

Fully-automated quality assurance in multi-center studies using MRI phantom measurements



Mathias Davids^a, Frank G. Zöllner^{a,*}, Michaela Ruttorf^a, Frauke Nees^b, Herta Flor^b,
Gunter Schumann^{c,d}, Lothar R. Schad^a
the IMAGEN Consortium

^a Computer Assisted Clinical Medicine, Medical Faculty Mannheim, Heidelberg University

^b Department of Cognitive and Clinical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University

^c MRC Social, Genetic and Developmental Psychiatry Centre, London, UK

^d Institute of Psychiatry, King's College London, UK

ARTICLE INFO

Article history:

Received 14 June 2013

Revised 20 January 2014

Accepted 24 January 2014

Keywords:

MRI

Automated quality assurance

Phantom measurement

Multicenter study

ABSTRACT

Phantom measurements allow for investigating the overall quality characteristics of an MRI scanner. Especially within multicenter studies, these characteristics ensure the comparability of the results across different sites, in addition to the performance stability of a single scanner over time. This comparability requires consistent phantoms, sequence protocols, and quality assurance criteria. Within the scope of this work, a software library was implemented for fully-automated determination of important quality characteristics (comprising **signal-to-noise ratio, image uniformity, ghosting artifacts, chemical shift and spatial resolution and linearity**) including methods for data preparation, automated pre- and postprocessing as well as visualization and interpretation. All methods were evaluated using both synthetic images with predefined distortions and a set of 44 real phantom measurements involving eight sites and three manufacturers. Using the synthetic phantom images, predefined levels of distortion that were incorporated artificially were correctly detected by the automated routines with no more than 2.6% of relative error. In addition, the methods were applied to real phantom measurements - all data sets could be evaluated automatically considering all quality parameters as long as the acquisition protocols are followed. Shortcomings of the processability only occurred in the ghosting artifacts (39/44 evaluable) and the spatial linearity (43/44 evaluable) analysis due to gross misalignments of the phantom during image acquisition. Based on evaluation results, the accuracy of the evaluation appears to be robust to misalignments, artifacts, and distortions affecting the images, allowing for objective fully-automated evaluation and interpretation of large data set numbers.

© 2014 Elsevier Inc. All rights reserved.

1. Introduction

Longitudinal multicenter studies allow for including larger patient cohorts and thus provide population based results unlike studies with specialized patient groups [1–4]. In general this is important in standardized image guided diagnostics and therapy [5] or in studies aiming at correlating findings from genetics and behavioral development to functional magnetic resonance imaging (fMRI) [6]. Therefore, certain quality standard protocols should be followed to allow for comparing results across sites [7]. This includes identical protocols, phantoms, and repeated measurements at certain time intervals.

In magnetic resonance imaging (MRI), many influences based on the principles of the imaging process exist that may degrade the

image quality. **Incorrect coil shimming yielding an inhomogeneous B_0 field may lead to geometric distortions. Different sources such as the electrical system itself or the imaging sequence parameters such as the slice thickness influence the level of noise present in the image and hence its signal-to-noise ratio (SNR). Varying penetration depths of the surface coils may lead to inhomogeneities of the signal intensities and decreased spatial resolution. In addition, the resonance frequency shift of fat compared to water (up to some hundreds of Hertz) may induce alleged shifts between the corresponding tissues as well as diffuse borders.** These different influences can be detected and quantified using quality measurements based on dedicated phantoms [8].

Several approaches are available, providing manual or at most partially automated methods for determining basic quality parameters such as the SNR based on phantom data sets [9–13]. Pursuing this concept, the aim of this study was to develop a software library to strictly automatically extract important quality parameters from

* Corresponding author. Tel.: +49 6213835117.

E-mail address: frank.zoellner@medma.uni-heidelberg.de (F.G. Zöllner).

phantom data sets to facilitate a fast and objective quality assurance in MRI. This includes SNR, image uniformity, ghosting artifacts, and spatial resolution and linearity. In addition, methods for data preparation, automated pre- and postprocessing as well as visualization and interpretation are provided. Commercial packages are available that allow for automatic evaluation of quality measurement data to some extent (e.g., *Radia Diagnostic* package, Radiological Imaging Technology, Colorado Springs, CO, USA). However, we provide an extendable open-source library that allows for full automation of such an analysis.

2. Material and Methods

2.1. MR imaging and phantom

For the determination of the quality parameters, an MRI phantom (*JM Specialty Parts Inc.*, San Diego, CA, USA) recommended in the American College of Radiology (ACR) accreditation program [8] was used. The cylindrical phantom with a diameter of 20.4 cm/19.0 cm and a height of 16.5 cm/14.8 cm (for the outer and inner dimensions, respectively) is filled with 10 mmol nickel chloride solution, including 45 mmol sodium chloride simulating biological conductivity. Several structures in the phantom allow for assessment of the scanner's image quality (cf. Fig. 1, A). The measurement protocol to acquire phantom data complied with the ACR specifications (T_2 -weighted imaging, repetition time 2000 ms, echo times 20 ms and 80 ms, 5 mm slice thickness, 5 mm gap, 11 slices, 25 cm² field of view (FOV), 256 × 256 matrix), except for use of parallel imaging factor 2 which was added in correspondence with the measurement protocol used by the multicenter study [14]. The first slice contains a vertical bar allowing for measuring the phantom's rotation along its cylindrical axis (tilt angle). Two squares in the same slice (filled with 10 mmol nickel chloride solution and vegetable fat, respectively) that are arranged catty-corner can be used to determine the chemical shift and the receiving bandwidth. In addition, the slice contains three small grids of 31 holes with square basis per grid (diameter 1.1 mm, 1.0 mm, and 0.9 mm, respectively) to quantify the scanner's spatial resolution capacity (see Fig. 1, B). A slice without any structures is used to quantify the image uniformity and the level of artifacts being present (cf. Fig. 1, C). The level of geometric distortions caused by the scanner's imperfections is obtained from the fifth slice containing a large grid structure (cf. Fig. 1, D). Low contrast inserts of thin polycarbonate membranes of different thickness (cf. Fig. 1, top of sagittal view) allow for quantification of the scanner's contrast resolution.

2.2. Fully-automated assessment of quality parameters

The software library was implemented in MATLAB R2011b (*The MathWorks Inc.*, Natick, MA, USA) including the *Curve Fitting*, *Image*

Processing, and *Statistics* toolboxes, providing methods for fully-automated processing of the data sets (cf. Fig. 2). First, all data sets are imported using a graphical user interface. The data set is stored in a single file along with all information from the DICOM header required for further processing. Afterwards, all data sets are evaluated fully-automatically accordingly to the quality parameters, including pre- and postprocessing. The preprocessing includes localization of the phantom within the images and verification of the sequence parameters according to the ACR protocol, whereas in postprocessing preparation of the data for subsequent handling and saving of the evaluation results is carried out. Several methods are available for visualization and interpretation of the results, allowing for site and scanner-wide comparison and detection of temporal changes of quality characteristics. The actual evaluation of the quality parameters follows the Phantom Test Guidance for the ACR MRI Accreditation Program (06/05) and is subdivided in several stand-alone methods described in the following.

2.2.1. Phantom Localization

For all quality parameters described in the following sections, an **automatic localization of the phantom** in the image is performed. This is achieved by detecting the first and the last row of the image that contains a pixel representing the phantom, with the threshold being determined by a *k-means* clustering (with $k = 2$, one gray value class representing the background and a second class representing the phantom) [15]. The two points denoting the upper (x_u, y_u) and lower (x_l, y_l) edge of the phantom are defined by the row index y and the position x of the bright pixel in the corresponding row yielding an estimated position of the phantom center

$$(c_x, c_y) \approx \left(\frac{1}{2}(x_u + x_l), \frac{1}{2}(y_u + y_l) \right)$$

and the radius of the phantom

$$r \approx \frac{1}{2} \sqrt{(x_u - x_l)^2 + (y_u - y_l)^2}.$$

The exact values are determined by sampling points representing the upper and lower edge of the phantom, by determining the first and last pixel in a column greater than the *k-means*-threshold (cf. Fig. 3). Since the phantom is circular, the sample points of both the upper and lower phantom edge shape a circle and can, therefore, be used to fit a corresponding function

$$c(x) = \pm \sqrt{\max(r^2 - (x - c_x)^2, 0)} + c_y, \quad (1)$$

yielding two separated fits. By averaging the results of these two fits, more accurate values for the phantom center (c_x, c_y) and the radius r are obtained. One of the slices containing a horizontal bar is used to

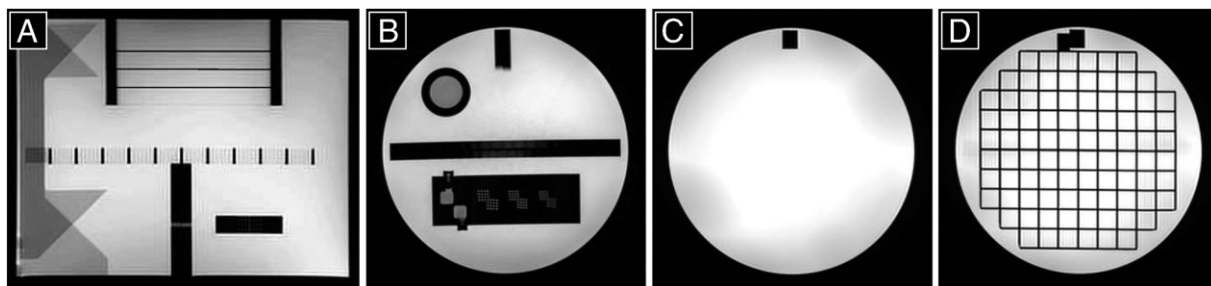


Fig. 1. Overview of the ACR phantom: sagittal localizer view (A), transaxial view of the slices used for angle determination, spatial resolution, and chemical shift (B), signal-to-noise ratio, image uniformity, and ghosting artifacts (C), and spatial linearity (D).

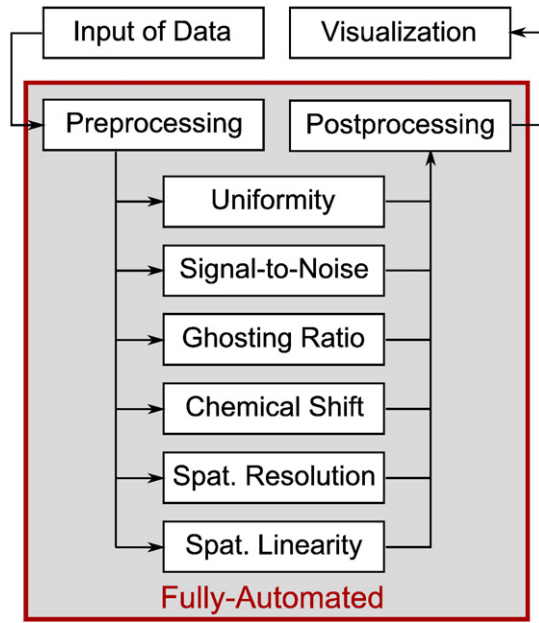


Fig. 2. Flowchart illustrating the pipeline of processing for phantom datasets; the steps within the gray box are performed fully-automated.

determine the **tilt angle α of the phantom along its z-axis**. Again, sample points are extracted, representing the upper and lower edge of the bar. By fitting a linear function to both the upper and lower set of sample points and averaging both angles, the final tilt angle α is deduced. Assuming the phantom not to be deformed along its axis, this value α is used for all other slices. In addition the tilt of the phantoms z-axis is obtained from the center points of all slices along with the slice gaps extracted from the DICOM header. If the tilt angle along one orientation exceeds a predefined threshold (default value 1°), the user is notified by a warning message.

2.2.2. Integral Image Uniformity

The image uniformity, characterizing the homogeneity of ideally uniform regions, is determined using a slice of the phantom without structures (see Fig. 1, C) [16]. A region of interest (ROI) centered at

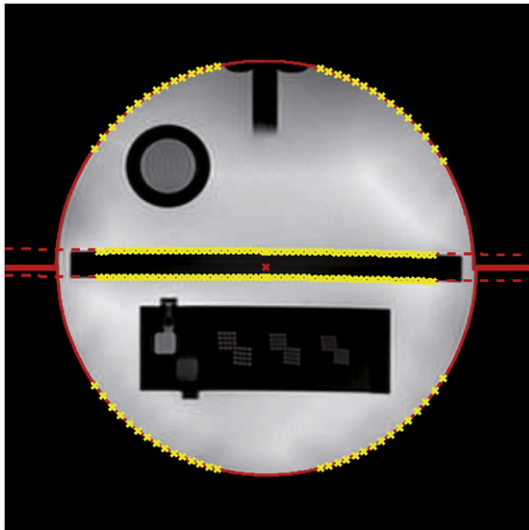


Fig. 3. By fitting the sample points of the phantom edges and the center bar (yellow crosses) with the circular function (solid line) and the linear function (dashed line), the center and the tilt angle are determined.

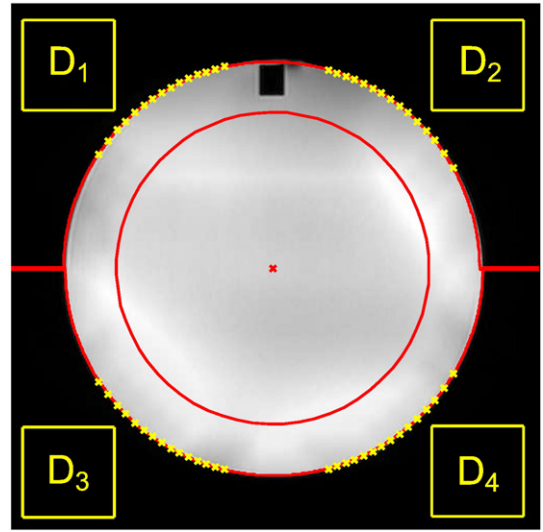


Fig. 4. The central ROI with a radius of 80 % of the phantom radius r is used to determine the image uniformity and (including the four background ROIs D_1, \dots, D_4) the SNR.

(c_x, c_y) is defined, covering 80 % of the phantom radius r according to the ACR protocol (cf. Fig. 4). Taking into account p_{min} denoting the smallest pixel gray value of the ROI and p_{max} the largest one, respectively, **the uniformity U_Σ is defined incorporating span Δ and midrange \bar{S} with**

$$\Delta = \frac{p_{max} - p_{min}}{2} \quad \text{and} \quad \bar{S} = \frac{p_{max} + p_{min}}{2} \quad (2)$$

by

$$U_\Sigma = \left(1 - \frac{\Delta}{\bar{S}}\right). \quad (3)$$

Note, that the image is smoothed with a Gaussian filter ($\sigma = 1.0$) before determining p_{min} and p_{max} to compensate for possible noise. A uniformity $U_\Sigma > 0.82$ for 3 T scanners is considered to pass the quality assurance according to the ACR [8].

2.2.3. Signal-to-Noise Ratio

The determination of the SNR uses the same slice and central ROI utilized in the image-uniformity routine. In addition, **the background deviation is computed using four ROIs D_1, \dots, D_4 located close to the corner of the image** (cf. Fig. 4). The ROIs are defined such that each region equals in size and distance to the phantom center. The SNR is then calculated using the mean of the centered ROI denoted by S_{mean} and the standard deviation σ of the background ROIs by

$$SNR = \frac{S_{mean}}{\sigma_{[D_1, D_2, D_3, D_4]}} \cdot \frac{1}{\sqrt{R}}. \quad (4)$$

The factor $1/\sqrt{R}$ corrects the SNR since parallel imaging was used in the ACR-based IMAGEN sequence, with $R = 2$ denoting the acceleration factor [17]. The acceleration factor is obtained from the DICOM header.

2.2.4. Ghosting Ratio

The ghosting ratio states the amount of artifacts, **mainly foldover artifacts**, superimposing the image. Again the same homogeneous slice and centered ROI are used. Four elongated ROIs are defined in the background of the image, right above and below the phantom as well as on the left and right side, respectively. The space between the

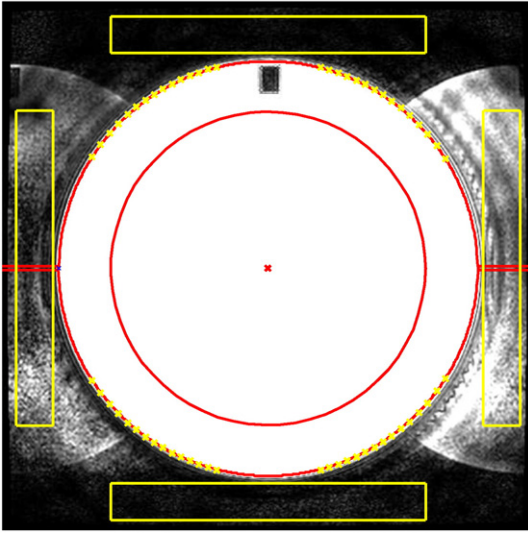


Fig. 5. Four elongated ROIs along with the central ROI are used to determine the level of artifacts surrounding the phantom, yielding the *ghosting ratio*.

ghost-ROIs and the edge of the phantom is equivalent to 2.5 % of the phantom radius r , the size of each ROI is 75% of the phantom diameter tangentially to the phantom edge and between 6% and 10% of the radius r perpendicular to it. Here, the varying dimensions result from the different alignments of the phantom, field of views etc., in the individual data sets (cf. Fig. 5). The ghosting ratio R_g is then calculated using S_{mean} and the mean values M_{top} , M_{bottom} , M_{left} and M_{right} of the four background ROIs by

$$R_g = \frac{|(M_{top} + M_{bottom}) - (M_{left} + M_{right})|}{2 \cdot S_{mean}}. \quad (5)$$

A ghosting ratio threshold of $R_g < 0.025$ is considered to pass the quality assurance according to ACR.

2.2.5. Chemical Shift

The chemical shift evaluation is no longer part of the ACR protocol but is kept as an optional test in this software package. Chemical shift evaluation is determined by the two square structures arranged catty-corner (cf. Fig. 1, D). By detecting position and correct size of both squares in the slice, the shift can be calculated. Therefore, a square sub-image of the slice containing both squares is extracted and filtered with a 3×3 gradient filter, yielding $G(x,y)$ of size $n \times n$. This image $G(x,y)$ is now separated into two images $g_1(x,y)$ and $g_2(x,y)$ of the same size as $G(x,y)$ where, for $g_1(x,y)$ the pixels below the diagonal (from bottom left corner to the top right) are set to 0 and for $g_2(x,y)$ the pixels above are cleared. The k -means clustering ($k = 2$) is applied to both images, yielding two thresholds ϕ_1 and ϕ_2 , respectively, separating the small gradients based on noise from the gradient pixels representing real structures. Both $g_1(x,y)$ and $g_2(x,y)$ are transformed into binary images using the thresholds ϕ_1 and ϕ_2 , setting the pixels to 1 if the gray value is greater ϕ_1/ϕ_2 or to 0 otherwise (cf. Fig. 6). The two images are now convolved with a mask M_g representing the gradient of a square with equal size which is rotated according to the tilt angle α and blurred in x - and y -direction to increase the weighting of the mask pixels close to the expected gradient. The position of the square structure in $g_1(x,y) \in \mathbb{R}^{n \times n}$ is given by

$$(x_{cs}^{(1)}, y_{cs}^{(1)}) = \arg \max_{1 \leq x_{cs}^{(1)}, y_{cs}^{(1)} \leq n} M_g * g_1 \quad (6)$$

and for the second square by $(x_{cs}^{(2)}, y_{cs}^{(2)})$ based on $g_2(x,y)$, i.e., the maximum pixel of the convolved images in g_1 and g_2 represents the position of the optimal match of the square and its mask. Taking into account the size d of the square structures and the pixel size s_{px} , the shift is defined by

$$\Delta x = (x_{cs}^{(2)} - x_{cs}^{(1)} - d) \cdot s_{px} \quad \text{and} \quad \Delta y = (y_{cs}^{(2)} - y_{cs}^{(1)} - d) \cdot s_{py}.$$

The receiving bandwidth is calculated by

$$BW = \frac{FOV_y}{\Delta y} \cdot 3.5 \text{ ppm} \cdot \frac{\gamma B_0}{2\pi} \cdot \frac{1}{N_{fc}}$$

with $\gamma B_0/2\pi$ being the scanner's operating frequency, FOV_y the field of view in y -direction, and N_{fc} the number of acquired frequency columns [8]. Comparing the measured BW value to the pixel bandwidth set in the acquisition sequence should optimally yield a ratio of 1.

2.2.6. Spatial Resolution

The spatial resolution is based on a slice of the phantom showing three small grid structures. The position of each grid inside the slice was stored in a raw template based on an MRI measurement which can be adjusted to the individual data set using the phantom center (c_x, c_y) , radius r , and tilt angle α . In addition, the arrangement of the 31 holes per grid was stored, defining a mask that the holes can be located with. Due to slight geometric distortions and uncertainties of the template or the localization of the phantom, the raw peak positions based on the template are optimized. For each peak, a region covering the whole peak is defined as well as a range of this region, where the peak is expected to be. The whole region is fitted by a 2D-Gaussian (with variable peak position, peak value and variance), yielding the optimized peak position in x and y (cf. Fig. 7, template position (circle), search area (square), adjusted peak (cross)). Along each of the 8 rows and columns per grid, the gray value paths traversing the peaks are extracted to determine the gray value differences between the peaks (created by the holes) and the valleys (representing the background) as well as the distance of the peaks. The differences of the average peak value and the average background value per path in relation to half of the average peak distances (i.e., the peak-valley-distances) yield a simplified representation of the scanner's modulation transfer function (MTF). Subdivided in horizontal and vertical resolution MTF, each consisting of 24 points (8 points from three grids), both

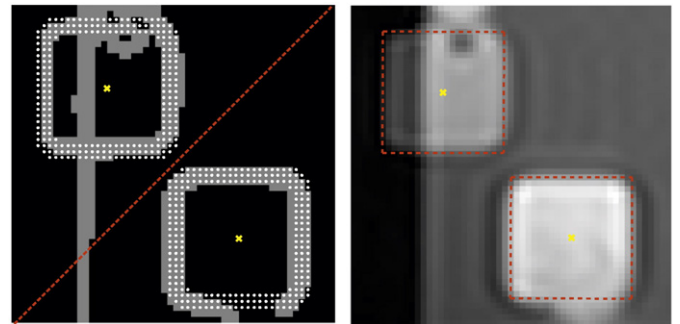


Fig. 6. Left: The diagonally divided binary gradient image (top half representing g_1 , bottom half g_2 ; background in black, segmentation in gray) overlaid with the square template (dotted white) and the resulting square positions (crosses). Right: The resulting location of the gradient template localization of the chemical shift structures in the image.

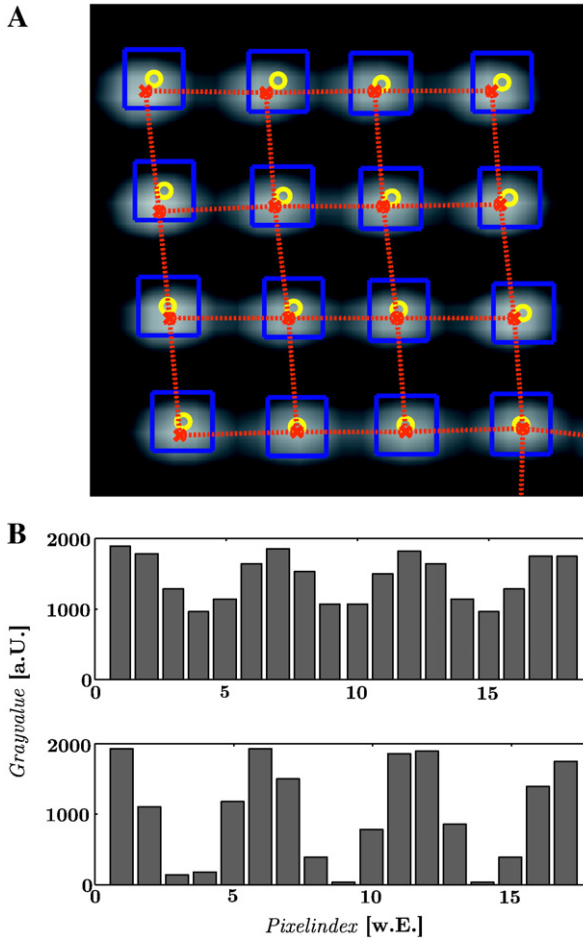


Fig. 7. Left: Extraction of the gray value paths: template positions of the peak (circle), area to be searched using the fit approach (square), adjusted peak positions including the extracted paths (dashed lines). **Right:** An exemplary horizontal (top) and vertical (bottom) gray value path, illustrating an enhanced blur in horizontal orientation (extracted at the right and at the bottom of the grid). Note that the image was acquired at a matrix size of 512×512 pixels for an enhanced insight in the applied methods.

curves of the MTFs are fitted linearly. To obtain the size of the smallest resolvable detail, the fits are intersected with a gray value threshold; in this case a threshold of 5% of the signal maximum was used, i.e., structures with a gray value distance from peak to valley larger than 5% are assumed to be distinguishable. For the signal maximum value we used the average of the signal intensity of the central ROI denoted by S_{mean} in the homogeneous slice that was computed within the SNR evaluation. According to ACR a resolution in both vertical and horizontal direction of less than 1.0 mm should be achieved.

2.2.7. Spatial Linearity

The level of geometric distortions is determined using the slice with the large grid structure (cf. Fig. 1, D). This structure shapes 88 squares, surrounded by the dark edge of the grid whereby width \tilde{w} and area \tilde{A} of each square can be estimated using the radius r by $\tilde{w} \approx 0.14 \cdot r$ and $\tilde{A} \approx \tilde{w}^2$. For all squares, subscripted by $i = 1, \dots, 88$, the expected positions $(\tilde{m}_x^{(i)}, \tilde{m}_y^{(i)})$ are given by a pre-defined template that is adjusted to the individual slice using its localization parameters. Each of the areas is now being segmented using a *Region-Growing* algorithm, with the seed point corresponding to $(\tilde{m}_x^{(i)}, \tilde{m}_y^{(i)})$. Again, the

Table 1

Summary of the five quality setups used in the evaluation; the specific parameter to be tested is kept constant in the different setups and altered ten times for each setup.

Parameter	Distortion		Rel. Error [%]
	min...max	unit	
Translation	0...40	[px]	0.4 ± 0.5
Rotation	0...10	[°]	0.7 ± 1.6
Uniformity	0.6...1.0	[a.U.]	0.7 ± 0.1
SNR	10...1000	[a.U.]	0.5 ± 0.3
Ghosting	0...0.2	[a.U.]	2.2 ± 1.6
Resolution*	0.6...1.2	[mm]	2.6 ± 2.7
Deformation	0.2...1.9	[px]	1.3 ± 1.6
(Chem. Shift)	-1.4... + 9.9	[px]	5.1 ± 2.8

* Evaluation has been performed on high resolution images (4167×4167 px) to compensate for blurring by image rescaling.

segmentation thresholds are computed using the *k-means* clustering applied to a circular ROI with a radius equal to the expected square width \tilde{w} and centered at $(\tilde{m}_x^{(i)}, \tilde{m}_y^{(i)})$. The barycenter of the binary segmentation result, calculated by averaging all x- and y-indices included in the segmented square, determines the accurate position $(m_x^{(i)}, m_y^{(i)})$ of each square as well as its area $A^{(i)}$. All square positions $(m_x^{(i)}, m_y^{(i)})$ with the corresponding area $A^{(i)}$ fulfilling $1/2 \cdot \tilde{A}^{(i)} < A^{(i)} < 3/2 \cdot \tilde{A}^{(i)}$ (i.e., excluding missegmented squares due to misalignment or artifacts) yield a set of points representing the entire grid. A second set representing an ideal grid without any deformations, given by the square positions $(d_x^{(i)}, d_y^{(i)})$, is now overlaid with the measured area positions such that the error in terms of least squares, namely

$$E = \sum_{i=1}^N \left(m_x^{(i)} - d_x^{(i)} \right)^2 + \left(m_y^{(i)} - d_y^{(i)} \right)^2, \quad (7)$$

is minimized. This ideal grid is uniquely defined by the grid center (g_x, g_y) , the tilt angle α_g and the spacing Δ_g between the rows and columns, respectively. The registration of the two point sets is done in two steps, extracting the raw parameters of the corresponding ideal grid from the measured points, and then iteratively matching them in terms of least squares. The raw center (g_x, g_y) of the ideal grid is given by the phantom center (c_x, c_y) . Tilt angle α_g and grid spacing Δ_g are determined by independently fitting each row and column of the measured points $(m_x^{(i)}, m_y^{(i)})$ linearly, yielding 20 fits overall. Hence, the raw tilt angle (taking into account the slopes of all linear fits) and the grid spacing (regarding the distance of one fit to the next one) are determined. The distance of two adjacent linear fits is estimated by the distance of the two intersection points, defined by a line perpendicular to the fits and intersecting the grid center (g_x, g_y) . The average of the overall 18 distances, including 9 in both vertical and horizontal direction, yields the raw grid spacing Δ_g . Since

Table 2

Summary of the testing results: processability states the number of data sets that were acquired sufficiently well to be automatically processed, Criterion fulfilled indicates the number of data sets per parameter fulfilling the performance criterion given by the ACR or further literature.

Parameter	Processability	Performance Criterion	Criterion fulfilled
Uniformity	44/44	> 0.82 a.U.	43/44
SNR	44/44	n.a.	
Ghosting	39/44	< 0.025 a.U.	38/39
Res. (hor.)	44/44	< 1.0 mm	39/44
Res. (vert.)	44/44	< 1.0 mm	42/44
Linearity	43/44	< 2.0 mm	40/43
(Chem. Shift)	44/44	n.a.)	

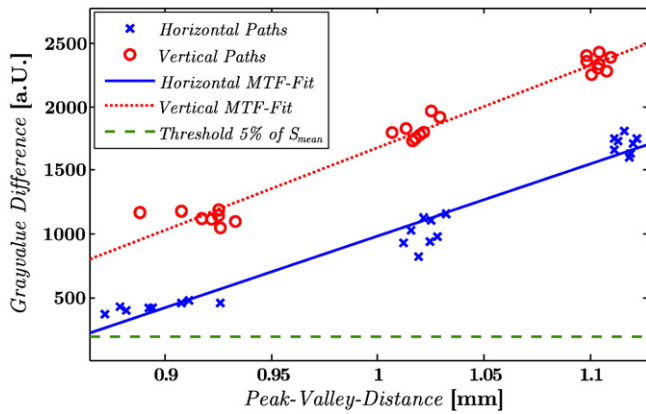


Fig. 8. The resulting MTF of the whole data set including all three resolution grids: the horizontal resolution is detected to be worse compared to the vertical one; the detail detectability is determined using the gray value threshold of 5% of S_{mean} (vertical dashed line).

these four parameters (g_x, g_y), α_g and Δ_g define a suitable ideal grid, but not necessarily the local optimum in terms of least squares, the parameters are now iteratively optimized by applying a *point-point-matching* algorithm. The algorithm performs a similarity transformation of the ideal grid points ($d_x^{(i)}, d_y^{(i)}$), i.e., translations, rotations, and dilations, to optimally match the measured grid points ($m_x^{(i)}, m_y^{(i)}$). Once the least squares error E is minimized, the optimal ideal grid is determined. Consequently, the remaining distance between measured and ideal points indicates the level of geometric distortions caused by the scanner. The greatest error shift of all grid points should be less than 2.0 mm [10].

3. Evaluation

In order to investigate the robustness of all methods, additional routines were implemented to evaluate the accuracy using synthetic data sets affected by predefined distortions. Vector graphics of the phantom slices without any distortions were created and rendered into gray scale images with the appropriate resolution accordingly to the ACR protocol (256×256 px). All types of distortions were artificially incorporated such that the level of degradation could be predicted and compared to the results determined by the automated methods. The following evaluation procedures were implemented:

3.1. Rotation and Translation

The phantom was rotated by an angle α and translated by Δx and Δy pixels, maintaining the original size of the image.

3.2. Image Uniformity

An incline plane P defined by the normal vector \mathbf{n} and a point \mathbf{p}_0 at the center of the phantom is defined, having a value of zero at \mathbf{p}_0 . This plane is multiplied with the normalized phantom image I and afterwards added to the original image in order to keep the background and the phantom center free of gray value changes but introduce a linear gray value gradient when moving towards the phantom edges.

3.3. Signal to Noise

Gaussian noise with mean μ and variance $\sqrt{\sigma}$ is incorporated by first subtracting the mean from the bright areas of the phantom and afterwards adding the noise to the entire image. This maintains the mean gray value in the center of the phantom but creates noise with an appropriate standard deviation. The variance $\sqrt{\sigma}$ of the noise is chosen such that a predefined SNR is established.

3.4. Ghosting Artifacts

The undistorted image is vertically halved, both parts are switched horizontally and recomposed. When added to the original image with certain weighting, the resulting image is superimposed by ghosting artifacts, as commonly known in MRI.

3.5. Chemical Shift

In order to simulate a chemical shift, the two filled squares that the shift evaluation is based on, are shifted by Δx and Δy against each other. The upper square is shifted by $-\frac{1}{2}\Delta x$ and $-\frac{1}{2}\Delta y$ and the lower one by $+\frac{1}{2}\Delta x$ and $+\frac{1}{2}\Delta y$. Note that both positive and negative shifts may occur, yielding squares that may be overlaid to some extent.

3.6. Blurring

The entire image is convolved with a symmetric Gaussian lowpass filter with standard deviation σ . From the characteristics of the filter kernel, the resulting spatial resolution of the blurred images is calculated. In this case, the evaluation is performed on high resolution images (4167×4167 px) instead of the default resolu-

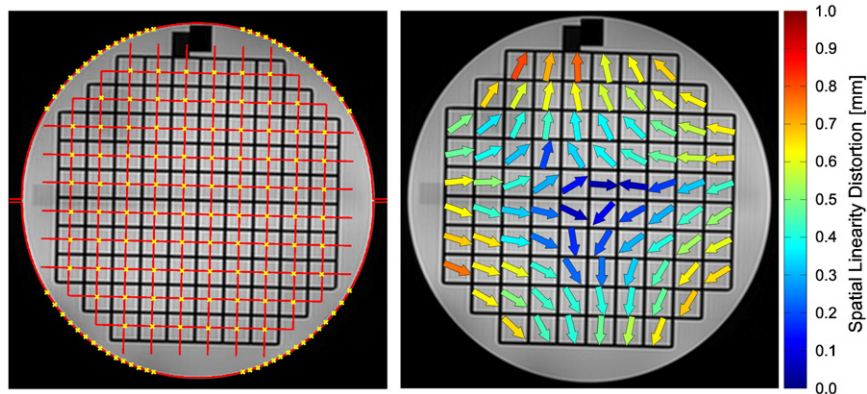


Fig. 9. **Left:** The grid structure with the square centers (crosses in the phantom) and the overlaid ideal grid (solid lines) **Right:** The resulting spatial linearity distortion of each square center from its ideal position in [mm] (arrow points towards the shift).

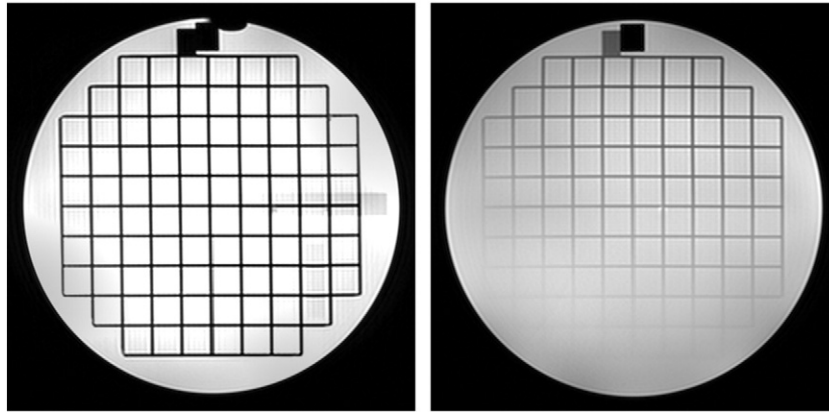


Fig. 10. Left: Evaluable spatial linearity slice: although the vertical block of the previous slice is slightly visible (center right), the grid structure can be extracted accurately. **Right:** Failed spatial linearity slice: the gross misalignment makes the grid structure disappearing at the bottom, preventing its successful extraction.

tion, as the image rescaling already introduces undesirable degradation of the resolution.

3.7. Deformation

To incorporate spatial deformations, a polynomial transformation is applied to the image using a basis $T \in \mathbb{R}^{6 \times 2}$, namely

$$(u \ v) = \begin{pmatrix} 1 & x & y & xy & x^2 & y^2 \end{pmatrix} \cdot T$$

where x and y denote the original pixel location and u and v the pixel in the transformed image. In addition, the deformation is applied to the ideal positions of the linearity grid centers. Thus, the accuracy of the linearity grid extraction based on the deformed images can be evaluated by comparison with the transformed ideal positions.

3.8. Procedure

Since the accuracy of the procedures highly depends on the phantom being located accurately, the corresponding localization method is tested separately for varying levels of translations and rotations. Subsequently, each remaining quality procedure is tested only towards robustness against the corresponding distortion. Note that the phantom is once again being automatically located before proceeding with the quality determination. The respective parameters are tested ten times with varying levels of distortion.

4. Results

4.1. Synthetic Datasets

The results of the evaluation are shown in Fig. 11, for each parameter the estimate of the automated methods is plotted against the synthetic value of the phantom. In addition, a brief summary of the evaluation is given in Table 1. The evaluation shows that our methods are capable of automatically and accurately determining all kinds of quality characteristics based on phantom measurements, even in presence of misalignments. Within all quality parameters, the obtained results only differed by no more than 2.6% from the expected values (excluding the chemical shift which is no longer part of the ACR protocol). Increased relative errors in the estimation of the chemical shift are mainly based on the fact that the two squares need to be localized with subpixel accuracy which is not always practical, especially in presence of blurring.

In contrast to the remaining parameters, the accuracy of the spatial resolution evaluation highly depends on using the specified

matrix size and pixel spacing according to the ACR protocol. Deviations from the protocol leading to higher or lower spatial resolutions introduce errors in the automatic evaluation, as the modulation transfer function can not be approximated sufficiently well in the relevant ranges. (cf. Fig. 11, Spatial resolution: valid range is approximately 0.6,...,1.2 mm).

4.2. IMAGEN Datasets

In addition to synthetic data, the algorithms were tested on a set of 43 measurements acquired in eight different sites within the IMAGEN-project [6], comprising the following 3 T scanner types: Siemens Verio and TimTrio; General Electric Signa Excite, and Signa HDx; Philips Achieva. Additionally one phantom data set was acquired on a 3 T Siemens Skyra resulting in 44 measurements. The processability and accuracy of our automatic quality assurance was similar to the results obtained using the synthetic data (cf. Table 2). Once again, the methods appear to be robust against noise, misalignments of the phantom, and even artifacts. The phantom was localized reliably in all data sets without being influenced by blurring and foldover artifacts, allowing for automatically setting ROIs for SNR and uniformity evaluation as well as the localization of sub-structures in the phantom. Both SNR and uniformity evaluation yield good indicators with regard to the scanner's basic performance (processability 44/44). This includes signal strength, noise level as well as coil and shimming inhomogeneities being present inside the phantom area. The level of foldover artifacts is computed by the ghosting ratio routine illustrated in Fig. 5. However, both SNR/uniformity and ghosting validation require an acceptable positioning of the phantom in the first place, as a sufficiently large area of the background has to be present in the images in order to place the corresponding ROIs (which was not the case in 5 out of the 44 data sets for the ghosting ratio). The chemical shift evaluation yields good

Table 3

Comparison of automatic and manual QC analysis for one data set. For the inner diameter of the phantom, the vertical and horizontal diameter is measured by the manual analysis. For the spatial linearity, the maximal value is given for the automated analysis method. For the spatial resolution, the pixel size that passed the test is given.

QC test	manual	automatic
Inner diameter	18.34 cm, 18.94 cm	19.00 cm
Linearity	0.15 cm	0.10 cm
Uniformity	0.90	0.90
Ghosting	0.004	0.005
Resolution	1.1 mm ² , 1.0 mm ²	1.1 mm ² , 1.0 mm ²

results even if the phantom was poorly aligned during acquisition, leading to other structures of the phantom being visible in the image. Using the gradient template approach used in this work, the two squares were still located accurately in all data sets (cf. Fig. 6). The extraction of the gray value paths along the resolution structures allows quantitative characterization of the real resolution perfor-

mance of the scanner by evaluating the underlying MTF (cf. Figs. 7, 8). Slight deviations of the peak positions (e.g., due to geometric distortions) are compensated by adjusting the peak positions using the Gaussian fit approach (44/44 could be evaluated). The spatial linearity evaluation described in this work yields a detailed characterization of the geometric inhomogeneities including spatial

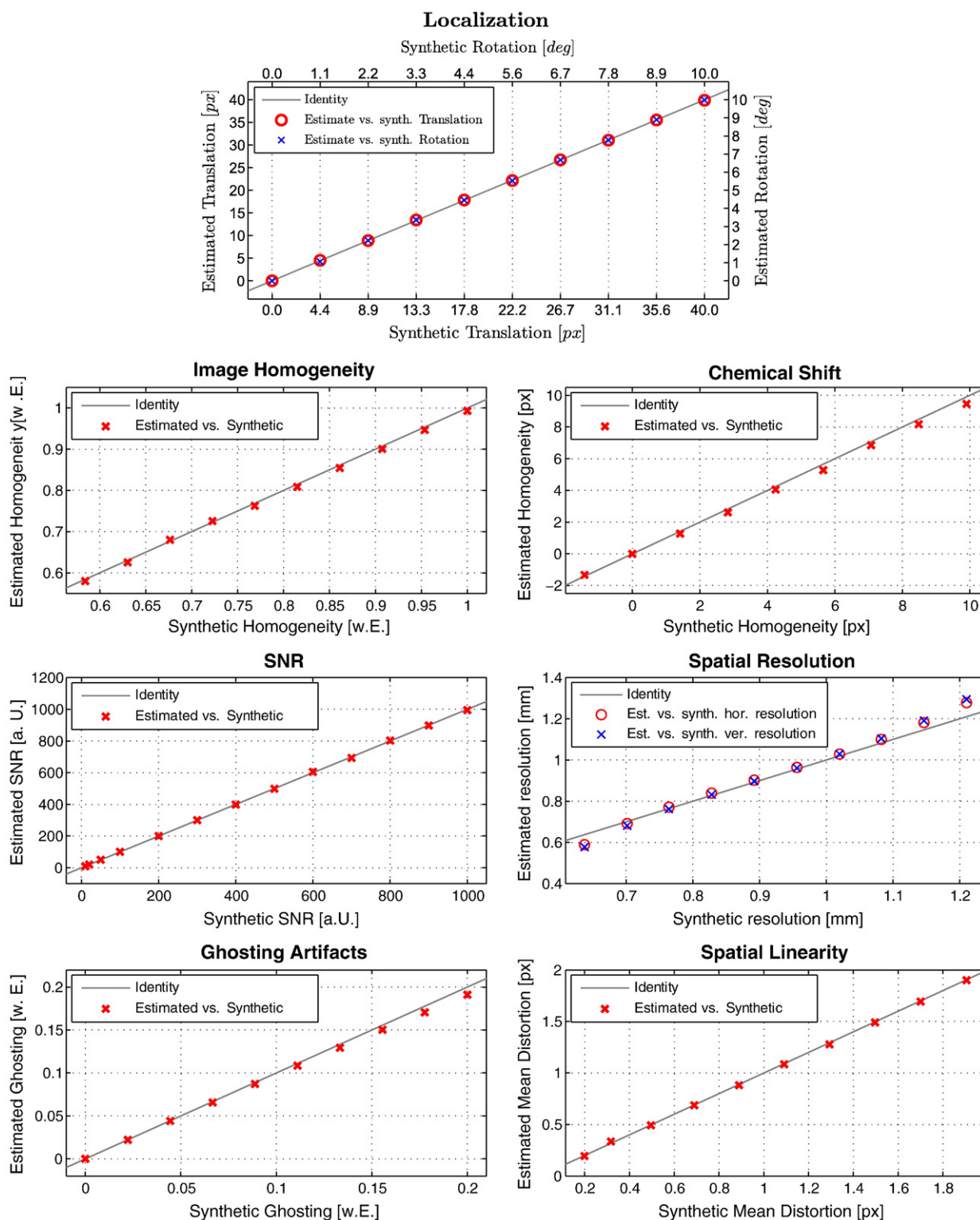


Fig. 11. Evaluation results using synthetic datasets: for each method the estimated quality parameter of the automated procedures is plotted against the predefined ideal value.

dependencies of the distortion level (cf. Fig. 9). Only one data set was not processable due to bad alignment of the slice and thus poor visibility of the grid structure (cf. Fig. 10, showing an average slice (left) and the only slice that was not processable (right)).

Furthermore, the automatic analysis by our software was compared to a quality analysis performed manually according to the ACR guidelines [10] in one data set exemplarily. Table 3 summarizes the results for the specific tests. Overall, our algorithms perform similar to the manual QC analysis. The chemical shift test was not compared as it is no longer part of the ACR guidelines.

5. Discussion

In this work, we implemented methods for fully-automatic evaluation of MRI quality measurements. The results of this work demonstrate that fully-automated evaluation is feasible and, therefore, allows for objective automated processing of large numbers of phantom data sets. Only for two out of seven quality measurements implemented in this study, maximal five out of 44 data sets failed the processing. We observed that both the accuracy of the automated processing and the processability of the data sets itself strictly depend on whether the measurement protocol was followed rather than due to the algorithms. The ability of evaluating the ghosting ratio mainly depends on setting the correct FOV to have sufficiently large areas in the background to extract the foldover artifacts which was not the case in five data sets. Spatial resolution and linearity as well as the chemical shift evaluation require a correct slice positioning, since deviations from the protocol allow for structures from neighboring slices being visible on the slice, thus invalidating the analysis. For large deviations, no processing was possible, whereas small deviations lead to poor results of the quality characteristics. Although the chemical shift evaluation has been removed from the ACR protocol it is kept available in our library to be optionally used.

Our methods were mainly evaluated on synthetic data sets allowing for controlled distortions of image quality. Therefore, our approach gives direct access to whether a test fails or passes correctly and/or whether false positive or negative results are produced. Regarding the results (see Fig. 11), we could show that our methods are sensitive and specific to the tests proposed by the ACR. Since a manual analysis of the QC phantom is regarded as gold standard, we also compared the calculated results of one data set to a manual analysis. The results did not deviate strongly between methods, and differences between methods always lie within the ranges prescribed for fulfilling the ACR test criteria. From this, we conclude that our results are comparable to manual analysis, and additionally exploit the full benefit of automated analysis. While manual analysis is always biased by the reader/operator our computational approach only relies on correct image acquisition. Furthermore, automatic data analysis of all tests on one data set can be performed in less than a minute while manual analysis usually takes several minutes per single test.

Several approaches to process quality measurement data were reported in the past [9–13], however, to the best of the authors' knowledge none of these studies implemented a fully-automatic processing pipeline. As opposed to these manual or partially automated approaches, we expanded the intention of phantom based quality characterization towards full automation, enabling comfortable handling and comparison of a large number of data sets. The library is available free-of-charge and open source allowing for adjustment of methods to meet specific requirements and addition of further algorithms (www.oposumm.de/projects/ummacrqe).

Taking into account performance criteria for each quality parameter one can, in addition to the comparability within one site, define study individual quality requirements. For our data, the

number of data sets that passed the given performance criteria varied. This might be due to different scanner types and slightly varying protocol settings. A detailed analysis of this is currently performed by our group. Also, for one criterion (SNR), no standard performance value was given by the ACR nor found in the literature. Here, the respective study group has to define internal standards to judge on the data quality. However, this is independent of the calculations by our approach.

When analysing QC phantom data one should always check for the correct acquisition of the data. Not following the acquisition protocol and, even more critical, the positioning of the ACR phantom within the scanner can hamper the results, regardless of the analysis (manual or automatically by software tools) being performed. In our implementation, we account for this. If, however, too high a tilt angle is detected during processing of the data set, a warning message is presented to the user.

A limitation of this study is the missing implementation of the test for low contrast object detectability. The reason was that the slices showing the polycarbonate membranes were almost entirely superimposed by artifacts throughout all datasets. Due to these distortions, a robust automated evaluation compared to the other tests was not possible; the contrast resolution processing is, therefore, excluded from this work and will be added in a future version of the library.

In conclusion, we developed a robust tool for fast fully-automated and objective processing of MRI phantom data for quality assurance. In combination with pass criteria, quality measurements within single and multicenter studies can be easily performed. We envision that the comparability of acquired data in such studies might increase, allowing for population based results. The use of this fully-automated evaluation tool within the multicenter project IMAGEN allowed for a comparison of data sets across several sites using scanners from different manufacturers in an easy-to-handle and objective manner, and thus ensured the establishment of a reliable QC data set.

Acknowledgment

This work was supported by the European Union-funded FP6 Integrated Project IMAGEN (Reinforcement-related behavior in normal brain function and psychopathology) (LSHMCT- 2007-037286).

References

- [1] Bartolozzi C, Donati F, Cioni D, Procacci C, Morana G, Chiesa A, et al. Detection of colorectal liver metastases: a prospective multicenter trial comparing unenhanced MRI, MnDPDP-enhanced MRI, and spiral CT. *Eur Radiol* 2004;14:14–20.
- [2] van Haren NE, Cahn W, Pol HEH, Schnack HG, Caspers E, Lemstra A, et al. Brain volumes as predictor of outcome in recent-onset schizophrenia: a multi-center MRI study. *Schizophr Res* 2003;64(1):41–52.
- [3] Seiderer J, Zech CJ, Reinisch W, Lukas M, Diebold J, Wrba F, et al. A multicenter assessment of liver toxicity by MRI and biopsy in IBD patients on 6-thioguanine. *J Hepatol* 2005;43(2):303–9.
- [4] Lee VS. MRI: From science to society. *J Magn Reson Imaging* 2013;37(4):753–60.
- [5] Pruem H, Gerigk L, Hintze C, Thieke C, Floca R. Software-guided standardization of manual landmark data in medical images. *Z Med Phys* 2011;21(1):42–51.
- [6] Schumann G, Loth E, Banaschewski T, Barbot A, Barker G, Buechel C, et al. The IMAGEN study: reinforcement-related behaviour in normal brain function and psychopathology. *Mol Psychiatry* 2010;15(12):1128–39.
- [7] Friedman L, Glover GH. Report on a multicenter fMRI quality assurance protocol. *J Magn Reson Imaging* 2006;23(6):827–39.
- [8] Dixon R, Butler P, Sobol W, editors. Accreditation Programs and the Medical Physicist, Vol. 27 of AAPM Monograph, Medical Physics Publishing; 2001.
- [9] Mulkern RV, Forbes P, Dewey K, Osganian S, Clark M, Wong S, et al. Establishment and results of a magnetic resonance quality assurance program for the pediatric brain tumor consortium. *Acad Radiol* 2008;15(9):1099–110.
- [10] Ihalaainen TM, Loennroth NT, Peltonen JJ, Uusi-Simola JK, Timonen MH, Kuusela LJ, et al. MRI quality assurance using the ACR phantom in a multi-unit imaging center. *Acta Oncol* 2011;50(6):966–72.

- [11] Fu L, Fonov V, Pike B, Evans A, Collins D. Automated analysis of multi site MRI phantom data for the NIHPD project. In: Larsen R, Nielsen M, Sporring J, editors. Medical Image Computing and Computer-Assisted Intervention - MICCAI 2006, Vol. 4191 of Lecture Notes in Computer Science. Heidelberg: Springer Berlin; 2006. p. 144–51.
- [12] Chen C-C, Wan Y-L, Wai Y-Y, Liu H-L. Quality assurance of clinical MRI scanners using ACR MRI phantom: preliminary results. *J Digit Imaging* 2004;17:279–84.
- [13] Magnotta VA, Friedman L, FIRST BIRN. Measurement of signal-to-noise and contrast-to-noise in the fBIRN multicenter imaging study. *J Digit Imaging* 2006;19:140–7.
- [14] American College of Radiology. Phantom Test Guidance; 2005.
- [15] MacQueen JB. Proceedings of the fifth berkeley symposium on mathematical statistics and probability: some methods for classification and analysis of multivariate observations, no. Bd. 2. University of California Press; 1967.
- [16] Price RR, Axel L, Morgan T, Newman R, Perman W, Schneiders N, et al. Quality assurance methods and phantoms for magnetic resonance imaging: report of AAPM nuclear magnetic resonance Task Group No. 1. *Med Phys* 1990;17(2): 287–95.
- [17] Dietrich O, Raya JG, Reeder SB, Reiser MF, Schoenberg SO. Measurement of signal-to- noise ratios in MR images: influence of multichannel coils, parallel imaging, and reconstruction filters. *J Magn Reson Imaging* 2007;26(2): 375–85.