. During the experiment, clinical constraints were also considered (patient and RT equipment set-up time).

### ***Signal-to-Noise ratio study***

Circular ROIs were placed in the organs-at-risk (brainstem, up to corpus callosum white matter, cerebellum, tongue and scalene muscle) of the volunteers. SNR (Eq. 1) were computed inside the ROIs, and averaged for all T<sub>1</sub> and T<sub>2</sub> images. The SNR was used to select the optimal parameter set.

### ***Contrast-to-Noise ratio study***

Circular ROIs were placed in the scalene muscle and the brainstem of the volunteers. CNR (Eq. 2) were computed inside the ROIs, and averaged for each T<sub>1</sub> and T<sub>2</sub> images. The CNR was used to select the optimal parameter set.

### ***Mean Opinion Score study***

The MOS was obtained from eight radiotherapists and five physicists. They ranked the volunteer images for each  $T_1$  and  $T_2$  parameter sets.

## **Pseudo-CT generation**

To evaluate performances of our optimized protocol, pseudo-CTs were generated from the patient MRIs. Each pseudo-CT was obtained by using a GAN with perceptual loss<sup>15, 16</sup>. This DLM was composed of two networks: a generator (G) and a discriminator (D), which were trained in competition. Training and validation of this method was performed by using a leave-one-out scheme.

### ***Generator network***

The generator network aimed to provide pseudo-CTs from MRIs. In this study, the generator was a U-Net<sup>13</sup>. Its architecture was composed of two networks called encoding and decoding. The encoding aimed to extract multi-scale features from the input MRI. This network was composed of 12 convolutional layers, followed by batch normalization and ReLu activation functions<sup>25</sup>.

The decoding part aimed to gradually reconstruct the pseudo-CT using the features computed during the encoding. This network was a mirror version of the encoding network.

Design of the encoding and decoding networks were detailed in supplementary materials 1.

To train the generator, a perceptual loss functions was implemented<sup>26</sup>. This loss mimics the human visual system to compare CT and pseudo-CT using similar features as opposed to only the intensities<sup>12, 26</sup>. The VGG16 network<sup>27</sup> was pretrained from the ImageNet data set<sup>28</sup>, and used to compute the features inside the CT and pseudo-CT. The perceptual loss ( $L_G$ ) function was defined as:

$$L_G(I, C) = \|VGG(C) - VGG(G(I))\|_2^2 \quad (4)$$

where  $I$  is the MRI,  $C$  is the corresponding CT,  $G(I)$  is the pseudo-CT provided by the generator,  $\|\cdot\|_2^2$  is the L2 norm, and  $VGG$  is the output of the 7<sup>th</sup> VGG16 convolutional layer.

### ***Discriminator network***

The discriminator network <sup>29</sup> aimed to classify the generated pseudo-CT as real or fake CT. Thus, the output of this network is a probability value ranging between 0 (fake) and 1 (real). The discriminator architecture was composed of six convolutional layers and one fully connected layer. Each convolutional layer was followed by batch normalization and Leaky-ReLu activation functions. The fully connected layer was followed by a sigmoid activation function. The loss function of the discriminator ( $L_D$ ) was a L2 loss.

Design of the discriminator networks was detailed in supplementary materials 1.

The generator ( $L_G$ ) and discriminator ( $L_D$ ) losses were combined to form the following adversarial loss:  $L_{adversarial}(I, C) = \lambda_1 L_D(I, C) + \lambda_2 L_G(I, C)$ , where  $I$  is the MRI,  $C$  is the corresponding CT,  $L_D(I, C)$  is the discriminator loss,  $L_G(I, C)$  is the generator loss, and  $\lambda_1$  and  $\lambda_2$  are the weights for the discriminator and generator losses (Eq. 4).

Hyper-parameter setting of the GAN <sup>15, 16</sup> was detailed in supplementary materials 1.

### ***Pseudo-CT evaluation***

To evaluate the accuracy of the DLM <sup>15, 16</sup>, a voxel-wise comparison of the HU from CT and pseudo-CT was performed. For this aim, the mean absolute error (MAE) and the mean error (ME) were calculated for the whole body. These endpoints were defined as:

$$MAE = \frac{1}{n} \sum_{i=1}^n |HU_{CT}(i) - HU_{pCT}(i)| \quad (5)$$

and

$$ME = \frac{1}{n} \sum_{i=1}^n HU_{CT}(i) - HU_{pCT}(i) \quad (6)$$

## **Results**

### **Comparison of the DIAG and RT setups**

Table 2 gives the SNR value and ratio between the manufacturer phantom and volunteer images. The SNR was always higher for the DIAG coil system than for the RT coil system. The SNR ratios of both coil systems were 1.3 for phantom and volunteers. As the SNR increases with the square root of the acquisition time, it will require an increase of 70% of the acquisition time

$((\text{SNR ratio})^2 = 1.69)$  to reach the diagnostic image quality with the RT coil system. Table 3 shows the SNR values obtained on the home-made phantom filled with different concentrations of NaCl. The SNR ratios were 1.6, 1.3 and 1.1 with respectively 0, 3.2 and 5 g/L of NaCl. For this phantom, the SNR was also higher for the DIAG coil system compared to that of the RT coil system. Figure 2 illustrates the image non-uniformity provided by both coil systems before and after PURE and N4 corrections. The PIU on raw 3D  $T_1$ -weighted images is higher for the DIAG coil system compared to that of the RT coil system (38.8 % against 33.5 %). The N4 algorithm provided higher values compared to the PURE (DIAG: 86.7 % against 41.7%, RT: 76.6 % against 49.0 %).

## Optimization of the $T_1$ and $T_2$ sequences for the RT setup

Fig 3 illustrates the axial reconstructed  $T_2$ -weighted MR images (parameter sets v1, v2, v3) and the native  $T_2$ -weighted MR image (parameter set v4), for one volunteer. Table 4 gives the SNR values from the volunteer images acquired with the distinct  $T_1$  and  $T_2$  parameter sets. Higher SNR values were obtained for the v3  $T_1$  and v1  $T_2$ -weighted images. Table 4 gives also the CNR values for the volunteer images. CNR values were higher for the v3  $T_1$  and v4  $T_2$ -weighted images compared to other parameter sets. The lowest values were found for the v2  $T_2$ -weighted images. Additional Table 1 shows the MOS values given by 8 radiotherapists and 5 physicists. The maximum MOS values were found for the v3  $T_1$  and v4  $T_2$ -weighted images. These values correspond to the appreciation “good”. The scores range from 2.64 to 4.0. Fig 4 illustrates the MR images obtained with the v3  $T_1$  and v4  $T_2$  sequence parameters, for one volunteer wearing a head and shoulder mask within the RT setup. These sequence parameter sets were retained for the clinical protocol. The patient set-up time with radiotherapy device was around 15 minutes (compared to 5 minutes for diagnostic set-up). The acquisition times were 3.21 and 6.22 minutes and the full protocol time was 15 minutes.

## Pseudo-CT evaluation

Fig 5. illustrates for one patient the MRI, the real CT and the pseudo-CT generated by the GAN using the perceptual loss. The MAE and ME (Eq. 5 and Eq. 6) obtained by using the GAN DLM on our optimized patient MRI were 82.8 ( $\pm 48.6$ ) HU and -3.9 ( $\pm 12.8$ ) HU. The computation time to generate one pseudo-CT was approximately 35 seconds.

## Discussion

In a context of head-and-neck MRI-only radiotherapy, we quantify the loss of image quality caused by using a RT setup for acquisition in treatment position. The 30% image quality decrease found would require a 70% increase in acquisition time to be fully compensated (as in diagnostic). To compensate this decrease, an experimental design was proposed to optimize a clinical MRI protocol. The obtained MR images are suitable for dose planning (acquisition in treatment position and dose calculation).

To compare the performances of the DIAG and RT coil systems, 3D isotropic  $T_1$  images were acquired on phantoms and healthy volunteers, covering all the head-and-neck anatomy.

SNR ratios of 1.3 between the DIAG and RT coil systems were found for the manufacturer phantom and volunteers. In another study, Liney et al. <sup>4</sup> reported ratios equal to 1.09, 1.55 and 1.9 (respectively in axial, coronal and sagittal planes) between the DIAG and RT coil systems. These values are close to our SNR ratio. However, they are hardly comparable because the authors performed the imaging in 2D with a different RT setup. Moreover, the coverage of these RT coil (head only) was too small to image correctly the lower neck region, limiting its use to some tumor sites.

The comparison between different saline solutions showed the importance of the conductivity losses for head-and-neck imaging at 1.5 T. As expected, the DIAG coil system is by design far more optimized for head-and-neck imaging than the flexible RT coil system. This is demonstrated by the 1.6 ratio of SNR between the two systems for the pure water phantom. When the salt concentration increases, losses in the phantom increase, reducing the impact of the coil design. In our case, the SNR of the two coil systems are equivalent for a 5 g/L salt concentration. At 3.2 g/L, the observed ratio of 1.3 is similar to the ratio observed on volunteer, supporting the use of this salt concentration in phantoms mimicking the human body <sup>22</sup>.

The DIAG coil system is less affected by image non-uniformity compared to the RT coil system. This system is designed with coil elements surrounding all the patient head, thus providing a very uniform signal. The flexible RT coil system is designed to accommodate different patient anatomies, this explains its non-uniformity. 3D PIU values computed on the 3D  $T_1$  image were 38.8% and 33.5 % for the DIAG and RT systems. In another study, Wong et al. <sup>5</sup> reported 2D PIU values equal to 92.7% and 93.4% for the DIAG and RT setups. These values are hardly comparable with our PIU results because they were only computed using one axial slice of 2D head images. The N4 non-uniformity correction algorithm provided better results compared to the manufacturer method PURE. Nevertheless, the N4 method has a lot of tuning parameters requiring strong knowledge in computer vision making it unappealing for clinical routine. The PURE method doesn't need tuning parameters. Indeed, the PURE method appears more appropriate for clinical

routine while the N4 method can be used for pseudo-CT generation <sup>30</sup>.

A design experiment was performed on the 3D  $T_1$  and  $T_2$  volunteer images, to select the best sequence parameters for delineation and MRI dose calculation (pseudo-CT generation) during the RT planning.

On the  $T_1$ -weighted MR images, the whole quantitative and subjective criteria provided the same results. The v3  $T_1$  parameter set gave the best results with mean SNR, mean CNR and MOS respectively equal respectively to 20.6, 6.16 and 3.91. Conversely, for the  $T_2$ -weighted images these criteria do not converge to the same result. As tissue contrast is more important than noise for delineation tasks during the RT planning, CNR and MOS were mostly considered to select the optimal  $T_2$  parameter set. The v4  $T_2$  parameter set provided the best results with mean CNR and MOS equal to 44.46 and 4.0.

During our experiment, clinical constraints as patient and RT equipment set-up time, and patient discomfort (caused by the head and shoulder mask), were also considered. The patient RT set-up time was around 15 minutes compared to 5 minutes for the DIAG set-up time. The acquisition time of the whole protocol was 15 minutes which is shorter than a standard radiotherapy treatment session (~ 30 min).

Pseudo-CT generation was investigated in a limited number of studies for head-and-neck site. Guerreiro et al. <sup>31</sup> used an ABM to generate pseudo-CTs in the head-and-neck area and reported a MAE = 90.7 HU. Johansson <sup>32</sup> performed a Gaussian mixture and obtained a MAE = 137 HU. Our GAN method with perceptual loss <sup>15, 16</sup> compared favorably with these studies. Our optimized MRI allow us to obtained accurate pseudo-CTs.

Our study presents some limitations. Firstly, MRI geometrical distortions were not evaluated. However, for head-and-neck site geometrical distortions are generally low and inferior to 2 mm, which is acceptable for RT planning <sup>3, 33, 34</sup>. Secondly, the efficient and low-SAR (Specific-Absorption-Rate) 3D gradient echo  $T_1$ -weighted sequence used in this study could be sensible to dental artifacts. A fast spin echo alternative was added as an option to address this issue. Thirdly, we do not quantify the impact of using an RT-setup on tumor and organs-at-risk delineations. This quantification could be very challenging to perform because of bias from IRM-DIAG IRM-RT deformable registrations required to compare these delineations. Moreover, only three volunteers were scanned for practical reasons (fabrication of personalized head and shoulder mask, double acquisitions performed five times for each phantom and healthy volunteer, availability of the clinical MRI scanner). Finally, a dosimetric evaluation of the obtained pseudo-CTs was not

performed, it is part of future works.

## **Conclusion**

This study aims to improve patient positioning and dose targeting in head-and-neck MRI-only radiotherapy by using MRI acquired in treatment position (with RT-setup). To compensate loss of MR image quality generated by RT-setup, a design experiment was proposed to optimize a head-and-neck MRI protocol. The obtained optimized MR images allow accurate and fast pseudo-CT generation by a generative adversarial network using perceptual loss. These images could be therefore used for dose planning in clinical practice.

## **Acknowledgement**

We would like to thank the MRI radiographers of our clinical center for assistance during the data acquisition. The volunteers were included in the OSS-IRM study, supported by the University Hospital of Rennes and the University of Rennes. This work was financed, in part, by the French Institut National du Cancer (INCa) in the context of a PhysiCancer project.



## References

- <sup>1</sup> Devic S. MRI simulation for radiotherapy treatment planning. *Med. Phys.* 2012; **39**(11): 6701–6711.
- <sup>2</sup> Paulson ES, Erickson B, Schultz C, Allen Li X. Comprehensive MRI simulation methodology using a dedicated MRI scanner in radiation oncology for external beam radiation treatment planning. *Med. Phys.* 2015; **42**(1): 28–39.
- <sup>3</sup> Paulson ES, Crijns SPM, Keller BM, Wang J, Schmidt MA, Coutts G, et al. Consensus opinion on MRI simulation for external beam radiation treatment planning. *Radiother. Oncol.* 2016; **121**(2): 187–192.
- <sup>4</sup> Liney GP, Owen SC, Beaumont AKE, Lazar VR, Manton DJ, Beavis AW. Commissioning of a new wide-bore MRI scanner for radiotherapy planning of head and neck cancer. *Br. J. Radiol.* 2013; **86**(1027): 20130150.
- <sup>5</sup> Wong OL, Yuan J, Yu SK, Cheung KY. Image quality assessment of a 1.5T dedicated magnetic resonance-simulator for radiotherapy with a flexible radio frequency coil setting using the standard American College of Radiology magnetic resonance imaging phantom test. *Quant. Imaging Med. Surg.* 2017; **7**(2): 205–214.
- <sup>6</sup> *Phantom Test Guidance for the ACR MRI Accreditation Program. The American College of Radiology, Reston, VA. Available online: <http://www.acraccreditation.org/~media/ACRAccreditation/Documents/MRI/LargePhantomGuidance.pdf?la=en>, accessed June 2016. - Recherche Google. n.d.;*
- <sup>7</sup> Lambert J, Greer PB, Menk F, Patterson J, Parker J, Dahl K, et al. MRI-guided prostate radiation therapy planning: Investigation of dosimetric accuracy of MRI-based dose planning. *Radiother. Oncol.* 2011; **98**(3): 330–334.
- <sup>8</sup> Dowling JA, Sun J, Pichler P, Rivest-Hénault D, Ghose S, Richardson H, et al. Automatic Substitute Computed Tomography Generation and Contouring for Magnetic Resonance Imaging (MRI)-Alone External Beam Radiation Therapy From Standard MRI Sequences. *Int. J. Radiat. Oncol. Biol. Phys.* 2015; **93**(5): 1144–1153.
- <sup>9</sup> Burgos N, Cardoso MJ, Guerreiro F, et al. Robust CT Synthesis for Radiotherapy Planning: Application to the Head and Neck Region, in *Med. Image Comput. Comput.-Assist. Interv. -- MICCAI 2015*, edited by N. Navab, J. Hornegger, W.M. Wells and A.F. Frangi (Springer International Publishing, 2015), pp. 476–484.
- <sup>10</sup> Andreasen D, Van Leemput K, Edmund JM. A patch-based pseudo-CT approach for MRI-only radiotherapy in the pelvis. *Med. Phys.* 2016; **43**(8): 4742.
- <sup>11</sup> Largent A, Barateau A, Nunes J-C, Lafond C, Greer PB, Dowling JA, et al. Pseudo-CT Generation for MRI-Only Radiation Therapy Treatment Planning: Comparison Among Patch-Based, Atlas-Based, and Bulk Density Methods. *Int. J. Radiat. Oncol. • Biol. • Phys.* 2019; **103**(2): 479–490.
- <sup>12</sup> Nie D, Trullo R, Lian J, Petitjean C, Ruan S, Wang Q, et al. Medical Image Synthesis with Context-Aware Generative Adversarial Networks, in *Med. Image Comput. Comput.-Assist. Interv. – MICCAI 2017*(Springer, Cham, 2017), pp. 417–425.
- <sup>13</sup> Han X. MR-based synthetic CT generation using a deep convolutional neural network method. *Med. Phys.* 2017; **44**(4): 1408–1419.
- <sup>14</sup> Wolterink JM, Dinkla AM, Savenije MHF, Seevinck PR, Berg CAT van den, Išgum I. Deep MR to CT Synthesis Using Unpaired Data, in *Simul. Synth. Med. Imaging*(Springer, Cham, 2017), pp. 14–23.
- <sup>15</sup> Largent A, Barateau A, Nunes J-C, et al. Comparison of deep learning-based and patch-based methods for pseudo-CT generation in MRI-based prostate dose planning. *Int. J. Radiat. Oncol. • Biol. • Phys.* 2019; **0**(0).

- 16 Largent A, Nunes J, Saint-Jalmes H, et al. Pseudo-CT Generation for Mri-only Radiotherapy: Comparative Study Between A Generative Adversarial Network, A U-Net Network, A Patch-Based, and an Atlas Based Methods, in *2019 IEEE 16th Int. Symp. Biomed. Imaging ISBI 2019*(2019), pp. 1109–1113.
- 17 Firbank MJ, Coulthard A, Harrison RM, Williams ED. A comparison of two methods for measuring the signal to noise ratio on MR images. *Phys. Med. Biol.* 1999; **44**(12): N261-264.
- 18 Murphy BW, Carson PL, Ellis JH, Zhang YT, Hyde RJ, Chenevert TL. Signal-to-noise measures for magnetic resonance imagers. *Magn. Reson. Imaging* 1993; **11**(3): 425–428.
- 19 Sled JG, Pike GB. Understanding intensity non-uniformity in MRI, in *Med. Image Comput. Comput.-Assist. Interv. — MICCAI'98*(Springer, Berlin, Heidelberg, 1998), pp. 614–622.
- 20 Ibanez L, Schroeder W, Ng L, Cates J. The ITK software guide. 2005;
- 21 Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, et al. N4ITK: improved N3 bias correction. *IEEE Trans. Med. Imaging* 2010; **29**(6): 1310–1320.
- 22 Giovannetti G, Frijia F, Menichetti L, Hartwig V, Viti V, Landini L. An efficient method for electrical conductivity measurement in the RF range. *Concepts Magn. Reson. Part B Magn. Reson. Eng.* 2010; **37B**(3): 160–166.
- 23 Ogasawara G, Inoue Y, Matsunaga K, Fujii K, Hata H, Takato Y. Image Non-Uniformity Correction for 3-T Gd-EOB-DTPA-Enhanced MR Imaging of the Liver. *Magn. Reson. Med. Sci.* 2016; **16**(2): 115–122.
- 24 Boyes RG, Gunter JL, Frost C, Janke AL, Yeatman T, Hill DL, et al. Intensity non-uniformity correction using N3 on 3-T scanners with multichannel phased array coils. *NeuroImage* 2008; **39**(4): 1752–1762.
- 25 Hahnloser RHR, Sarpeshkar R, Mahowald MA, Douglas RJ, Seung HS. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature* 2000; **405**(6789): 947–951.
- 26 Yang Q, Yan P, Zhang Y, , Yu H, Shi Y, Mou X, Kalra MK et al. Low-Dose CT Image Denoising Using a Generative Adversarial Network With Wasserstein Distance and Perceptual Loss. *IEEE Trans. Med. Imaging* 2018; **37**(6): 1348–1357.
- 27 Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *ArXiv14091556 Cs* 2014;
- 28 *ImageNet Large Scale Visual Recognition Challenge* | *SpringerLink*. n.d.;
- 29 Wang C, Xu C, Wang C, Tao D. Perceptual Adversarial Networks for Image-to-Image Transformation. *IEEE Trans. Image Process.* 2018; **27**(8): 4066–4079.
- 30 Johnstone E, Wyatt JJ, Henry AM, et al. Systematic Review of Synthetic Computed Tomography Generation Methodologies for Use in Magnetic Resonance Imaging–Only Radiation Therapy. *Int. J. Radiat. Oncol.* 2018; **100**(1): 199–217.
- 31 Guerreiro F, Burgos N, Dunlop A, Wong K, Petkar I, Nutting C, et al. Evaluation of a multi-atlas CT synthesis approach for MRI-only radiotherapy treatment planning. *Phys. Med.* 2017; **35**: 7–17.
- 32 Johansson A, Karlsson M, Nyholm T. CT substitute derived from MRI sequences with ultrashort echo time. *Med. Phys.* 2011; **38**(5): 2708–2714.
- 33 Torfeh T, Hammoud R, Perkins G, McGarry M, Aouadi S, Celik A, et al. Characterization of 3D geometric distortion of magnetic resonance imaging scanners commissioned for radiation therapy planning. *Magn. Reson. Imaging* 2016; **34**(5): 645–653.
- 34 Walker A, Liney G, Metcalfe P, Holloway L. MRI distortion: considerations for MRI based radiotherapy treatment planning. *Australas. Phys. Eng. Sci. Med.* 2014; **37**(1): 103–113.
- 35 Kingma DP, Ba J. Adam: A Method for Stochastic Optimization. *ArXiv14126980 Cs* 2014;

	<b>V1 T<sub>1</sub></b>	<b>V2 T<sub>1</sub></b>	<b>V3 T<sub>1</sub></b>	<b>V1 T<sub>2</sub></b>	<b>V2 T<sub>2</sub></b>	<b>V3 T<sub>2</sub></b>	<b>V4 T<sub>2</sub></b>
<b>FOV (mm)</b>	380	380	380	380	380	380	410
<b>Matrix (px)</b>	260x260	260x260	260x260	192x190	192x190	192x190	256x256
<b>Voxel size (mm<sup>3</sup>)</b>	1.5 <sup>3</sup>	1.5 <sup>3</sup>	1.5 <sup>3</sup>	2.0 <sup>3</sup>	2.0 <sup>3</sup>	2.0 <sup>3</sup>	1.6 <sup>3</sup>
<b>Number of slices</b>	176	176	124	224	224	224	242
<b>TE (ms)</b>	4.2	1.12	4.2	74.6	131.9	95.5	100.3
<b>TR (ms)</b>	7.0	3.5	7.0	2000	2000	2000	3000
<b>Flip angle (°)</b>	15	10	15	90	90	90	90
<b>Bandwidth (Hz/px)</b>	195.3	325.5	195.3	325.5	244.1	244.1	122.1
<b>Acceleration factor</b>	2	2	1	2	2	1	1.5
<b>Acquisition time (min:s)</b>	2:22	1:11	3:21	3:17	2:13	3:39	6:31
<b>Acquisition orientation</b>	Sagittal	Sagittal	Sagittal	Sagittal	Sagittal	Sagittal	Axial

Table 1. 3D T<sub>1</sub> and T<sub>2</sub>-weighted sequence parameters. Three T<sub>1</sub> and four T<sub>2</sub>-weighted sequence parameters were tested for the comparison of the coil systems and the design optimization. Each acquisition was performed twice and repeated five times.

Table 1. Paramètres des séquences 3D pondérées en T<sub>1</sub> et T<sub>2</sub>. Trois paramètres de séquences pondérées en T<sub>1</sub> et quatre paramètres de séquences pondérées en T<sub>2</sub> ont été testées pour la comparaison des systèmes d'acquisition et la réalisation du plan d'expérience. Chaque séquence a été acquise deux fois avec cinq répétitions.

	<b>Manufacturer phantom</b>		<b>Volunteer</b>	
<b>Coil systems</b>	<b>DIAG</b>	<b>RT</b>	<b>DIAG</b>	<b>RT</b>
<b>SNR</b>	173.8 ( $\pm 10.0$ )	133.8 ( $\pm 12.3$ )	21.2 ( $\pm 5.2$ )	16.3 ( $\pm 3.3$ )
<b>Ratio</b>	1.3		1.3	

Table 2. SNR from 3D T<sub>1</sub>-weighted MR images of the manufacturer phantom and the volunteers. The ratios between the SNR of the both coil systems were computed to quantify their differences in term of image quality. Each acquisition was performed twice and repeated five times.

Table 2. SNR provenant d'images IRM 3D pondérées en T<sub>1</sub> d'un fantôme et de volontaires. Le ratio entre les SNR des deux systèmes d'acquisition a été calculé pour quantifier leurs différences en termes de qualité d'images. Chaque séquence a été acquise deux fois avec cinq répétitions.

<b>NaCL (g/L)</b>	<b>DIAG SNR</b>	<b>RT SNR</b>	<b>Ratio</b>
<b>0</b>	143.6 ( $\pm 4.3$ )	91.9 ( $\pm 2.7$ )	1.6
<b>3.2</b>	90.9 ( $\pm 1.6$ )	72.2 ( $\pm 3.6$ )	1.3
<b>5</b>	79.7 ( $\pm 3.3$ )	71.7 ( $\pm 2.8$ )	1.1

Table 3. Influence of the conductivity on the SNR measurements from DIAG and RT coil systems.

SNR from 3D T<sub>1</sub>-weighted MR images of a home-made phantom filled with different concentrations of NaCl. Each acquisition was performed twice.

Table 3. Influence de la conductivité sur les SNR provenant des systèmes d'acquisition de diagnostic et de radiothérapie. Ces SNR proviennent d'images IRM 3D pondérées en T<sub>1</sub> de fantômes remplis avec différentes concentrations de NaCl. Chaque acquisition a été réalisée deux fois.

	T <sub>1</sub> -weighted SNR			T <sub>2</sub> -weighted SNR				T <sub>1</sub> -weighted CNR			T <sub>2</sub> -weighted CNR			
	V1	V2	V3	V1	V2	V3	V4	V1	V2	V3	V1	V2	V3	V4
<b>Volunteer 1</b>	16.3 (± 3.3)	10.2 (± 3.3)	20.5 (± 2.9)	51.9 (± 6.8)	23.9 (± 8.6)	28.1 (± 10.8)	29.3 (± 5.3)	4.1 (± 2.9)	2.0 (± 1.3)	5.3 (± 4.2)	19.9 (± 5.8)	14.1 (± 4.2)	16.5 (± 3.5)	50.6 (± 10.8)
<b>Volunteer 2</b>	18.7 (± 2.7)	9.8 (± 1.6)	20.4 (± 5.5)	46.3 (± 10.2)	29.7 (± 4.2)	49.5 (± 9.4)	25.0 (± 3.5)	5.1 (± 3.0)	2.2 (± 1.1)	6.0 (± 3.1)	15.9 (± 3.7)	12.1 (± 2.4)	14.7 (± 3.9)	45.9 (± 8.5)
<b>Volunteer 3</b>	18.0 (± 2.8)	10.6 (± 2.0)	20.9 (± 2.7)	47.9 (± 10.6)	25.0 (± 5.4)	39.3 (± 14.1)	22.6 (± 7.5)	6.1 (± 2.4)	2.6 (± 1.2)	7.2 (± 3.4)	20.2 (± 2.8)	17.4 (± 4.0)	18.2 (± 6.0)	36.3 (± 10.8)

Table 4. SNR and CNR from 3D T<sub>1</sub> and T<sub>2</sub>-weighted MR images of the three volunteers (with RT equipment for RT acquisitions). Three T<sub>1</sub> and four T<sub>2</sub>-weighted sequence parameters were tested for the comparison of the coil systems and the design optimization. Each acquisition was performed twice and repeated five time

Table 4. SNR et CNR provenant d'images IRM 3D pondérées en T<sub>1</sub> et T<sub>2</sub> de trois volontaires. Trois paramètres de séquences pondérées en T<sub>1</sub> et quatre paramètres de séquences pondérées en T<sub>2</sub> ont été testées pour la comparaison des systèmes d'acquisition et la réalisation du plan d'expérience. Chaque séquence a été acquise deux fois avec cinq répétitions.

	<b>V1 T<sub>1</sub></b>	<b>V2 T<sub>1</sub></b>	<b>V3 T<sub>1</sub></b>	<b>V1 T<sub>2</sub></b>	<b>V2 T<sub>2</sub></b>	<b>V3 T<sub>2</sub></b>	<b>V4 T<sub>2</sub></b>
<b>MOS</b>	3.45 (± 0.6)	2.64 (± 0.7)	3.91 (± 0.8)	3.25 (± 0.6)	3.08 (± 0.8)	3.08 (± 0.7)	4.0 (± 0.6)

Additional Table 1. The Mean Opinion Score (MOS) given by 8 radiotherapists and 5 physicists.

The score label is defined by bad = 1, poor = 2, fair = 3, good = 4, excellent = 5.

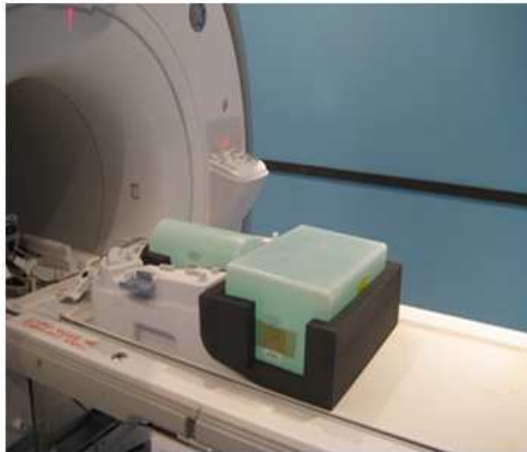
Additional Table 1. La note d'opinion moyenne (MOS) donnée par huit radiothérapeutes et cinq physiciens. Le label associé à chaque note est : très mauvais = 1, mauvais = 2, acceptable = 3, bien = 4, excellent = 5.



(a)



(b)



(c)



(d)

Fig 1. DIAG (a) and RT (b) coil systems, manufacturer phantom (c), and one volunteer wearing a five points head and shoulder mask (d)

Fig 1. Systèmes d'acquisitions de diagnostic (a) et de radiothérapie (b), fantôme (c), et un volontaire portant un masque thermoplastique a cinq points (d)



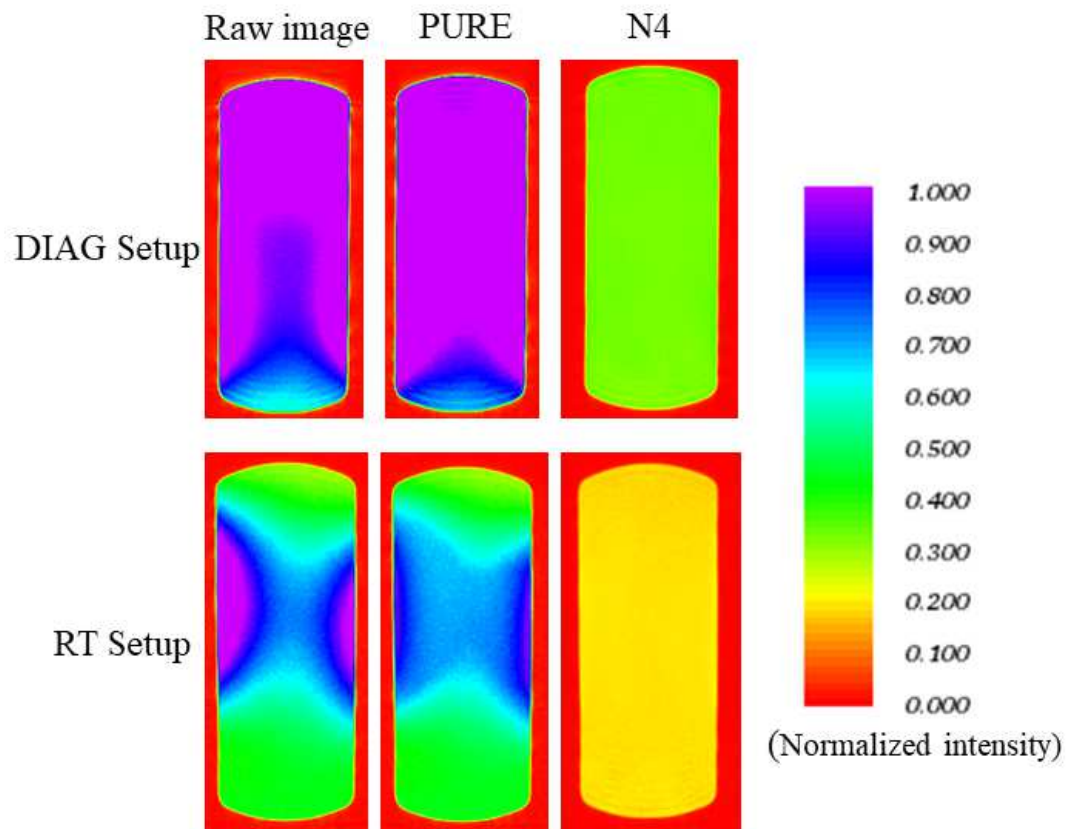


Fig 2. Illustration of the image uniformity on the 3D  $T_1$ -weighted MR images from the manufacturer phantom (before and after PURE and N4 image uniformity correction algorithms)

Fig 2. Illustration de l'uniformité d'images IRM 3D pondérées en  $T_1$  d'un fantôme (avant et après application des algorithmes PURE et N4)

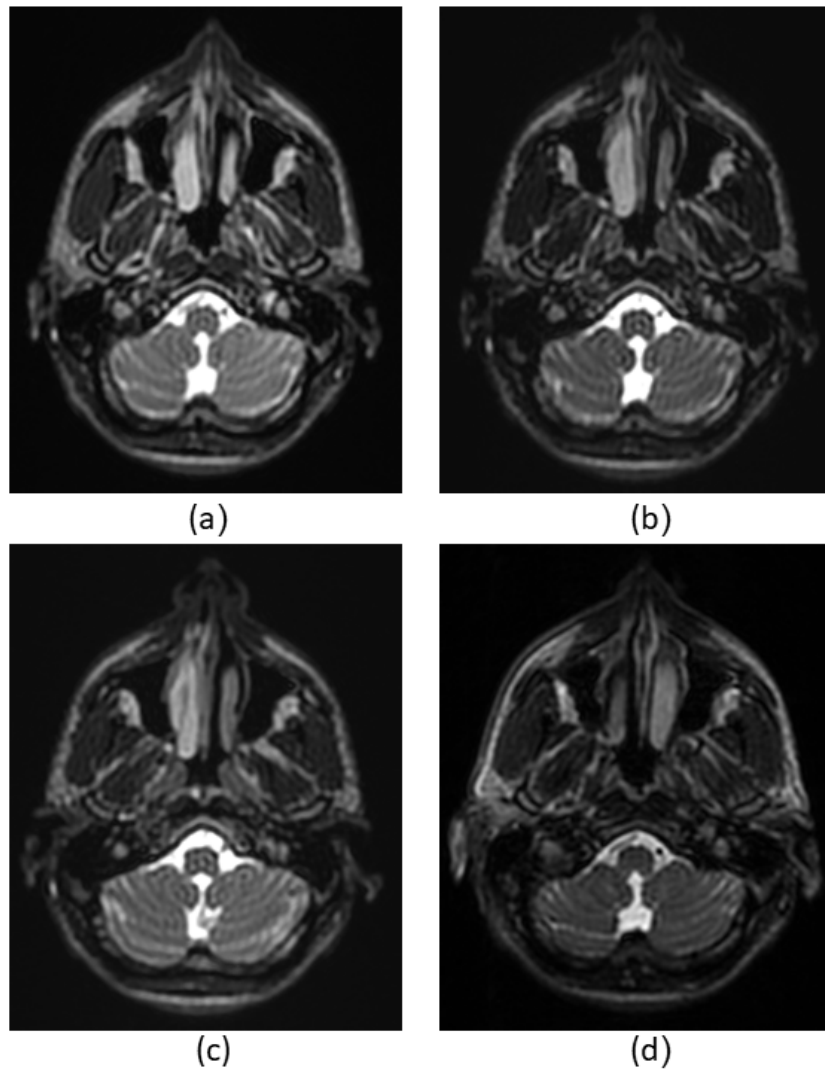


Fig 3. Illustration of the axial reconstruction of the 3D  $T_2$ -weighted MR images v1 (a), v2 (b), v3 (c), and the native 3D  $T_2$ -weighted MR image v4 (d), for one volunteer

Fig 3. Illustration de coupes axiales d'images IRM 3D pondérées en  $T_2$  reconstruites v1 (a), v2 (b), v3 (c), et d'une coupe axiale d'une image IRM 3D pondérée en  $T_2$  native v4 (d), pour un volontaire

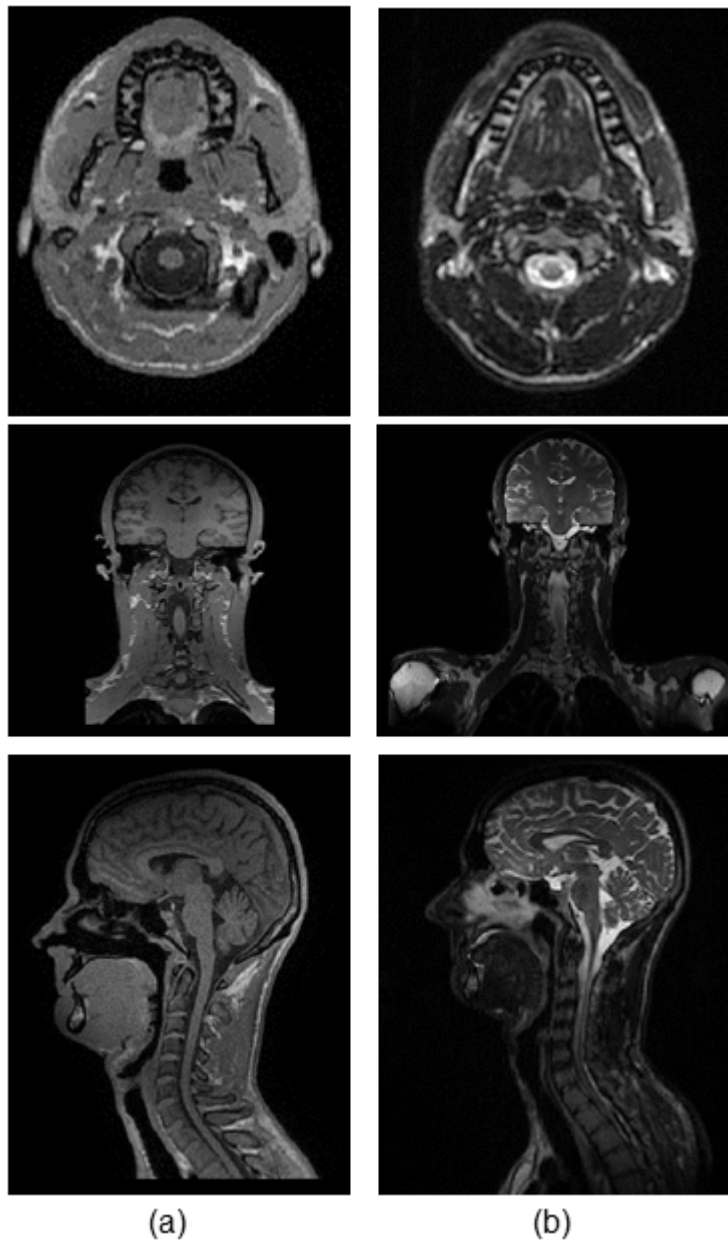


Fig 4. Illustration of the optimized 3D  $T_1$  and  $T_2$ -weighted MR images for one volunteer (in the 3 view): v3  $T_1$  (a) and v4  $T_2$  (b)

Fig 4. Illustration d'images IRM 3D optimisées pondérées en  $T_1$  et  $T_2$  pour un volontaire (dans les trois directions): v3  $T_1$  (a) et v4  $T_2$  (b)

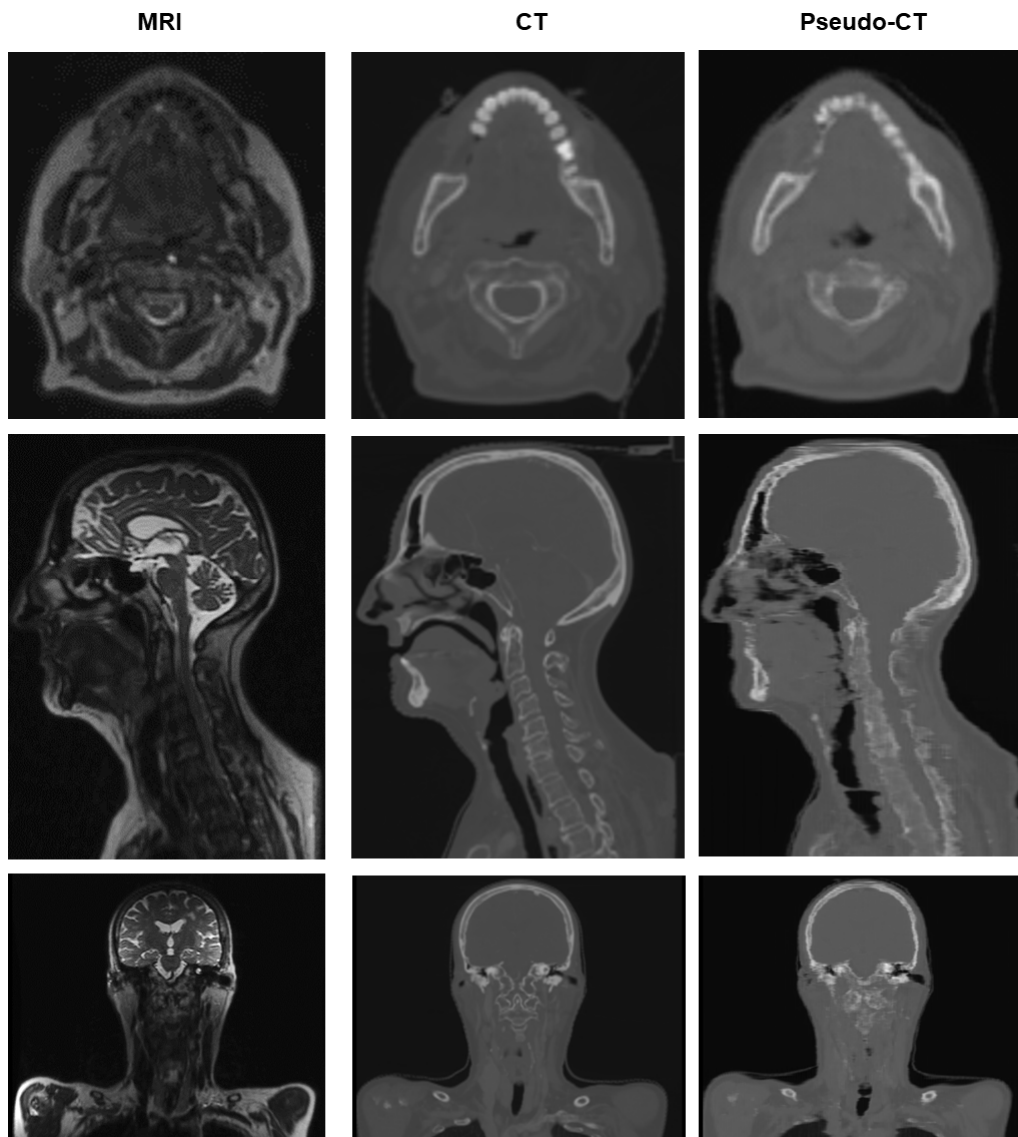


Fig 5. Illustration for one patient of the  $T_2$ -weighted MRI, real CT and pseudo-CT generated by the GAN using a perceptual loss

Fig 5. Illustration pour un patient d'une image IRM pondérée en  $T_2$ , d'une image scanner (CT-scan), et un pseudo-CT généré par une méthode d'apprentissage profond appelée GAN utilisant une fonction de coût perceptuelle

## Supplementary materials 1: Pseudo-CT generation by a GAN with perceptual loss

Each pseudo-CT were obtained by using a GAN<sup>15, 16</sup> on the patient MRIs. This DLM was composed of two networks: a generator (G) and a discriminator (D), which were trained in competition.

### *Generator network*

The generator network aimed to provide pseudo-CTs from MRIs. In this study, the generator was a U-Net<sup>13</sup>. Its architecture was composed of two networks called encoding and decoding. The encoding aimed to extract multi-scale features from the input MRI. This network was composed of 12 convolutional layers, followed by batch normalization and ReLu activation functions<sup>25</sup>. The filter numbers of these layers were 64, 64, 128, 128, 256, 256, 256, 512, 512, 512, 512, and 512, and their filter size was  $3 \times 3$  (stride = 1). For down-sampling, convolutional layers with a filter size of  $2 \times 2$  (stride = 2) were used.

The decoding part aimed to gradually reconstruct the pseudo-CT using the features computed during the encoding. This network was a mirror version of the encoding network. For feature up-sampling, transposed 2D convolutional layers were used with a filter size of  $2 \times 2$  (stride = 2). To obtain the pseudo-CT, the last layer of the decoding network was a convolution layer with one filter (size =  $1 \times 1$ ).

To train the generator a perceptual loss functions was implemented<sup>26</sup>. This loss mimics the human visual system to compare CT and pseudo-CT using similar features as opposed to only the intensities<sup>12, 26</sup>. The VGG16 network<sup>27</sup> was pretrained from the ImageNet data set, and used to compute the features inside the CT and pseudo-CT. The perceptual loss function ( $L_G$ ) was defined as:

$$L_G(I, C) = \|VGG(C) - VGG(G(I))\|_2^2 \quad (4)$$

where  $I$  is the MRI,  $C$  is the corresponding CT,  $G(I)$  is the pseudo-CT provided by the generator,  $\|\cdot\|_2^2$  is the L2 norm, and  $VGG$  is the output of the 7<sup>th</sup> VGG16 convolutional layer.

### ***Discriminator network***

The discriminator network aimed to classify the generated pseudo-CT as real or fake CT. Thus, the output of this network is a probability value ranging between 0 (fake) and 1 (real). The discriminator architecture was composed of six convolutional layers and one fully connected layer. Each convolutional layer was followed by batch normalization and Leaky-ReLu activation functions. The filter numbers of these layers were 8, 16, 32, 64, 64 and 64. The filter size was  $3 \times 3$  (stride = 2) for the first four layers and  $1 \times 1$  (stride = 1) for the remaining layers. The fully connected layer was followed by a sigmoid activation function. The loss function of the discriminator  $L_D$  was a L2 loss.

The generator and discriminator losses were combined to form the following adversarial loss:  $L_{adversarial}(I, C) = \lambda_1 L_D(I, C) + \lambda_2 L_G(I, C)$ , where  $I$  is the MRI,  $C$  is the corresponding CT,  $L_D(I, C)$  is the discriminator loss,  $L_G(I, C)$  is the generator loss, and  $\lambda_1$  and  $\lambda_2$  are the weights for the discriminator and generator losses.

### ***Training of the GAN with perceptual loss***

The GAN DLM was trained using anatomically paired data: axial 2D slices of the training CT and MR images. Data augmentation was performed to increase the size of the training cohort. It was conducted by randomly applying affine registrations on the slices (translated by -5% to 5% per axis, rotated by  $-10^\circ$  to  $+10^\circ$ , sheared by  $-10^\circ$  to  $10^\circ$ ). A mini-batch size of 10 slices and 200 epochs was considered. The network parameters were optimized using the Adam algorithm<sup>35</sup>. The parameters of this algorithm parameters were:  $\alpha = 1 \times 10^{-4}$ ,  $\beta_1 = 0.9$ , and  $\beta_2 = 0.9$ . The weights of the discriminator and generator loss functions were:  $\lambda_1 = 100$  and  $\lambda_2 = 1$ , respectively. The computation time of the GAN DLM training was approximately 30 hours with a Nvidia GTX 1070 TI 8 Go. The computation time to generate one pseudo-CT from a new MRI (not belong to the training cohort) was approximately 35 seconds.