Received: 7 April 2025 | Revised: 1 August 2025 | Accepted: 5 August 2025

DOI: 10.1002/mrm.70050

RESEARCH ARTICLE

Magnetic Resonance in Medicine

Motion-robust T_2^* quantification from low-resolution gradient echo brain MRI with physics-informed deep learning

Hannah Eichhorn[1,2] | **Veronika Spieker**[1,2] | **Kerstin Hammernik**[2] | **Elisa Saks**[3,4] | **Lina Felsner**[2] | **Kilian Weiss**[5] | **Christine Preibisch**[3,4,6] | **Julia A. Schnabel**[1,2,7]

[1]Institute of Machine Learning in Biomedical Imaging, Helmholtz Munich, Neuherberg, Germany

[2]School of Computation, Information & Technology, Technical University of Munich, Munich, Germany

[3]Department of Diagnostic and Interventional Neuroradiology, School of Medicine & Health, TUM Klinikum Rechts der Isar, Technical University of Munich, Munich, Germany

[4]TUM-Neuroimaging Center, School of Medicine & Health, TUM Klinikum Rechts der Isar, Technical University of Munich, Munich, Germany

[5]Philips GmbH Market DACH, Hamburg, Germany

[6]Clinic of Neurology, School of Medicine & Health, TUM Klinikum Rechts der Isar, Technical University of Munich, Munich, Germany

[7]School of Biomedical Engineering & Imaging Sciences, King's College, London, UK

Correspondence
Hannah Eichhorn, Institute of Machine Learning in Biomedical Imaging, Helmholtz Munich, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany.
Email: hannah.eichhorn@tum.de

Abstract

Purpose: T_2^* quantification from gradient echo magnetic resonance imaging is particularly affected by subject motion due to its high sensitivity to magnetic field inhomogeneities, which are influenced by motion and might cause signal loss. Thus, motion correction is crucial to obtain high-quality T_2^* maps.

Methods: We extend PHIMO, our previously introduced learning-based physics-informed motion correction method for low-resolution T_2^* mapping. Our extended version, PHIMO+, utilizes acquisition knowledge to enhance the reconstruction performance for challenging motion patterns and increase PHIMO's robustness to varying strengths of magnetic field inhomogeneities across the brain. We perform comprehensive evaluations regarding motion detection accuracy and image quality for data with simulated and real motion.

Results: PHIMO+ outperforms the learning-based baseline methods both qualitatively and quantitatively with respect to line detection and image quality. Moreover, PHIMO+ performs on par with a conventional state-of-the-art motion correction method for T_2^* quantification from gradient echo MRI, which relies on redundant data acquisition.

Conclusion: PHIMO+'s competitive motion correction performance, combined with a reduction in acquisition time by over 40% compared to the state-of-the-art method, makes it a promising solution for motion-robust T_2^* quantification in research settings and clinical routine.

KEYWORDS

data-consistent image reconstruction, motion correction, motion detection, motion simulation, self-supervised optimization

1 | INTRODUCTION

Patient head motion during image acquisition remains a major challenge for brain magnetic resonance imaging (MRI). Motion-related artifacts-like blurring or ghosting-do not only result from positional inconsistencies during k-space acquisition but also from secondary effects, such as changes in the local magnetic field (B_0), gradient nonlinearities, as well as transmit and receive field heterogeneity affecting coil sensitivities.[1,2] The extent to which these secondary effects degrade image quality depends on the specific imaging sequence and hardware and their sensitivity to such perturbations.

To quantify the brain's oxygen metabolism, the multi-parametric quantitative BOLD (mqBOLD) technique uses noise-robust multi-echo T_2^*-weighted gradient echo (GRE) acquisitions at low spatial resolution ($2 \times 2 \times 3\ mm^3$).[3] However, T_2^*-weighted MRI is particularly sensitive to B_0 inhomogeneities and thus, highly susceptible to motion, as motion-induced B_0 variations can lead to signal loss.[4,5] The impact of motion-induced B_0 inhomogeneity changes on T_2^*-weighted GRE images increases with increasing echo times.[6] Since T_2^* quantification requires multiple echoes, motion may lead to erroneous T_2^* parameter maps, which can further affect derived parameter maps, such as the susceptibility related R2' relaxation rate in the mqBOLD technique.[3,7] Since mqBOLD is highly sensitive to motion effects in T_2^* parameter maps, redundant acquisitions of the k-space center are currently used for motion correction (MoCo), which significantly increases the overall acquisition time.[8] Various prospective and retrospective approaches have previously been proposed for rigid-body MoCo in brain MRI.[1] A few methods also take into account motion-induced B_0 inhomogeneity changes.[9–12] While retrospective methods are attractive as they do not rely on external hardware or sequence modifications for motion and field estimates, they are particularly challenged by the signal loss due to motion-induced B_0 inhomogeneity variations.[1] In recent years, deep learning has shown to be capable of learning complex patterns and dealing with inconsistent data and, thus, might offer promising solutions to address this challenge. Many learning-based MoCo approaches have already been proposed for brain MRI.[13] MoCo can be addressed with image denoisers, using convolutional[14,15] or generative adversarial networks.[16–18] Yet, relying exclusively on image data, these methods cannot guarantee consistency with the measured k-space data, which limits their clinical translation potential. Data consistency can only be ensured by integrating MoCo into the model- or learning-based image reconstruction process.[19–22] However, previous methods have been developed mainly for higher resolution qualitative MRI or, if developed in the context of T_2^* quantification, they do not enforce data consistency.[15] Furthermore, with respect to motion-induced B_0 inhomogeneity changes, Motyka et al.[23] employ deep learning to predict motion-related B_0 inhomogeneity variations. Yet, to the best of our knowledge, no learning-based MoCo method exists that incorporates information on these B_0 variations into the reconstruction.

We have recently introduced PHIMO, a **PH**ysics-**I**nformed **M**otion c**O**rrection technique for low-resolution GRE-based T_2^* mapping in mqBOLD applications,[24] which implicitly utilizes the above-described B_0 inhomogeneity-induced signal loss to detect motion events. In particular, PHIMO employs a physics-informed loss function to detect motion-corrupted k-space lines and exclude them from a data consistent reconstruction. This physics-informed loss has proven promising in detecting motion events, but the final image quality strongly depends on the performance of the reconstruction network, which is challenged by specific motion patterns, i.e., when central k-space lines need to be excluded. Additionally, varying strengths of B_0 inhomogeneities across the brain make a consistent estimation of exclusion masks challenging.

In this work, we leverage MR acquisition physics to enhance the robustness of PHIMO with two major extensions. We validate the underlying assumptions of these extensions and the physics-informed loss (i.e., its correlation with the severity of motion) and perform comprehensive evaluations. Our contributions are three-fold:

1. We improve the performance of PHIMO's reconstruction module for scenarios where the subject moves during acquisition of the k-space center.
2. We accelerate the self-supervised line detection and improve PHIMO's robustness to varying strengths of inhomogeneities throughout the brain.
3. We perform extensive evaluations and comparisons with other conventional and learning-based MoCo methods, using simulated and real motion data.

2 | THEORY

2.1 | T_2^* mapping from GRE MRI

The effective transverse relaxation time T_2^* can be determined from a series of GRE images $x = [x_1, \ldots, x_N]$ acquired at increasing echo times TE_n for $n = 1, \ldots, N$. Without motion and without B_0 inhomogeneities, the time evolution of the signal magnitude $s_n = |x_n|$ for an individual voxel can be approximated by[6]:

$$s_n = s_0 \cdot \exp\left(-\frac{\mathrm{TE}_n}{T_2^*}\right) \quad (1)$$

with s_0 representing the signal magnitude at $\mathrm{TE}_n = 0$. T_2^* and s_0 can thus be obtained by voxel-wise least-squares fitting of the multi-echo GRE data.

2.2 | MRI forward model with motion

The multicoil MRI forward model, accounting for motion, involves the sampling mask $\mathbf{S}_t$ (including the line-wise k-space acquisition pattern), the Fourier transform $\mathcal{F}$, the coil sensitivity maps $\mathbf{C}$, and the motion transform $\mathbf{U}_t$. To obtain the motion-affected k-space data $\hat{\mathbf{y}}$, these are applied to the motion-free image x for each time point t[25]:

$$\hat{\mathbf{y}} = \sum_{t=1}^{T} \mathbf{S}_t \mathcal{F} \mathbf{C} \mathbf{U}_t x \quad (2)$$

For rigid-body motion in the presence of field inhomogeneities, $\mathbf{U}_t$ consists of translation and rotation transforms, $\mathbf{T}_t$ and $\mathbf{R}_t$, as well as a phase shift which is induced by the position-dependent in-plane B_0 inhomogeneities ω_t and increases with echo time TE_n[26]:

$$\mathbf{U}_t = e^{-2i\omega_t \mathrm{TE}_n} \mathbf{T}_t \mathbf{R}_t \quad (3)$$

2.3 | Assumptions on motion patterns

Head motion is commonly modeled as rigid-body motion that occurs at random timings, e.g., caused by the patient's discomfort or lack of attention. We assume that the patient is able to stay still for most of the scan with only occasional movements. Thus, it is possible to define an undersampling mask $\mathbf{\Omega}$ that excludes these individual motion events, so that the masked motion-corrupted data closely approximate the masked motion-free data $\mathbf{y}$:

$$\mathbf{\Omega}\hat{\mathbf{y}} \simeq \mathbf{\Omega}\mathbf{y} \quad (4)$$

This allows us to split the MoCo into two subproblems: (1) finding the optimal exclusion mask $\mathbf{\Omega}$ and (2) reconstructing the undersampled k-space data $\mathbf{\Omega}\hat{\mathbf{y}}$. We have recently introduced a solution to these two subproblems,[24] which we briefly describe together with the current major improvements in Section 3.1.

If the subject moves, during the acquisition of the k-space center, Equation 4 leads to the exclusion of central k-space lines, which contains important information about the overall image contrast. Completely disregarding such lines severely hampers the correct reconstruction of image intensities, and thus falsifies the subsequent T_2^* quantification, depending of the extent of excluded data. To minimize the impact of excluding central k-space lines, we propose to utilize the fact that the central k-space point $\mathbf{y}_{0,0}$ corresponds to the zero spatial frequency, representing the mean image intensity: $\mathbf{y}_{0,0} = \bar{x}$. Under the assumption that the mean image intensity does not significantly change for small amounts of motion ($\bar{\hat{x}} \simeq \bar{x}$), we set the central entry of the exclusion mask, $\mathbf{\Omega}_{\mathbf{0,0}}$, to 1, independent of the central line's motion status. In practice, for an even matrix size, we have to set the smallest possible group of pixels (i.e., the four pixels around the k-space center) in the exclusion mask to 1.

3 | METHODS

3.1 | PHIMO

We have previously introduced a physics-informed MoCo method, which builds on Eq. 4 and splits the MoCo into (1) a physics-informed detection of motion-corrupted k-space lines and (2) a recon77struction of undersampled data.[24,27] Figure 1 illustrates these steps. In the following, we provide a short overview of PHIMO and elaborate on the refinements which we introduce in this extension and which we refer to as "PHIMO+" for the remainder of this article.

3.1.1 | Self-supervised k-space line detection

For detecting affected k-space lines in *motion-corrupted* data, PHIMO freezes the weights of the pretrained reconstruction model (Sec. 3.1.2) and optimizes for the best exclusion mask with a self-supervised physics loss[24] (Figure 1A and B). The physics loss is calculated as the empirical correlation coefficient between reconstructed signal intensities s_n^{rec} and mono-exponentially "fitted" intensities s_n^{fit} (Eq. 1):

$$L_{phys} = 1 - \frac{\sum_n (s_n^{rec} - \bar{s}^{rec})(s_n^{fit} - \bar{s}^{fit})}{\sqrt{\sum_n (s_n^{rec} - \bar{s}^{rec})^2 (s_n^{fit} - \bar{s}^{fit})^2}}, \quad (5)$$

with the mean over all echoes $\bar{s} = \frac{1}{n}\sum_n s_n$. L_{phys} makes use of motion-related B_0 inhomogeneity changes, which disturb the mono-exponential signal evolution.[8] L_{phys} is calculated voxel-wise across echoes and averaged within the brain mask.

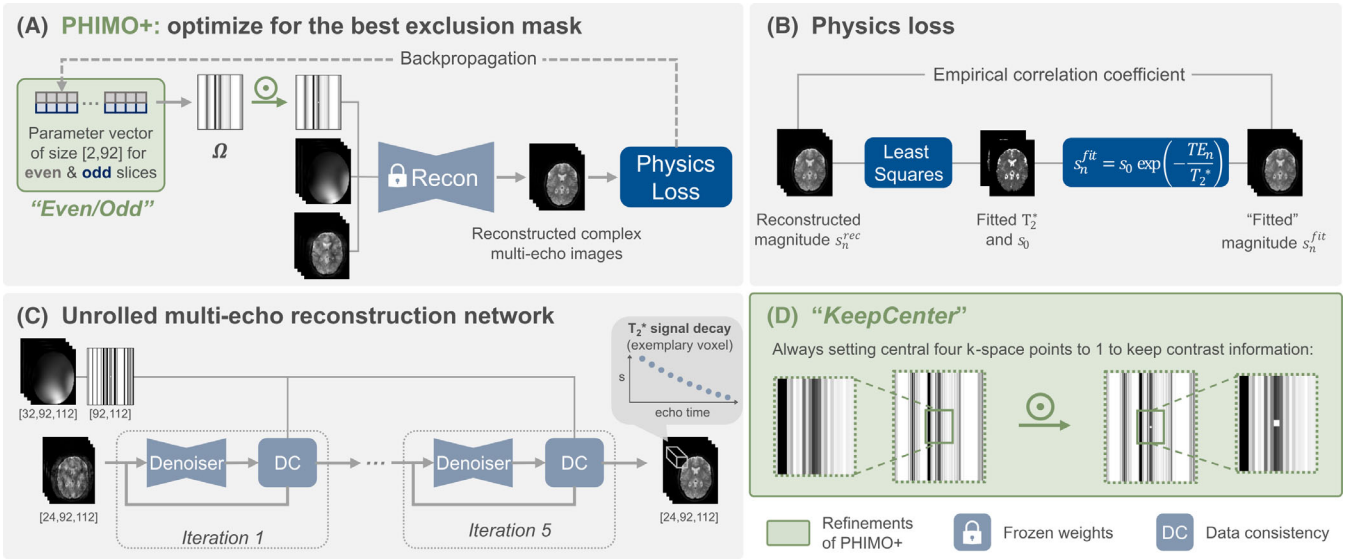

FIGURE 1 Illustration of PHIMO+ with the proposed refinements highlighted on green background. **(A)** Subject-specific line detection for *motion-corrupted* data: optimizing one exclusion mask, Ω, for even and odd slices, respectively - by minimizing the physics-informed loss illustrated in (B). **(B)** Empirical correlation coefficient as physics-informed loss for the self-supervised optimization of the exclusion masks in (A). **(C)** Training an unrolled reconstruction network with randomly undersampled *motion-free* multi-echo data. The pretrained weights of the reconstruction network are frozen for the optimization of the exclusion masks in (A). **(D)** *KeepCenter* extension: always keep the central four k-space points regardless of their motion status to avoid severe contrast loss.

Refinements of PHIMO+

To accelerate the self-supervised line detection and reduce reconstruction times, we directly optimize a parameter vector, representing each slice's exclusion mask, instead of using a multilayer perceptron for predicting the exclusion mask.[24] Furthermore, we incorporate knowledge of the interleaved MRI acquisition scheme and optimize only one exclusion mask for even and odd slices, respectively, (Figure 1A), instead of individual masks for all slices. These *Even/Odd* masks are optimized on the most inferior eight slices with mean susceptibility gradients $< 80\ \mu$T/m. We choose this criterion to select a consistent subset of slices with similarly moderate susceptibility gradients, enabling a robust selection of hyperparameters that remains effective across different subjects of our data set, without the need for fine-tuning. The resulting masks are then applied to all slices during inference.

The output of PHIMO+ is a mask with continuous values between 0 and 1 whose size corresponds to the number of phase encoding (PE) lines. We do not binarize the output to benefit from smoother loss minimization. The masks are directly multiplied with the motion-corrupted k-space, and thus, down-weight contributions from motion-affected lines. To achieve a stable optimization, we include a regularization term L_{reg} with weight λ_{reg} to enforce a minimal number of excluded lines:

$$L_{reg} = \lambda_{reg}\left(1 - \frac{1}{Y}\sum_{y=1}^{Y}\Omega_y\right), \tag{6}$$

with Y representing the number of PE lines. In order to particularly penalize the exclusion of central k-space, we add a second term with stronger weight λ_{center} on the central ten k-space lines:

$$L_{reg} = \lambda_{reg}\left[\left(1 - \frac{1}{Y}\sum_{y=1}^{Y}\Omega_y\right) + \lambda_{center}\cdot\left(1 - \frac{1}{10}\sum_{y=Y_1}^{Y_2}\Omega_y\right)\right], \tag{7}$$

with $Y_1 = \frac{Y}{2} - 5$ and $Y_2 = \frac{Y}{2} + 5$.

Training details

For PHIMO+, a parameter vector of size 184 is optimized for 100 epochs with Adam,[28] a learning rate of 0.01, a batch of 20 slices and the regularization term is weighted with $\lambda_{reg} = 0.005$ and $\lambda_{center} = 2$. Hyperparameters were empirically derived on the validation set (see Section 3.2), focusing on general adjustments rather than an exhaustive optimization. Differentiable least-squares fitting is implemented with `torch.linalg.lstsq`. The optimization takes less than three minutes per subject on a NVIDIA RTX A6000.

3.1.2 | Unrolled multi-echo reconstruction

For the reconstruction task, we utilize an unrolled network (Figure 1C), which is trained on randomly undersampled *motion-free* data, utilizing a mean squared error loss. The reconstruction alternates five times between a CNN-based

denoiser and gradient descent data consistency layer, with independent weights for each iteration. Real and imaginary components of the input image and multiple echoes are stacked in the channel dimension. For more implementation details, we refer to our previous publication.[24]

Refinements of PHIMO+

As outlined in Section 2.3, we adapt our previous implementation to improve the reconstruction performance in scenarios where the subject moves during the acquisition of the k-space center. For this *KeepCenter* setting, we set the four central pixels in the undersampling mask to 1, already during training of the reconstruction network (Figure 1D). For the comparison experiments (see Section 3.4), we train multiple networks with differently generated masks, e.g., with and without the *KeepCenter* setting.

3.2 | Data acquisition

We have acquired multi-coil k-space data from 19 volunteers (27.0 ± 2.8 years, 7 females) on a 3T Philips Elition X MR scanner (Philips Healthcare, Best, The Netherlands). The acquisition consists of an interleaved multi-slice 2D GRE sequence that acquires even and odd slices in two packages (12 echoes, 36 slices, 92 PE lines, $\mathrm{TE}_1 = \Delta\mathrm{TE} =$ 5 ms, TR = 2300 ms, voxel size: $2 \times 2 \times 3$ mm^3, field of view: $224 \times 184 \times 118.5$ mm^3, 32-channel head coil). The acquisition takes 3 min 39 s. The study has been approved by the local ethics committee (approval numbers 440/18 S-AS, 2023-386-S-SB).

We have performed repeated scans under two conditions, instructing the subject to (1) remain still and (2) to move occasionally in random directions (e.g., imitating sneezing or coughing). To evaluate PHIMO's motion detection performance, we have provided five subjects with precise timings on when to move (three to six times for five to ten seconds[1]). We refer to this as *motion timing experiment*. To explore different motion types, we instructed one subject ("*MoCo validation subject*") to perform four motion patterns with three different amplitudes each. Moreover, we have collected half- and quarter-resolution data for all subjects under all motion conditions to compare the performance of PHIMO+ with the state-of-the-art MoCo method.[8]

Preprocessing

The data are divided subject-wise into train/validation sets for reconstruction (6/3 subjects) and validation/test sets for MoCo (1/9 subjects), including only slices with more than 20% brain voxels. For input to the reconstruction network, data are normalized per 3D volume based on the maximum image magnitude.

3.3 | Motion simulation

To expand our evaluations and to train a supervised baseline method (Section 3.4.2), we simulate head motion in motion-free multi-coil images. The simulation of rigid-body motion and random B_0 inhomogeneity variations follows our previously presented physics-aware simulation pipeline.[26] The simulations are based on real head motion, i.e., 82 3D motion curves extracted from in-house fMRI time series.[2] These curves are divided into training, validation, and test sets (40/15/27 curves).

Principal component analysis is used to determine the largest modes of variations of the training motion curves. The largest 20% of all N principal components, $pc_i(t)$, are combined with random weights α_i and added to the mean curve $\overline{T(t)}$ to generate 90 augmented training samples: [29]

$$T_{train}(t) = \overline{T(t)} + \sum_{i=1}^{0.2 \cdot N} \alpha_i \cdot pc_i(t). \tag{8}$$

Implementation details

We simulate motion only for PE lines, where the average displacement of points within a sphere with radius 64 mm-modelling the patient's head-exceeds a threshold of 2 mm, the voxel size of our data. To circumvent the requirement of registration when calculating full-reference metrics of the simulated images relative to the original motion-free image, the motion curves are shifted so that the median motion state-measured by average displacement-is positioned at zero.

3.4 | Experiments

3.4.1 | Validation of assumptions

PHIMO+ relies on assumptions for the reconstruction and line detection steps, including the proposed *KeepCenter* and *Even/Odd* extensions. To ensure the validity of these assumptions, we conduct experiments using the MoCo validation subject, i.e., the subject performing four different motion patterns with three different amplitudes. Regarding the reconstruction choices, we test if the central four k-space points are affected by motion (Section 2.3). For this, we compare reconstructions of only the central four k-space points from acquisitions without and with

intentional motion. Additionally, we investigate the impact of *KeepCenter* on the reconstruction quality when excluding several central k-space lines. To assess this, we compare the mean absolute error (MAE) of T_2^* maps obtained from networks trained without and with *KeepCenter*.

For the line detection choices, we first validate that the proposed physics-informed loss function correlates with the degree of motion that affects the GRE data. For this, we compare the loss value for motion-free images and three levels of intentional motion. Furthermore, we evaluate the effects of the *Even/Odd* extension on the detection robustness across the brain by analyzing the line detection accuracy relative to the reference mask for different slice numbers. Lastly, we investigate if the line detection performance depends on the line's specific k-space location. For this, we compare the detection accuracy across all slices and motion experiments for three k-space regions: low frequencies (30 central k-space lines), high frequencies (32 peripheral lines, 16 on each side), and medium frequencies (30 lines in between).

3.4.2 | Comparative evaluation on test set

We compare the proposed refinements of PHIMO+ to two learning-based alternatives, outlier-rejecting bootstrap aggregation (ORBA)[18] and supervised line detection (SLD),[22,26] as well as to the previous version, PHIMO,[24] and the state-of-the-art MoCo for T_2^* quantification from GRE MRI with redundant k-space acquisition (HR/QR).[8]

ORBA[18]
Approaches motion as a probabilistic undersampling problem and averages reconstructions with 15 random bootstrap masks that each randomly exclude some lines as potential motion events. For a fair comparison, we implement ORBA by training a separate unrolled reconstruction network with variable density masks at a fixed exclusion rate of 0.5, utilizing the *KeepCenter* setting.

SLD[22,26]
Utilizes simulated data for training a convolutional neural network to predict exclusion masks for a given motion-corrupted k-space in a supervised fashion. The training of the line detection network follows the details in our previous publication,[26] but we reconstruct the resulting undersampled data with the same reconstruction network as used for PHIMO+.

PHIMO[24]
Does not use the *KeepCenter* setting during training and inference of the reconstruction network and optimizes an individual exclusion mask for every slice. For a robust optimization, PHIMO requires a second regularization term that enforces a small variation of the masks for adjacent slices. For more details, refer to Eichhorn et al.[24]

HR/QR[8]
Leverages additional half- and quarter-resolution acquisitions and calculates a weighted average of the three/two acquisitions of each PE line in the central half/quarter of k-space, in order to suppress individual motion events in one of the acquisitions. The reliance on redundant acquisition of the k-space center prolongs the acquisition time from 3 min 39 s to 6 min 25 s.

3.5 | Evaluation metrics

3.5.1 | Line detection

We evaluate the line detection performance of PHIMO+ by calculating the MAE between predicted and reference masks, and the accuracy, quantified by the fraction of correctly classified lines when thresholding the predictions at a level 0.5. Additionally, we analyze the precision-recall-curve for various thresholds. Precision measures the fraction of lines that are correctly predicted as motion-corrupted, relative to the total number of lines predicted as motion-corrupted. Recall measures the fraction of lines that are correctly predicted as motion-corrupted, relative to the total number of actually motion-corrupted lines. For simulated data, reference masks are derived from the time points where motion was simulated (Section 3.3). For real motion data, reference masks are only available for the motion timing experiment, where the motion timings are converted into masks using the known k-space acquisition scheme. Note that these reference masks are subject to some uncertainty, since they are generated based on the instructed motion timings and do not incorporate any measurement of the actual motion amplitude or timing deviations.

3.5.2 | Image quality

We quantitatively evaluate the quality of the T_2^* maps based on MAE, structural similarity (SSIM), feature similarity (FSIM), and perceptual image patch similarity (LPIPS), which have been shown to correlate with radiological evaluation of image quality.[30,31] All metrics are calculated per slice relative to the motion-free T_2^* map for brain tissue voxels with susceptibility gradients smaller than 100 μT/m.

In the case of real motion (Section 4.3), we align all acquisitions to the motion-free acquisition via 3D

registration of the stacked slices using SPM12[32] `coregister`. To differentiate between the methods' performances for more severe and minor motion, we categorize the nine test subjects based on visual assessment of the level of motion artifacts in the motion-corrupted GRE images: data where motion artifacts were clearly visible in earlier echoes are categorized as "severe" and data where artifacts were only visibly for later echoes as "minor." This results in five subjects with severe and four subjects with minor motion.

3.5.3 | Statistical analysis and implementation

We use Wilcoxon signed rank tests and False-Discovery Rate correction for statistical testing. HR/QR motion correction and segmentation of anatomical scans are performed in MATLAB (R2022b) and SPM12 with custom programs.[33] All other computations are performed in Python 3.8.12, using PyTorch 2.0.1 and MERLIN.[34] Our code is publicly available at https://github.com/compai-lab/2025-mrm-eichhorn.

4 | RESULTS

4.1 | Validation of assumptions

We investigate the assumption that the proposed *KeepCenter* setting improves PHIMO's performance for cases where the subject moved during the acquisition of the k-space center. Our analysis is based on the motion-free acquisition of the MoCo validation subject. The exemplary reconstructed images from only four central k-space points, resulting from data with and without intentional motion, exhibit some differences in details but their overall intensity distributions are comparable (Figure 2A). Accordingly, for unrolled network reconstructions with masks that exclude several central k-space lines, the *KeepCenter* extension reduces the MAE of the corresponding T_2^* maps, compared to reconstructions with the original full masks (Figure 2B). Reconstructions for both settings are provided in the Supporting Information (Figure S1). These examples illustrate that the *KeepCenter* extension avoids severe signal loss and recovers the contrast more correctly.

We further test if our previously introduced physics loss[24] reflects the degree of motion-corruption in the GRE images, using the repeated acquisitions of the MoCo validation subject. As illustrated in Figure 2C, the value of the loss increases from motion-free images over low and medium to high motion amplitudes. Additionally, the larger variance of loss values for all slices compared to selected slices (based on susceptibility gradient strengths) confirms our assumption that the high variability of susceptibility strengths across the brain makes a robust optimization challenging. Moreover, Figure 2D demonstrates that the *Even/Odd* extension, which is optimized for selected slices only, results in a robust line detection accuracy across slices, with only small differences between even and odd slices. In contrast, optimizing on individual slices leads to lower accuracies for inferior and superior slices. This quantitative observation is also obvious in the example exclusion masks for different

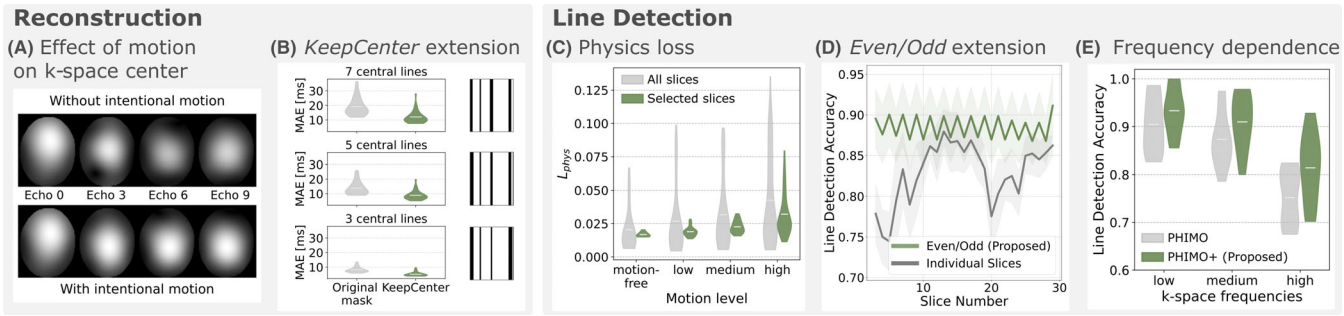

FIGURE 2 Investigating assumptions regarding reconstruction and line detection on the MoCo validation subject. (A) Exemplary reconstructions of only four central k-space points for acquisitions without and with intentional motion (different echoes). (B) MAE of T_2^* maps resulting from the *KeepCenter* and the original mask reconstructions for masks that exclude seven, five, and three central k-space lines. Examples of reconstructed images can be found in the Supporting Information (Figure S1). (C) Values of the physics loss for the motion-free acquisition compared to the acquisitions with low, medium, and high motion amplitudes. The loss is once shown for all slices (gray) and once for only the inferior eight slices with mean susceptibility gradients < 80 μT/m (green), as we propose with the *Even/Odd* extension. (D) Slice-wise average line detection accuracy (relative to reference mask), comparing the *Even/Odd* setting (green) to optimizing each slice individually (gray). Example exclusion masks are provided in the Supporting Information (Figure S2). (E) Line detection accuracy of PHIMO and PHIMO+ for distinct k-space regions: low frequencies (30 central k-space lines), medium frequencies (30 lines), and high frequencies (32 peripheral lines).

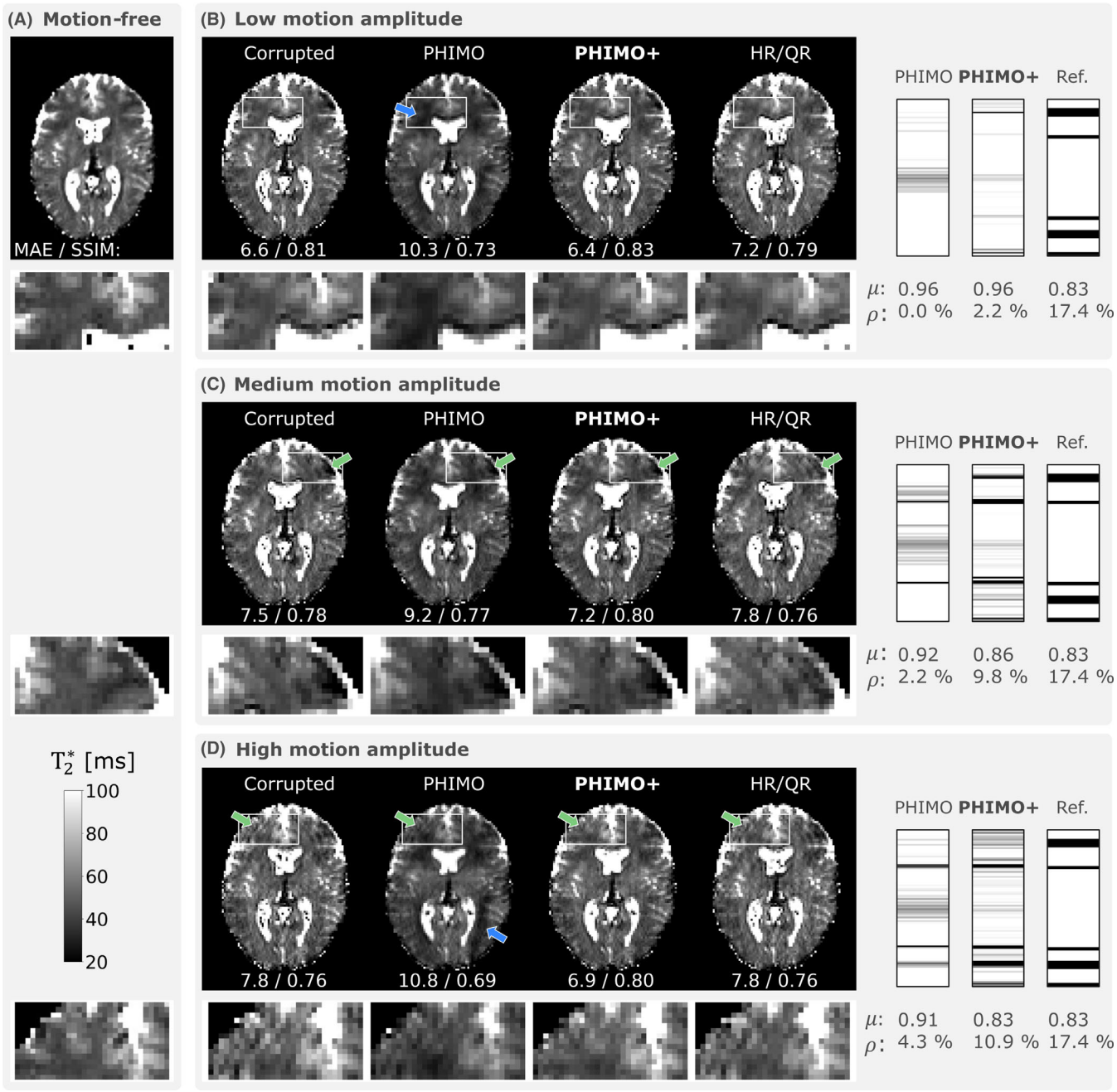

FIGURE 3 Testing the adaption-ability of PHIMO+ to different motion levels for the MoCo validation subject. The resulting T_2^* map from PHIMO+ is compared to the uncorrected, motion-corrupted data, PHIMO, HR/QR, and the separately acquired motion-free data (A). The corresponding exclusion masks and the reference mask are presented on the right. The three rows show different acquisitions with the same prescribed motion pattern but different motion amplitudes: low (B), medium (C), and high (D). Green arrows indicate motion artifacts that are best suppressed by PHIMO+, blue arrows indicate visible T_2^* estimation errors for PHIMO. MAE (in ms) and SSIM values for the visualized slices relative to the motion-free T_2^* map are provided below each image. Average mask values (μ) and fraction of excluded lines (ρ) are provided below each mask.

slices in the Supporting Information (Figure S2), which illustrates that the *Even/Odd* extension leads to a more consistent line detection throughout the brain. Lastly, Figure 2E showcases PHIMO+'s robust line detection performance across different k-space regions. PHIMO+ outperforms PHIMO throughout the entire k-space, with only a slight decrease in detection performance from low to high k-space frequencies.

Additionally, we use the repeated acquisitions of the MoCo validation subject to compare PHIMO+ to PHIMO and test if PHIMO+ is able to adapt the predicted masks to varying levels of motion corruption. Figure 3 shows

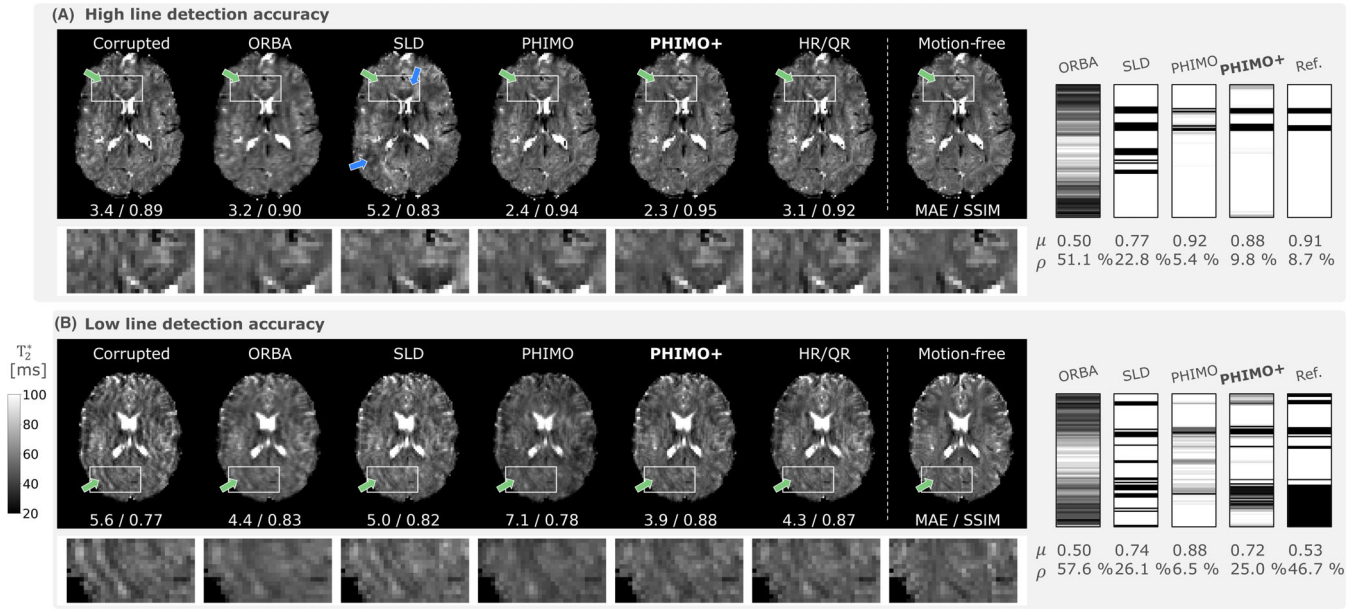

FIGURE 4 Qualitative examples for test subjects with simulated motion. The T_2^* map resulting from PHIMO+ is compared to the uncorrected, motion-corrupted data (left), ORBA, SLD, PHIMO, HR/QR, and the separately acquired motion-free data (right). The corresponding exclusion masks and the reference mask are additionally provided at the right. The top row (A) shows data from one of the subjects with the highest line detection accuracies (0.989) and the bottom row (B) data from one of the subjects with the lowest accuracies for PHIMO+ (0.739). Green arrows indicate severe motion artifacts that are reduced by PHIMO+ and partially by PHIMO and HR/QR. Blue arrows indicate artifacts introduced by SLD. MAE (in ms) and SSIM values for the visualized slices relative to the motion-free T_2^* map are provided below each image. Average mask values (μ) and fraction of excluded lines (ρ) are provided below each mask.

examples for three repeated acquisitions with different motion amplitudes but approximately the same pattern. The mask examples validate that PHIMO+ is able to adapt the predicted masks to different motion levels. PHIMO+'s predictions match the reference more closely than PHIMO. Furthermore, PHIMO+ recovers the T_2^* values more accurately than PHIMO and outperforms HR/QR in removing ringing artifacts for this motion pattern.

4.2 | Evaluation on simulated test data

We compare example T_2^* maps from different reconstructions of data without and with simulated motion (Figure 4). PHIMO+ significantly reduces motion artifacts which mostly appear as ringing or wave-like patterns. The top row shows an example where the exclusion mask estimated by PHIMO+ closely resembles the prescribed reference. In the bottom row, where PHIMO+ does not detect all lines, the corresponding exclusion mask yet outperforms the previous version, PHIMO, and generally matches the pattern of the reference mask, which confirms that PHIMO+ can detect the dominant motion events. HR/QR also reduces severe motion artifacts, but shows more residual wave-like artifacts than PHIMO+, particularly for the subject in the top row. The T_2^* maps resulting from ORBA show an increased blurring but no clear reduction of the wave-like artifacts. SLD partially reduces the wave-like motion artifacts but introduces additional artifacts in some regions. The exclusion masks from SLD approximate the patterns of the reference masks but particularly overestimate motion events in the k-space center, especially for the subject in the top row, and miss more lines than PHIMO+ for the subject in the bottom row.

To analyze PHIMO+'s performance quantitatively, we compare PHIMO+'s line detection results to PHIMO, ORBA, and SLD on the test data with simulated motion in Figure 5. PHIMO+ outperforms all comparison methods in terms of MAE and accuracy as well as in the precision-recall-curve. Similarly, the quantitative evaluation of the resulting T_2^* maps in Figure 6 demonstrates that our proposed extension, PHIMO+, outperforms PHIMO as well as the learning-based baselines, ORBA and SLD, and approaches the performance of HR/QR.

4.3 | Evaluation on real motion test data

We show example T_2^* maps from test subjects with real motion in Figure 7. For the subject exhibiting strong

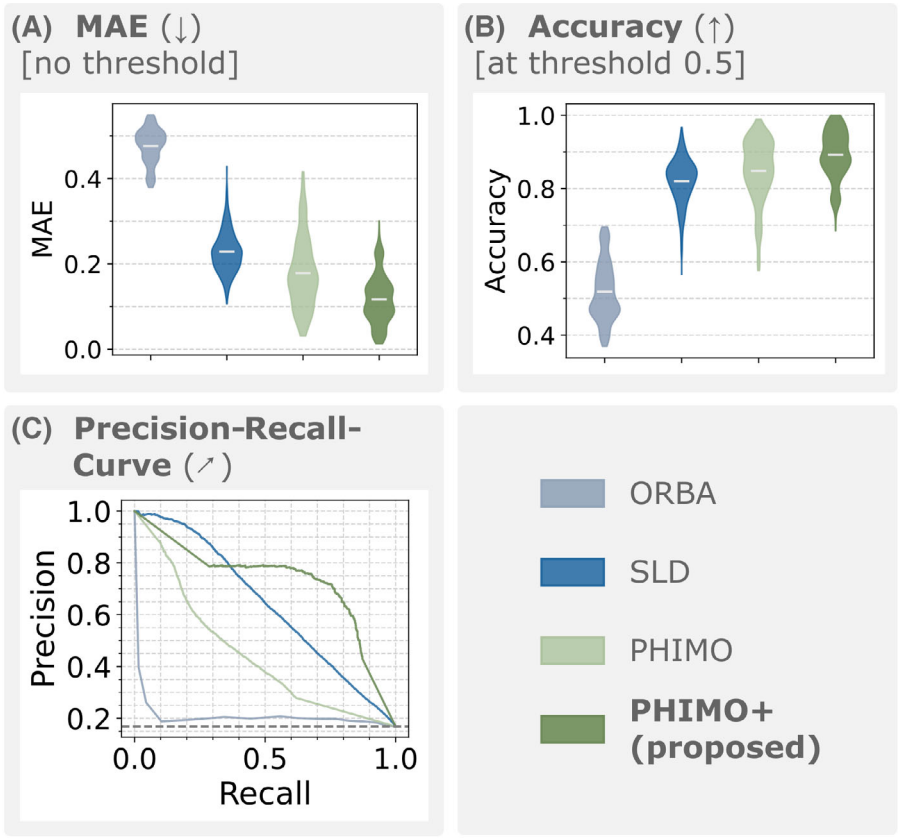

FIGURE 5 Line detection results for simulated test data. (A) MAE, (B) accuracy, and (C) precision-recall-curve for all PE lines aggregated. All comparisons in (A) and (B) are statistically significant ($p<0.001$).

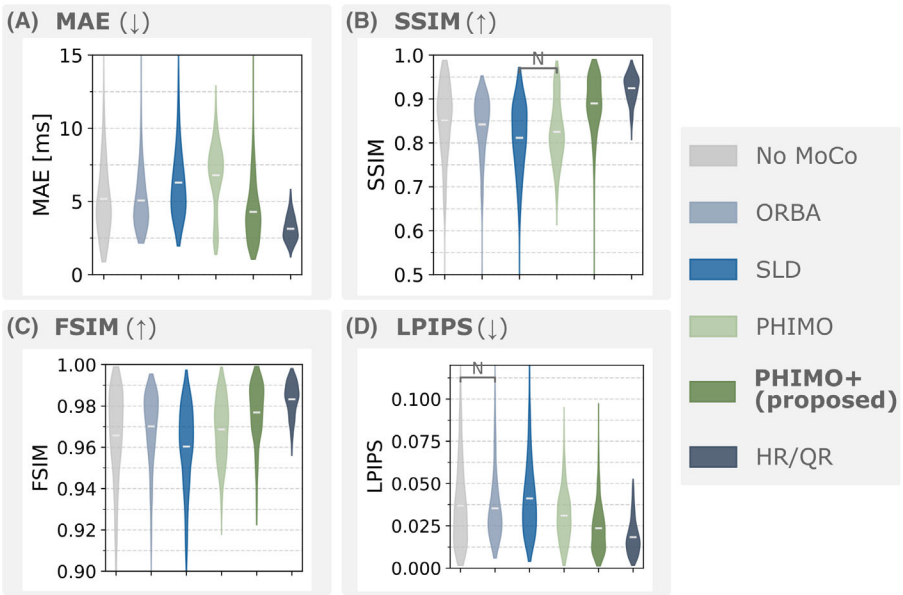

FIGURE 6 Quantitative evaluation of T_2^* maps for test data with simulated motion. Image quality metrics (A) MAE, (B) SSIM, (C) FSIM, and (D) LPIPS are calculated relative to the motion-free T_2^* map. The proposed PHIMO+ improves compared to the uncorrected T_2^* map and outperforms PHIMO as well as the learning-based baselines. Gray brackets indicate comparisons with no statistical significance ($p>0.001$). Outlier values (ranging up to 62.9/0.18 for MAE/LPIPS and down to 0.13/0.85 for SSIM/FSIM) were cut off to better visualize the differences of the mean values.

motion, PHIMO+ outperforms all comparison methods and effectively suppresses severe motion artifacts that otherwise cause regional T_2^* misestimation errors. However, similar to all other methods, PHIMO+ is unable to fully recover all fine details compared to the separately acquired motion-free T_2^* map for this level of head motion. The estimated exclusion mask matches the pattern of the reference mask (extracted from the prescribed timings), which confirms PHIMO+'s ability to detect real motion events. Similar to the simulated motion examples, SLD somewhat overestimates the exclusion mask, particularly in the k-space center. For the subject exhibiting minor motion, the corrupted T_2^* map shows only subtle motion artifacts and all methods either preserve its quality or lead

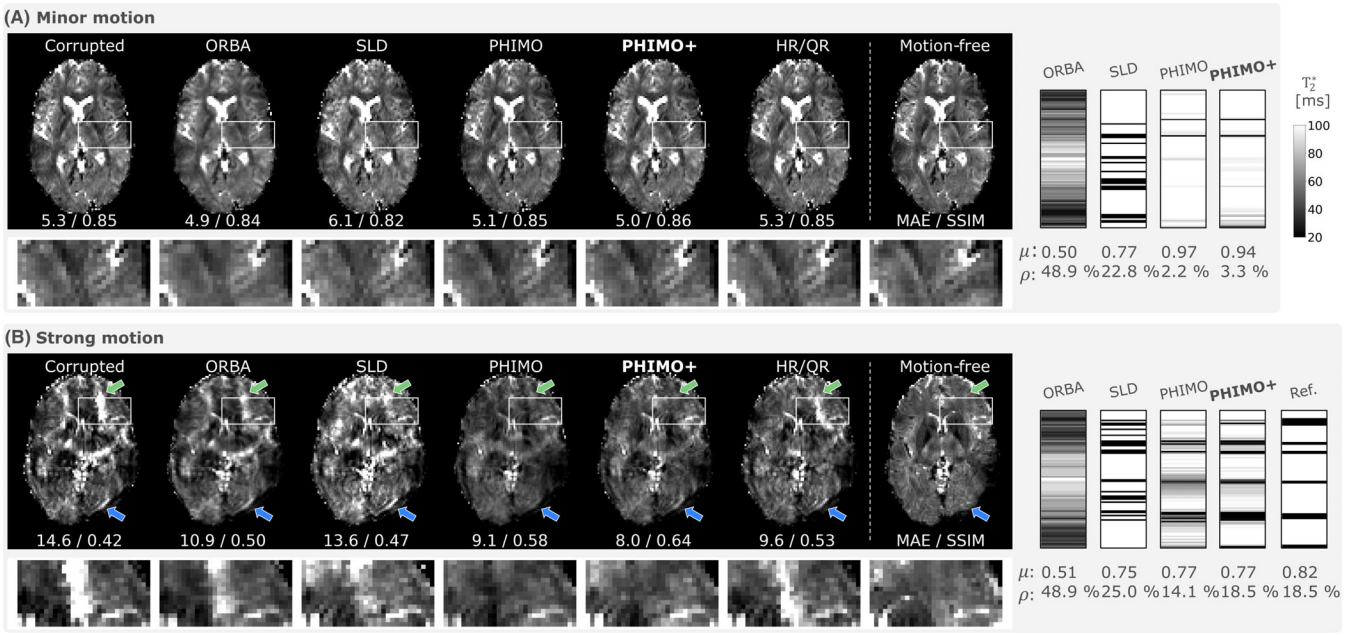

FIGURE 7 Qualitative examples for test data with real motion. The resulting T_2^* map from PHIMO+ is compared to the uncorrected data (left), ORBA, SLD, PHIMO, HR/QR, and the separately acquired motion-free data (right). We additionally provide the corresponding exclusion masks and the reference mask (if available). The top row (A) shows one subject with minor motion (no instructions on timing given, thus, no reference mask available) and the bottom row (B) one subject with strong motion (motion timing experiment). Green arrows indicate severe motion artifacts that are suppressed by PHIMO, PHIMO+, and partially by HR/QR, blue arrows indicate subtle residual artefacts for all methods. MAE (in ms) and SSIM values for the visualized slices relative to the motion-free T_2^* map are provided below each image. Average mask values (μ) and fraction of excluded lines (ρ) are provided below each mask.

to minor improvements in MAE and SSIM that are difficult to detect visually. For this subject no reference mask is available, which precludes a direct evaluation of the predicted exclusion masks. However, PHIMO+ and PHIMO detect only a few motion-corrupted lines, as expected for such minor motion artifacts, while SLD excludes a large number of lines, especially in the k-space center.

Figure 8 quantitatively compares the line detection performance of PHIMO+ on real motion data for the four test subjects that performed the motion timing experiment. In line with the qualitative observations from Figure 7, PHIMO+ shows superior performance compared to all other methods based on MAE, accuracy, and the precision-recall-curve.

The quantitative evaluation of the T_2^* maps based on different image quality metrics in Figure 9 supports our qualitative observations. For the subjects with strong motion, PHIMO+ improves all image quality metrics compared to the uncorrected maps and the learning-based baseline methods. PHIMO+ outperforms the previous version, PHIMO, in terms of MAE and SSIM and approaches the performance of HR/QR. Minor motion data generally show better metric values with small but consistent improvements by PHIMO+ and HR/QR. SLD, and to some extent PHIMO, rather degrades image quality for minor motion, particularly in terms of MAE. A potential reason for this could be that they overestimate the amount of motion in the k-space center. For the subjects with minor motion, the average value of the exclusion mask for the central 10 lines is 0.95 for PHIMO+, 0.84 for PHIMO, and 0.65 for SLD.

We additionally evaluate how PHIMO+ and the comparison methods deal with acquisitions without intentional motion. Example T_2^* maps for their applications to apparently motion-free data are provided in the Supporting Information (Figure S3). In these examples, PHIMO+, PHIMO, and HR/QR preserve the quality of the motion-free data, whereas ORBA introduces blurring and SLD excludes many central k-space lines, partially resulting in severe artifacts. This observation is also reflected in the summary of predicted exclusion masks by ORBA, SLD, PHIMO, and PHIMO+ for the motion-free data in Table 1. Assuming that the cooperative volunteers remained still for the majority of the acquisitions, PHIMO+'s predicted exclusion masks, with an average of 1.3% excluded lines, appear more realistic compared to SLD's average of 13.3%.

5 | DISCUSSION

We have proposed two major extensions to our previously introduced self-supervised and physics-informed

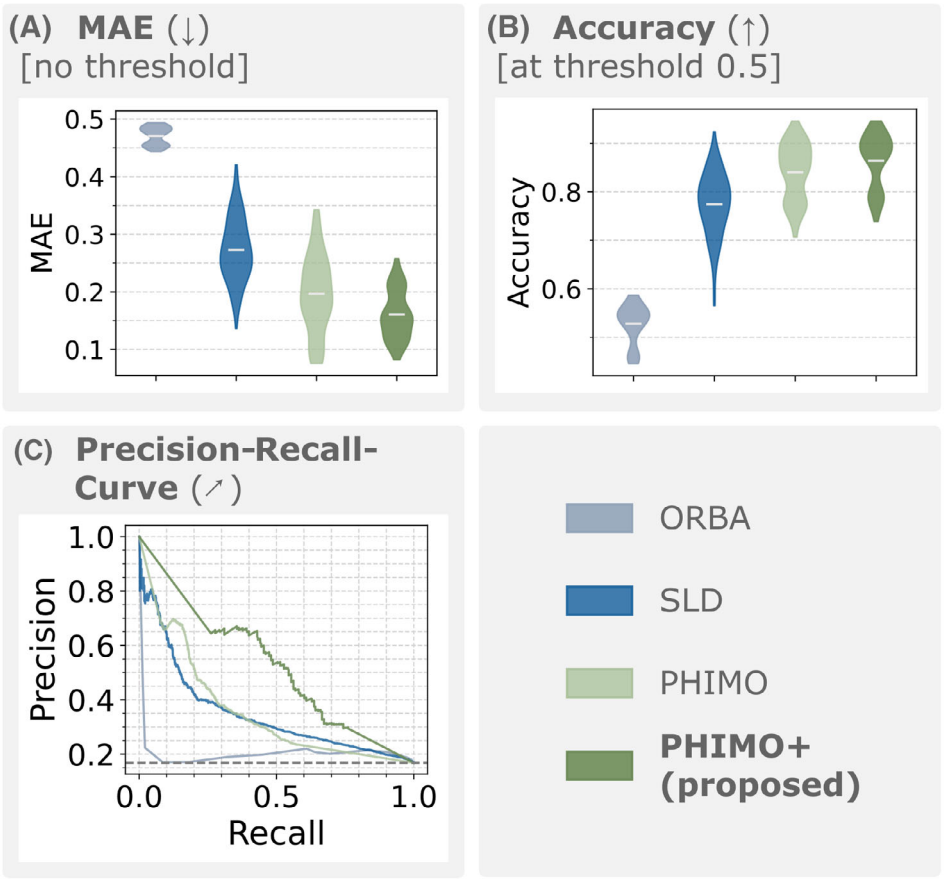

FIGURE 8 Line detection performance on real motion test data for the four subjects of the motion timing experiment. (A) MAE, (B) accuracy, and (C) precision-recall-curve. Reference masks were generated from the prescribed motion timings using the known k-space acquisition scheme. All comparisons in (A) and (B) are statistically significant ($p<0.001$).

learning-based MoCo method.[24] This extended version, PHIMO+, improves the robustness of the self-supervised line detection by optimizing only one mask for even and odd slices, respectively. Additionally, the *KeepCenter* extension improves the undersampled reconstruction and reduces T_2^* quantification errors for challenging motion patterns affecting the k-space center. We have validated these extensions and their underlying assumptions, performing comprehensive validation experiments. Moreover, our comprehensive evaluations using simulated and real motion data have confirmed the superiority of PHIMO+ to learning-based baseline methods. Our evaluations have further confirmed that PHIMO+ performs on par with the state-of-the-art HR/QR MoCo method for T_2^* quantification from GRE MRI,[8] while substantially reducing the acquisition time, since PHIMO+ does not rely on redundant acquisitions of the k-space center.

5.1 | PHIMO+ vs. PHIMO

Our validation experiments have supported our assumption that the reconstruction benefits from keeping the central four k-space points regardless of their motion status (Figure 2). The information in the k-space center appears to not be severely affected by motion but aids the correct contrast reconstruction and subsequent T_2^* quantification. Furthermore, we have demonstrated that the proposed physics loss-based on the T_2^* decay fit-is sensitive to the level of motion-corruption. Moreover, the loss values are more tightly clustered when calculated on a selected subset of slices (based on susceptibility gradient strengths), which confirms our assumption that optimizing the exclusion mask only on a subset of slices allows for a more robust optimization. Together with our *Even/Odd* extension, which groups slices that are acquired within a sufficiently narrow time frame, we achieve a more consistent detection performance throughout the brain. Moreover, the visual and quantitative comparisons of PHIMO+ and PHIMO[24] demonstrate that the *KeepCenter* and *Even/Odd* extensions significantly improve the overall MoCo performance (Figure 3).

5.2 | PHIMO+ vs. learning-based baselines

We have compared PHIMO+ to two learning-based baseline methods, ORBA[18] and SLD.[22,26] These methods are also based on excluding motion-corrupted k-space lines,

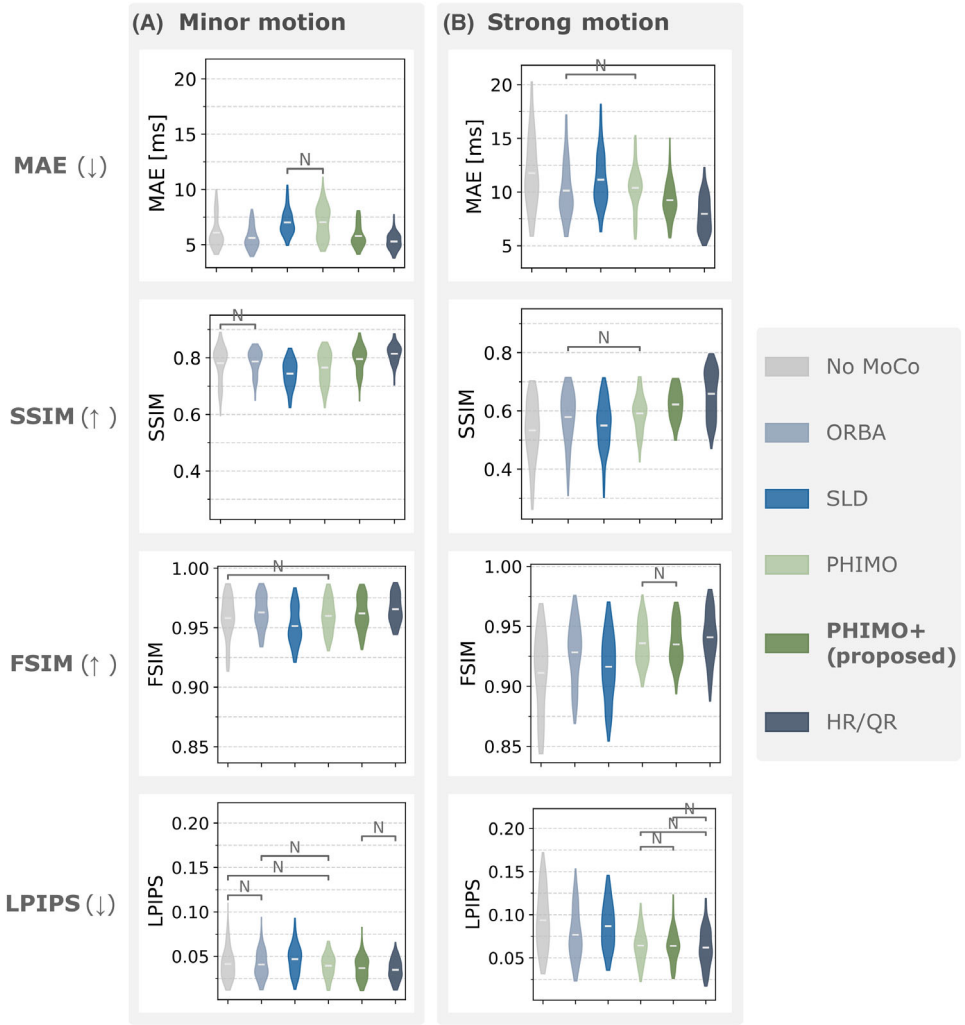

FIGURE 9 Quantitative evaluation of T_2^* maps for real motion test data. Image quality metrics MAE, SSIM, FSIM, and LPIPS relative to the motion-free T_2^* map for four subjects with minor motion (A) and five subjects with strong motion (B). The proposed PHIMO+ improves compared to the uncorrected T_2^* map and outperforms PHIMO as well as the learning-based baselines (ORBA and SLD). Gray brackets indicate comparisons with no statistical significance ($p>0.001$).

TABLE 1 Summary of exclusion masks predicted for motion-free data by ORBA, SLD, PHIMO, and PHIMO+.

	ORBA	SLD	PHIMO	PHIMO+
Average value of masks	0.503	0.867	0.958	0.967
Fraction of excluded lines	52.1%	13.3%	0.8%	1.3%

Note: For calculating the fraction of excluded lines, the masks were thresholded at 0.5.

and thus are applicable to T_2^*-weighted GRE MRI with B_0 inhomogeneities as secondary motion effects. For both simulated and real motion, ORBA results in blurry T_2^* maps and cannot correct the observed motion artifacts, in particular wave-like patterns and ringing artifacts. This is not surprising, since - in contrast to PHIMO+ - ORBA does not specifically exclude motion-corrupted lines but averages the outcomes of several random masks. Our results demonstrate that PHIMO+'s explicit self-supervised line detection is superior to this probabilistic aggregation of random masks.

The second baseline method, SLD, is trained to exclude specific k-space lines in a supervised fashion on simulated data. However, compared to PHIMO+, the predicted masks for unseen data are less accurate across all metrics, especially for real motion. We have observed that SLD tends to exclude too many lines in the k-space center, which particularly challenges the reconstruction and needs to be avoided in the absence of actual motion. In particular for motion-free data, SLD has partially introduced severe artifacts, while PHIMO+ has preserved the high quality of the motion-free maps (Figure S3). This overestimation of motion in the k-space center certainly contributes to SLD's lower image quality metrics, where PHIMO+ outperforms SLD for simulated and real motion data. The superiority of PHIMO+ to SLD confirms the benefit of self-supervised scan-specific optimization, which avoids some generalization issues of supervised training with limited or potentially not fully realistically

simulated training data. Furthermore, the scan-specific design makes PHIMO+ robust across different acquisition settings, improving its generalizability to new datasets, as long as the assumptions of the physics-based loss function remain valid.

5.3 | PHIMO+ vs. state-of-the-art HR/QR

We have additionally compared PHIMO+ to HR/QR, which samples the k-space center three times, combining the motion-corrupted acquisition with additional half- and quarter-resolution acquisitions. Quantitatively, HR/QR achieves slightly higher values for most image quality metrics in simulated and real motion data; but PHIMO+ more closely approaches HR/QR's performance than any other comparison method. HR/QR's superior metric values must be considered in the context that it combines three acquisitions into a weighted average, which inherently improves SNR, even in the absence of motion, at the cost of longer acquisition times. However, when the subject moves during the acquisition of peripheral k-space lines, HR/QR cannot correct the corresponding artifacts, since it only reacquires the k-space center. The examples in Figures 3, 4 and 7 demonstrate PHIMO+'s superiority for these scenarios. Considering quantitative and qualitative results, PHIMO+ performs on par with HR/QR, while significantly reducing the overall scan time by over 40%, which is crucial for clinical translation.

5.4 | Limitations

Our work is not without limitations. First, similar to all comparison methods, PHIMO and PHIMO+ assume that the subject remains mostly still, with only occasional random movements. Our evaluation revealed some variation in line detection accuracy across test subjects (Figures 5 and 8). Beyond the uncertainty of the reference masks themselves, this could indicate that the performance of PHIMO+ depends on the specific motion pattern. Furthermore, in the Supporting Information (Figure S4), we provide an example that was excluded from the main analysis due to exaggerated motion in the k-space center (ten seconds or nine lines), which challenges the undersampled reconstruction even with the *KeepCenter* extension. To improve the resilience to such severe motion patterns, one potential solution could involve randomizing the acquisition order in the PE direction, as proposed in the DISORDER framework,[35] and thus distribute potentially long motion events across the entire k-space. A detailed analysis of how many k-space lines can be excluded from which region without degrading reconstruction quality remains a direction for future work. Second, in the clinical routine, short reconstruction times are essential. PHIMO+ currently requires up to three minutes for the scan-specific optimization of exclusion masks and final reconstruction. While this is already practical in research settings, where reconstructions can be performed offline, additional efficiency enhancements will be crucial to further reduce reconstruction times for clinical applications. Third, our analysis is limited to multi-echo GRE data at 2 mm resolution. While sufficient for noise-robust T_2^* quantification in the context of quantifying the oxygen metabolism with the mqBOLD technique,[3] higher resolution GRE acquisitions are commonly used for instance for quantitative susceptibility mapping.[36] These higher resolution acquisitions have benefited from advances in parallel imaging and learning-based accelerated imaging.[37] The generalizability of PHIMO+ to high-resolution data remains to be tested, as the physics-informed loss leverages contrast information which is more prevalent in the k-space center. However, our validation experiments suggest that PHIMO+'s line detection is relatively robust across k-space frequencies (Figure 2E). Lastly, PHIMO+ does not directly correct motion-related B_0 inhomogeneity changes; rather, it excludes k-space lines that are affected by such motion-induced magnetic field changes, thereby ensuring more consistent B_0 inhomogeneities across the acquisition. However, static B_0 inhomogeneities are still present and need to be addressed in a separate post-processing step.[38] In order to disentangle motion and static B_0 inhomogeneity artifacts in our analysis, we only evaluated the T_2^* maps for voxels with susceptibility gradients smaller than 100 μT/m.

6 | CONCLUSION

In conclusion, we have developed two key extensions to our previously introduced physics-informed motion correction method, PHIMO. With these extensions, we enhance PHIMO+'s robustness to highly variable magnetic susceptibility gradients across the brain and enhance its reconstruction performance for challenging motion patterns, in particular motion events occurring in the k-space center. Our comprehensive evaluation convincingly demonstrates that PHIMO+ outperforms two learning-based baseline methods and matches the performance of a state-of-the-art MoCo technique, while significantly reducing the total acquisition time. This makes PHIMO+ a promising option for motion-robust T_2^* quantification from GRE MRI in clinical and research applications.

ACKNOWLEDGMENTS

Hannah Eichhorn and Veronika Spieker are partially supported by the Helmholtz Association under the joint research school "Munich School for Data Science - MUDS." This work is supported by the Munich Center for Machine Learning and the DAAD programme Konrad Zuse Schools of Excellence in Artificial Intelligence, both sponsored by the Federal Ministry of Research, Technology, and Space. Open Access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTEREST STATEMENT

Kilian Weiss is an employee of Philips GmbH Market DACH.

DATA AVAILABILITY STATEMENT

Our code is publicly available at https://github.com/compai-lab/2025-mrm-eichhorn. An anonymized version of the dataset is publicly available at https://doi.org/10.15134/2kek-3553.

ENDNOTES

[1]One of the subjects moved 30 s in the beginning of the scan, since they missed the "stop" instruction.

[2]Unpublished data from two ongoing studies with approval numbers 472/16 S and 15/20 S, independent cohorts from above imaging cohorts.

ORCID

Hannah Eichhorn https://orcid.org/0000-0001-6980-9703
Veronika Spieker https://orcid.org/0000-0001-7720-7569
Kerstin Hammernik https://orcid.org/0000-0002-2734-1409
Elisa Saks https://orcid.org/0009-0002-2766-8054
Lina Felsner https://orcid.org/0000-0001-7695-2612
Kilian Weiss https://orcid.org/0000-0003-4295-4585
Christine Preibisch https://orcid.org/0000-0003-4067-1928
Julia A. Schnabel https://orcid.org/0000-0001-6107-3009

REFERENCES

1. Godenschweger F, Kägebein U, Stucht D, et al. Motion correction in MRI of the brain. *Phys Med Biol*. 2016;61:R32-R56.
2. Zaitsev M, Maclaren J, Herbst M. Motion artifacts in MRI: A complex problem with many partial solutions. *J Magn Reson Imaging*. 2015;42:887-901.
3. Hirsch NM, Toth V, Förschler A, Kooijman H, Zimmer C, Preibisch C. Technical considerations on the validity of blood oxygenation level-dependent-based MR assessment of vascular deoxygenation. *NMR Biomed*. 2014;27:853-862.
4. Liu J, de Zwart JA, van Gelderen P, MurphyBoesch J, Duyn JH. Effect of head motion on MRI B0 field distribution. *Magn Reson Med*. 2018;80:2538-2548.
5. Marques JP, Bowtell R. Application of a Fourier-based method for rapid calculation of field inhomogeneity due to spatial variation of magnetic susceptibility. *Concepts Magn Reson Part B: Magn Reson Eng*. 2005;25B:65-78.
6. Magerkurth J, Volz S, Wagner M, et al. Quantitative T2* -mapping based on multi-slice multiple gradient echo flash imaging: Retrospective correction for subject motion effects: Movement Correction in T2* Mapping. *Magn Reson Med*. 2011;66:989-997.
7. Eichhorn H, Hammernik K, Epp SM, Karampinos DC, Schnabel JA, Preibisch C. Investigating the impact of motion and associated B0 changes on oxygenation sensitive MRI through realistic simulations. *Proceedings of the Annual Meeting of ISMRM*; 2023 Abstract 3410.
8. Nöth U, Volz S, Hattingen E, Deichmann R. An improved method for retrospective motion correction in quantitative T2* mapping. *NeuroImage*. 2014;92:106-119.
9. van der Kouwe AJW, Benner T, Dale AM. Real-time rigid body motion correction and shimming using cloverleaf navigators. *Magn Reson Med*. 2006;56:1019-1032.
10. Matakos A, Fessler JA. Joint estimation of image and fieldmap in parallel MRI using single-shot acquisitions. *Proceedings of the 7th IEEE International Symposium on Biomedical Imaging*; IEEE; 2010:984-987.
11. Brackenier Y, CorderoGrande L, TomiTricot R, et al. Data-driven motion-corrected brain MRI incorporating pose-dependent B0 fields. *Magn Reson Med*. 2022;88:817-831.
12. Hewlett M, Oran O, Liu J, Drangova M. Motion and B0 correction for multiparametric mapping with jointly acquired FID and spherical navigators. *Proceedings of the Annual Meeting of ISMRM*; 2024 Abstract 2662.
13. Spieker V, Eichhorn H, Hammernik K, et al. Deep learning for retrospective motion correction in MRI: A comprehensive review. *IEEE Trans Med Imaging*. 2024;43:846-859.
14. Pirkl CM, Cencini M, Kurzawski JW, et al. Learning residual motion correction for fast and robust 3D multiparametric MRI. *Med Image Anal*. 2022;77:102387.
15. Xu X, Kothapalli SVVN, Liu J, et al. Learning-based motion artifact removal networks for quantitative R2* mapping. *Magn Reson Med*. 2022;88:106-119.
16. Johnson PM, Drangova M. Conditional generative adversarial network for 3D rigid-body motion correction in MRI. *Magn Reson Med*. 2019;82:901-910.
17. Küstner T, Armanious K, Yang J, Yang B, Schick F, Gatidis S. Retrospective correction of motion-affected MR images using deep learning frameworks. *Magn Reson Med*. 2019;82:1527-1540.
18. Oh G, Lee JE, Ye JC. Unpaired MR motion artifact deep learning using outlier-rejecting bootstrap aggregation. *IEEE Trans Med Imaging*. 2021;40:3125-3139.
19. Haskell MW, Cauley SF, Bilgic B, et al. Network Accelerated Motion Estimation and Reduction (NAMER): Convolutional neural network guided retrospective motion

correction using a separable motion model. *Magn Reson Med*. 2019;82:1452-1461.
20. Hossbach J, Splitthoff DN, Cauley S, et al. Deep learning-based motion quantification from K-space for fast model-based MRI motion correction. *Med Phys*. 2022;50:2148-2161.
21. Singh NM, Dey N, Hoffmann M, et al. Data consistent deep rigid MRI motion correction. *Proceedings of machine learning research, medical imaging with deep learning*. Vol 227; 2024:368-381.
22. Oksuz I, Clough JR, Ruijsink B, et al. Deep learning-based detection and correction of cardiac MR motion artefacts during reconstruction for high-quality segmentation. *IEEE Trans Med Imaging*. 2020;39:4001-4010.
23. Motyka S, Weiser P, Bachrata B, et al. Predicting dynamic, motion-related changes in B0 field in the brain at a 7T MRI using a subject-specific fine-trained U-net. *Magn Reson Med*. 2024;91:2044-2056.
24. Eichhorn H, Spieker V, Hammernik K, et al. Physics-informed deep learning for motion-corrected reconstruction of quantitative brain MRI. *Medical image computing and computer assisted intervention-MICCAI 2024. MICCAI*. Lecture Notes in Computer Science. Vol 15007. Springer Nature Switzerland; 2024.
25. Atkinson D. Retrospective correction of motion in MR images. In: van der Kouwe AJW, Andre JB, eds. *Motion Correction in MR*. Academic Press; 2022:259-267.
26. Eichhorn H, Hammernik K, Spieker V, et al. Physics-aware motion simulation For T2*-weighted brain MRI. *Simulation and Synthesis in Medical Imaging. SASHIMI*. Lecture Notes in Computer Science. Vol 14288. Springer Nature Switzerland; 2023.
27. Eichhorn H, Hammernik K, Spieker V, et al. PHIMO: Physics-informed motion correction of GRE MRI for T2* quantification. *Proceedings of the Annual Meeting of ISMRM*; 2024 Abstract 0391.
28. Kingma DP, Ba J. Adam: A method for stochastic optimization. *Proceedings of the 3rd ICLR*; 2015.
29. Cootes TF, Taylor CJ, Cooper DH, Graham J. Active shape models - their training and application. *Comput Vis Image Underst*. 1995;61:38-59.
30. Marchetto E, Eichhorn H, Gallichan D, Schwarz ST, Shekhrajka N, Ganz M. Assessing image quality metric alignment with radiological evaluation in datasets with and without motion artifacts. *Proceedings of the Annual Meeting of ISMRM*; 2024 Abstract 3019.
31. Eichhorn H, Marchetto E, Gallichan D, Schnabel JA, Ganz M. Assessing the influence of preprocessing on the agreement of image quality metrics with radiological evaluation in the presence of motion. *Proceedings of the Annual Meeting of ISMRM*; 2025 Abstract 7983.
32. Wellcome Centre for Human Neuroimaging. UCL queen square institute of neurology. *Statistical Parametric Mapping*; 2014. https://www.fil.ion.ucl.ac.uk/spm/software/spm12/
33. Kaczmarz S, Hyder F, Preibisch C. Oxygen extraction fraction mapping with multi-parametric quantitative BOLD MRI: Reduced transverse relaxation bias using 3D-GraSE imaging. *NeuroImage*. 2020;220:117095.
34. Hammernik K, Küstner T. Machine Enhanced Reconstruction Learning and Interpretation Networks (MERLIN). *Proceedings of the Annual Meeting of ISMRM-ESMRMB*; 2022 Abstract 1051.
35. CorderoGrande L, Ferrazzi G, Teixeira RPAG, O'Muircheartaigh J, Price AN, Hajnal JV. Motion-corrected MRI with DISORDER: Distributed and incoherent sample orders for reconstruction deblurring using encoding redundancy. *Magn Reson Med*. 2020;84:713-726.
36. Committee QCO,Bilgic B, Costagli M, et al. Recommended implementation of quantitative susceptibility mapping for clinical research in the brain: A consensus of the ISMRM electro-magnetic tissue properties study group. *Magn Reson Med*. 2024;91:1834-1862.
37. Heckel R, Jacob M, Chaudhari A, Perlman O, Shimron E. Deep learning for accelerated and robust MRI reconstruction. *MAGMA*. 2024;37:335-368.
38. Hirsch NM, Preibisch C. T2* mapping with background gradient correction using different excitation pulse shapes. *Am J Neuroradiol*. 2013;34:E65-E68.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

Data S1. Supporting Information

Figure S1. Examples of reconstructed images (first echo) for networks trained conventionally (with original masks and no *KeepCenter*) and with the *KeepCenter* extension, compared to the fully sampled image. From top to bottom, the masks exclude three, five, and seven central k-space lines. Green arrows indicate brain areas with more correctly recovered contrast in the *KeepCenter* reconstructions.

Figure S2. Example exclusion masks estimated from inferior, middle, and superior slices, comparing the *Even/Odd* extension to an optimization of individual slices for two different motion patterns of the MoCo validation subject. For these acquisitions, the slice range based on the susceptibility gradient strength for the *Even/Odd* optimization is [10, 17]. These examples demonstrate a more stable-line detection profile across the brain for the *Even/Odd* optimization compared to masks optimized for individual slices.

Figure S3. Qualitative examples for applying PHIMO+ and the comparison methods to apparently motion-free data. The resulting T_2^* maps (and, if available, exclusion masks) are compared for the original data without intentional motion, ORBA, SLD, PHIMO, PHIMO+, and HR/QR for two different subjects. Blue arrows indicate T_2^* quantification errors introduced by SLD.

Figure S4. Acquisition excluded from the main analysis due to excessive motion in the k-space center (10 s/nine lines). The T_2^* maps are compared for ORBA, SLD, PHIMO, PHIMO+, and HR/QR to the uncorrected acquisition (left) and the separate motion-free acquisition (right). The respective exclusion masks are shown on the right. Green arrows indicate areas where PHIMO and PHIMO+, and

to some extent HR/QR, clearly mitigate the extent of wave-like motion artifacts, blue arrows indicate missing details across all methods. PHIMO and PHIMO+ overestimate the exclusion mask, likely due to excessive motion in the k-space center, which challenges the reconstruction network even with the *KeepCenter* extension. Note that the current standard, HR/QR, is also challenged by such an extreme motion case, which, in clinical applications, may ultimately require reacquisition.

How to cite this article: Eichhorn H, Spieker V, Hammernik K, et al. Motion-robust T_2^* quantification from low-resolution gradient echo brain MRI with physics-informed deep learning. *Magn Reson Med*. 2026;95:346-362. doi: 10.1002/mrm.70050