1 Analysis of the number of components $R$ in the low-rank motion model

To test the hypothesis that the low-rank motion model can accurately represent motion-fields with a small number of components $R$, a singular value analysis was performed on 3D+t abdominothoracic and head motion-fields. Motion-fields with a large degree of freedom in the reconstructions were reconstructed from Eq. [9] by setting the number of components to 20, and the regularization to zero. Then, singular values $\sigma_j$ were computed by means of an SVD on the Casorati matrix of the resulting spatio-temporal motion-fields (spatio-temporal motion-fields ordered in a matrix with rows as space, and columns as time). Finally, the variance of the motion-fields captured by models with a reduced number of components $R$ was analyzed by computing $\sum_{j=1}^{R} \sigma_j^2/\sum_{j=1}^{20} \sigma_j^2$. See Supporting Information Figure S1 for the results of the analyses.

For both 3D+t abdominothoracic and head motion-fields, the resulting variance plots show a steep increase. This indicates that a low-rank assumption on the motion-fields is reasonable. The figure also shows that models with $R = 3$ and $R = 6$ for respectively 3D+t respiratory and head motion capture a significant amount of the variance (97.8% and 99.9% respectively) in spatio-temporal motion-fields that were reconstructed with a large degree of freedom.

In addition to the singular value analysis, the effects of the model order $R$ on the results were quantitatively analyzed for 3D+t respiratory-resolved motion-fields. For this analysis the relative error norm (REN), and normalized mutual information (NMI) were computed between MR-MOTUS reconstructions with $R = 1, \ldots, 20$ and image reconstruction on respiratory-resolved data. For the MR-MOTUS reconstructions, the reconstruction parameters listed in Table 1 were used. The means over all 20 reconstructed respiratory phases are shown in Supporting Information Figure S2. Only minor changes in the metrics can be observed between the different $R$-values. Based on these analyses, combined with visual inspections of the results, we have set $R = 3$ for respiratory motion, and $R = 6$ for head-and-neck motion.
2 Extension of MR-MOTUS to multi-coil acquisitions

In (34), the model in Eq. [3] was derived by substituting
\[ q_t(x) = q_0(U_t(x)) \text{det} \nabla(U_t(x)) \]  
[\text{S1}]
in the conventional MRI signal equation
\[ s_t(k) = \int_{\Omega} q_t(x)e^{-i2\pi k \cdot x}dx, \]  
[S2]
followed by a change-of-variables. In a multi-coil acquisition, k-space data \( s_j^t(k) \) from coil \( j \) with coil sensitivity \( S^j(x) \in \mathbb{C} \) can be modelled as
\[ s_j^t(k) = \int_{\Omega} S^j(x)q_t(x)e^{-i2\pi k \cdot x}dx. \]  
[S3]
Substitution of Eq. [S1] in Eq. [S3], followed by a change-of-coordinates \( x \rightarrow U_1^{-1}(x) \) results in a signal model similar to the one used in this work. The change-of-coordinates will make the intensity variations due to the coil sensitivities dependent on the unknown motion-fields:
\[ s_j^t(k) = \int_{\Omega} S^j(U_1^{-1}(x))q_0(x)e^{-i2\pi k \cdot U_1^{-1}(x)}dx. \]  
[S4]
Unfortunately, there is a practical problem with the multi-coil signal model Eq. [S3]: the reconstruction times are increased by a factor \( N_c \) for \( N_c \) coils, since \( N_c \) times more NUFFT are required. Additionally, it will require a warping \( S^j(U_1^{-1}(x)) \) at every dynamic, which further increases the reconstruction times. Altogether, this makes the reconstruction less practical.

In theory, the model could be combined with any available k-space-based coil compression technique that gives access to the coil sensitivities that modulate the compressed k-space data. In practice, this requires knowledge of the coil sensitivities of the compressed data, which are possibly different at every dynamic. We have therefore opted for a linear coil compression with coefficients \( c^j \in \mathbb{C} \) that produces a trade-off between homogeneous coil sensitivities and SNR-loss. The resulting approximately homogeneous coil sensitivity allows to neglect the coil sensitivities in the signal model, since \( \sum_j c^j S^j(U_1^{-1}(x)) \approx 1 \) for all \( x \). Combining both requirements, the compression coefficients can be obtained by solving the following minimization problem
\[ \min_{c} ||Sc - 1||_2^2 + \lambda_{cc}||Uc||_2^2, \]  
[S5]
Here \( S \in \mathbb{C}^{N \times N_c} \) denotes the matrix of coil sensitivities, \( 1 \in \mathbb{C}^{N \times 1} \) denotes a vector with all entries equal to one, and \( U \) denotes Cholesky factor of the noise covariance matrix \( \Sigma \) such that \( \Sigma = U^H U \). The minimization problem in Eq. [S5] has a unique solution given by
\[ c = (S^H S + \lambda_{cc} \Sigma)^{-1} S^H 1, \]  
[S6]
41
where the superscript $H$ denotes the conjugate transpose.

Applying the compression coefficients in Eq. [S6] to both sides of Eq. [S3] yields:

$$
\tilde{s}_t = F(D_t | \tilde{q}_0).
$$

Here $\tilde{s}_t := \sum_j c^j s^j_t$ and $\tilde{q}_0 := \sum_j c^j S^j(x)q_0(x) \approx q_0(x)$ denote the coil-compressed k-space data and reference image, respectively, and can be pre-computed before the motion-field reconstruction. Note that this includes the same operator as in Eq. [2], such that the same reconstruction strategy as for single-channel data can now be employed to reconstruct motion-fields from multi-coil acquisitions. This thus allows for similar reconstruction times with multi-coil data, but comes at the cost of SNR loss. For all experiments considered in this work we have set the compression parameter $\lambda_{CC} = 5 \cdot 10^5$; this yielded an empirically good trade-off between SNR loss and homogeneous sensitivities.

As mentioned above, the proposed compression is not optimal in terms of SNR (63, 64). To this extent, Supporting Information Figure S3 analyzes the loss in SNR between a Roemer coil combination (63) and the proposed coil compression on the 2D respiratory data. This shows an SNR loss factor between 1.5 and 2.5 in most of the body, which increases towards the boundary of the body. k-Space-based parallel imaging methods such as GRAPPA could possibly be used to improve the SNR after the compression.

3 Reference image reconstruction from free-breathing, multi-coil acquisitions

MR-MOTUS requires a motion-free reference image with slowly varying coil sensitivity. To be consistent with the forward model in Eq. [S7], the reference image $\tilde{q}_0$ should be reconstructed from the data $\tilde{s}_t$ after coil compression. This data was, however, acquired during free-breathing which makes the reconstruction of a motion-free image not trivial. The extreme positions in the respiratory motion (i.e. end-inhale, end-exhale) can be assumed relatively constant over different breathing cycles. To extract motion-free k-space data from the free-breathing data, we retrospectively selected only the readouts acquired at end-inhale; this yielded consistently better results than end-exhale. To enable this selection, we performed phase-binning on every readout based on the magnitude of the signal at $k = 0$ (denoted as $k_0$-values). A low-pass filter was applied to the time-series of $k_0$-values prior to the binning, based on the expected breathing
frequency. Next, phase-binning was performed with 10 phases per breathing cycle. Finally, the reference image was reconstructed from the end-inhale bin, using a density-compensated preconditioned conjugate-gradient reconstruction.

4 Motion reconstruction details

We followed the approach outlined in Section 2.2, and reconstructed motion-fields from multi-coil k-space data acquired during motion by solving the minimization problem [9]. The low-rank MR-MOTUS workflow is schematically summarized in Figure 1. The reconstruction problem [9] was solved with L-BFGS (46), using the MATLAB implementation from (47). The L-BFGS memory parameter was set to 20 and sampling density compensation (48, 49) was applied to improve the conditioning of the reconstruction. The elements in both of the matrices \( \Phi \) and \( \Psi \) were initialized as uniformly random draws from \([-0.5, 0.5]\), after which the matrices were scaled to unit Frobenius norm. This initialization yields equally sensitive updates in \( \Phi \) and \( \Psi \) during the starting phase of the reconstruction, and lead to consistently good results both in simulations and experimental settings despite the non-convex nature of the reconstruction problem. The multi-coil data was compressed to a single channel prior to all reconstructions, see Supporting Information Section 1 for more details. The regularization parameter \( \lambda_R \) was chosen according to \( \lambda_R \sim 1/M \) and the data \( s_t \) were scaled by the norm of the density-compensated k-space data in order to obtain consistent values between experiments. The reconstruction parameters were determined with a heuristic parameter search (see Supporting Information Section 3). The forward and adjoint operators were implemented using the type-3 NUFFT from (50). We refer to Table 1 for all parameter settings and to Supporting Information Section 4 and the Supporting Information in (34) for more implementation details. Code that produces similar results as presented in this study is openly available at https://github.com/nrfhuttinga/LowRank_MRMOTUS.git.

The forward model in [3], and corresponding derivatives (see Supporting Information in (34)) required for the motion-field reconstruction problem [9] are computed using the NUFFT-operator from (50). We observe that the evaluation of the data-fidelity term \( E[.] \) in the objective function [4] includes a sum over \( M \) independent terms. We found that a so-called ‘embarassingly parallel’ evaluation of the data-fidelity term can result in a reduced computation time. This requires a single-thread compilation of the NUFFT, and a parallelization over the \( M \) independent terms in the summation. The parallization was realized in MATLAB with a ‘parfor-loop’ over the
dynamics. For 3D this resulted in a significant improvement in computation times, while for 2D the speed-up due to parallelization was smaller than the thread initialization overhead. Hence, for 2D we consider the non-parallelized evaluation and for 3D we consider the parallelized evaluation.

5 Heuristic search for reconstruction parameter settings

We have heuristically determined the reconstruction parameter settings based on the quality of the respiratory-resolved MR-MOTUS reconstructions described in Section 3.2. The number of dynamics was set equal to the number of bins, i.e. 20, and as a quality metric we have taken the maximum REN between the reconstructions over all 20 dynamics. This allowed to select the parameters with the best worst-case performance over all respiratory phases. The reconstruction parameters included the resolution of the reference image, the binning-phase of the reference image, the spline order for the motion-field basis and the number of motion model components $R$. The optimal reconstruction parameters were determined as follows. First, the spline order was fixed to 9 (i.e. 9 spline functions over the whole FOV, in every direction), in order to have at least the 4 points required to correctly define a single third order B-spline on every resolution level. A grid search was then performed for the reference image resolution, binning-phase and the number of motion model components, with as metric the maximum REN over all dynamics. Next, the optimal parameters resulting from the grid search were selected, the spline order was increased to the maximum (i.e. 1 spline function per 4 grid-points), and a second grid search was performed for only the regularization parameter $\lambda_R$. To vary the spatial resolution of the reference image we have retrospectively varied the $k_{\text{max}}$ of the data, resulting in varying matrix sizes and, consequently, reconstruction times. To vary the respiratory-phase, the binning as described in Supporting Information Section 3 is performed to select only the required phase.

Supporting Information Figure S4 visualizes the quantitative results of the parameter search. Supporting Information Figure S4A shows the REN of the worst-case dynamic, minimized over the number of motion model components ($R = 1 - 6$), versus the reference image resolution. No regularization was applied yet. The best worst-case performance was obtained for a reference image resolution binned in end-inhale with a spatial resolution of 6.7 mm isotropic, and motion model with $R = 3$. This reference resolution was subsequently fixed, the motion resolution was set to the maximum of 1 spline function per 4 grid-points (i.e. spline order 16), and the optimal
value for the regularization parameter was determined with a grid-search. This resulted in an optimum value of $\lambda_R = 15$. Supporting Information Video S12 qualitatively compares the MR-MOTUS reconstructions with respiratory-resolved image reconstructions for different reference image binning-phases and reference image resolution. Although the visual differences between the reconstructions are minimal, it can be observed from Supporting Information Figure S4B that the reconstruction times are about a factor 3 lower for a reference image resolution of 6.7 mm, than for a resolution of 3.4 mm.
Supporting Figures

Supporting Information Video S1: *This is an animated figure and should be viewed under Supporting Information.* 2D-t compressed sensing reconstruction (left), MR-MOTUS warped reference images (middle), and pixel-wise absolute differences between the two reconstructions (right), as mentioned in Section 3.1 and Section 4.1. The top row shows reconstructions for volunteer 1, and the bottom row for volunteer 2.
Supporting Information Video S2: *This is an animated figure and should be viewed under Supporting Information.* MR-MOTUS warped reference images overlayed with reconstructed dynamic motion-fields from 2D time-resolved data, as mentioned in Section 3.1 and Section 4.1. The image shows a decomposition in the reconstructed components $\Phi^i$ (spatial) and $\Psi$ (temporal) for volunteer 1. For visualization purposes the components were scaled such that $\|\Phi^i\| = 1$. 
Supporting Information Video S3: This is an animated figure and should be viewed under Supporting Information. MR-MOTUS warped reference images overlayed with reconstructed dynamic motion-fields from 2D time-resolved data, as mentioned in Section 3.1 and Section 4.1. The image shows a decomposition in the reconstructed components $\Phi^i$ (spatial) and $\Psi$ (temporal) for volunteer 2. For visualization purposes the components were scaled such that $\|\Phi^i\| = 1$. 
Supporting Information Video S4: This is an animated figure and should be viewed under Supporting Information. Respiratory-resolved image reconstruction (Resp. resolved IR, left), MR-MOTUS warped reference images (middle), and pixel-wise absolute differences between the two reconstructions (right), as mentioned in Section 3.2 and Section 4.2. The visualization shows data from volunteer 1.
Supporting Information Video S5: This is an animated figure and should be viewed under Supporting Information. Respiratory-resolved image reconstruction (Resp. resolved IR, left), MR-MOTUS warped reference images (middle), and pixel-wise absolute differences between the two reconstructions (right), as mentioned in Section 3.2 and Section 4.2. The visualization shows data from volunteer 2.
Supporting Information Video S6: This is an animated figure and should be viewed under Supporting Information. MR-MOTUS warped reference images overlayed with reconstructed dynamic motion-fields from respiratory-sorted data, as mentioned in Section 3.2 and Section 4.2. The image shows a decomposition in the reconstructed components $\Phi^i$ (spatial) and $\Psi$ (temporal) for volunteer 1. For visualization purposes the components were scaled such that $\|\Phi^i\| = 1$. 
Supporting Information Video S7: *This is an animated figure and should be viewed under Supporting Information.* MR-MOTUS warped reference images overlayed with reconstructed dynamic motion-fields from respiratory-sorted data, as mentioned in Section 3.2 and Section 4.2. The image shows a decomposition in the reconstructed components $\Phi^i$ (spatial) and $\Psi$ (temporal) for volunteer 2. For visualization purposes the components were scaled such that $\|\Phi^i\| = 1$. 
Supporting Information Video S8: This is an animated figure and should be viewed under Supporting Information. MR-MOTUS warped reference images overlayed with reconstructed dynamic motion-fields from respiratory-sorted data, as mentioned in Section 3.2 and Section 4.2. The image shows a decomposition in the reconstructed components $\Phi^i$ (spatial) and $\Psi$ (temporal) for volunteer 1. For visualization purposes the components were scaled such that $\|\Phi^i\| = 1$. 
Supporting Information Video S9: *This is an animated figure and should be viewed under Supporting Information.* MR-MOTUS warped reference images overlayed with reconstructed dynamic motion-fields from respiratory-sorted data, as mentioned in Section 3.2 and Section 4.2. The image shows a decomposition in the reconstructed components $\Phi^i$ (spatial) and $\Psi$ (temporal) for volunteer 2. For visualization purposes the components were scaled such that $\|\Phi^i\| = 1$.

Supporting Information Video S10: *This is an animated figure and should be viewed under Supporting Information.* MR-MOTUS warped reference images resulting from the 3D head-and-neck motion reconstructions for volunteer 1, as mentioned in Section 3.3 and Section 4.3.
Supporting Information Video S11: This is an animated figure and should be viewed under Supporting Information. MR-MOTUS warped reference images resulting from the 3D head-and-neck motion reconstructions for volunteer 2, as mentioned in Section 3.3 and Section 4.3.
Supporting Information Video S12: This is an animated figure and should be viewed under Supporting Information. Respiratory-resolved image reconstruction (Resp. resolved IR, left), MR-MOTUS warped reference images (middle), and pixel-wise absolute differences between the two reconstructions (right), as mentioned in Supporting Information Section 5 and Section 3.2. The four blocks show reconstructions with different reconstruction parameter settings. ‘InhaleBinned’ denotes whether the reference image is binned in inhale (1) or exhale (0). ‘Ref. resolution’ denotes the resolution of the reference image in millimeters. All motion-fields were reconstructed without regularization and with 9 cubic spline functions in every direction.
Supporting Information Figure S1: Results of the singular value analyses in Supporting Information Section 1 for 3D+t respiratory motion (left), and 3D+t head-and-neck motion (right). This figure clearly indicates that 3D+t motion-fields possess the low-rank property; models with \( R = 3 \) and \( R = 6 \) can respectively capture 97.9% and 99.9% of the variance of 3D+t respiratory motion and 3D+t head-and-neck motion, allowing for a significant reduction in the number of unknowns.

Supporting Information Figure S2: This figure shows the effect of a different number of components \( R = 1, \ldots, 20 \) on the respiratory motion reconstructions, as discussed in Supporting Information Section 1. The metrics were evaluated on respiratory-resolved MR-MOTUS and image reconstructions, and the means over all 20 reconstructed respiratory phases are visualized in this figure. Only a minimal change can be observed in both metrics, showing that the effect of \( R \) on the results is minimal for respiratory motion. NMI = Normalized Mutual Information, REN = Relative Error Norm.
Supporting Information Figure S3: (A) Roemer reconstruction (63). (B) The coil compression with $\lambda_{CC} = 5 \cdot 10^5$, as discussed in Supporting Information Section 2. (C) SNR loss factor between (A) and (B). (D) The histogram of the SNR loss factor in (C). The SNR loss factor is between 1.5 and 2.5 in most of the body and increases towards the boundary of the body.

Supporting Information Figure S4: Results of the parameter search as mentioned in Supporting Information Section 5 and Section 3.2. (A) The effect of the reference image resolution and reference image respiratory binning-phase on the reconstruction quality. (B) The effect of the reference image resolution on the reconstruction time. In the figure, ‘InhaleBinned’ refers to the binning phase for the reference image (InhaleBinned=1 for inhale, InhaleBinned=0 for exhale), ‘Resolution’ denotes the spatial resolution of the reference image, and ‘SplineOrder’ denotes the number of spline basis functions defined per spatial dimension.