# The Relative Impact of Ghosting and Noise on the Perceived Quality of MR Images

Hantao Liu, *Member, IEEE*, Jos Koonen, Miha Fuderer, and Ingrid Heynderickx

*Abstract*—Magnetic resonance (MR) imaging is vulnerable to a variety of artifacts, which potentially degrade the perceived quality of MR images and, consequently, may cause inefficient and/or inaccurate diagnosis. In general, these artifacts can be classified as structured or unstructured depending on the correlation of the artifact with the original content. In addition, the artifact can be white or colored depending on the flatness of the frequency spectrum of the artifact. In current MR imaging applications, design choices allow one type of artifact to be traded off with another type of artifact. Hence, to support these design choices, the relative impact of structured versus unstructured or colored versus white artifacts on perceived image quality needs to be known. To this end, we conducted two subjective experiments. Clinical application specialists rated the quality of MR images, distorted with different types of artifacts at various levels of degradation. The results demonstrate that unstructured artifacts deteriorate quality less than structured artifacts, while colored artifacts preserve quality better than white artifacts.

*Index Terms*—MR, perceived image quality, ghosting, noise, human visual system.

## I. INTRODUCTION

**M**AGNETIC resonance (MR) imaging is a powerful and widely used clinical imaging modality, which is able to visualize detailed internal structures of the human body [1], [2]. It has some unique advantages over other imaging technologies, such as a high spatial resolution and high soft tissue contrast. Perhaps most importantly, unlike computed tomography (CT) scans or traditional X-rays, MR imaging does not rely on ionizing radiation, and therefore, it is safer for serial examinations, dynamic imaging studies and screening in asymptomatic subjects.

As with any other imaging modality, MR imaging is unfortunately vulnerable to artifacts, which potentially degrade the quality of images, and consequently, may cause inefficient and/or inaccurate diagnosis, may impact the visual search efficiency of the clinical specialists, and as such, may affect

their workflow [13], [14]. Sources of artifacts in MR imaging are non-ideal hardware characteristics, intrinsic tissue properties and their possible changes during scanning, assumptions underlying the data acquisition and image reconstruction process, and a poor choice of scanning parameters [3]–[6]. To minimize or eliminate these artifacts, many correction procedures have been developed. These methods typically involve one or more of the following strategies: improvement of hardware and scanning protocols, scan parameter and pulse sequence optimization, and advanced post-processing algorithms [7]–[10]. Nonetheless, reducing artifacts in MR imaging is not straightforward, and so far, the existing approaches hardly achieve an optimal image rendering from the user's point of view [11], [12]. One reason is that strategies coping with one type of artifact may induce another type of artifact. As a consequence, optimization of these strategies requires knowledge of the relative annoyance of different types of artifacts to diagnostic quality.

Due to the complexity of the human visual system (HVS) as well as the clinical task, measuring image quality in diagnostic imaging is not trivial [15]–[18]. In the literature, studies evaluating medical images typically concentrate on the diagnostic performance (i.e. on the errors made in diagnostic analysis) rather than on perceived image quality (i.e. rating the quality of an image without a direct or specific detection task involved). As such, subjective assessments in these studies are usually based on the receiver operating characteristic (ROC) method, in which images are assessed in terms of the ability of human observers to detect a disease phenomenon, i.e. to classify patients as "positive" or "negative" with respect to any particular disease [19]–[24]. The ROC method is very important, but does not take into account the quality of visual experience with which lesions or diseases are detected. Even at worse image quality, detection performance may still be good, though at higher perceptive and cognitive load of the assessor. As such, perceived image quality could be considered as an informative measure related to diagnostic performance. In addition, ROC is usually measured for a specific (detection) task; yet, MR imaging can produce a large variety of image contrasts, each with a large variety of clinical questions and each clinical question may link to a rather wide variety of relevant image-patterns. Hence, perceived image quality is considered as a measure that addresses performance averaged over all these (potential) diagnostic questions, and therefore is more and more studied in medical imaging [40]–[42]. Research linking perceived image quality to diagnostic performance, however, is very limited. Fundamental issues such

as how the measured differences in image quality affect the diagnostic performance remain largely unexplored, while this knowledge would be highly beneficial for further improvement of diagnostic imaging systems and clinical applications. This paper aims at contributing to this discussion through a better understanding and modelling of perceived image quality.

The increasing ubiquity of artifacts in MR images has pushed the demand for better image quality assessment and control strategies [3]–[5]. Progress has been made in studying the causes and characteristics of the artifacts, and in classifying these artifacts so that they can be recognized from relevant features [11], [12]. In general, the artifacts may be classified into two categories: unstructured and structured artifacts. In this paper, an unstructured artifact is referred to as random noise, while a structured artifact is defined as any type of coherent artifact that represents the anisotropy of the spectral content of local structure of the object being scanned. Ghosting, which is a cross-talk type of artifact generating a lower-intensity double image, spatially shifted with respect to the original content, is one example of such a structured artifact [7], [8]. In addition, random noise can be further classified into white noise and colored noise, according to its spectral density: white noise has a flat frequency spectrum, whereas the frequency spectrum of colored noise is not flat. A similar classification can be made for the structured artifacts; i.e., we can make a distinction between a white structured artifact and a colored structured artifact. The example of ghosting – explained above – can be considered as a colored structured artifact with the same distribution in the frequency spectrum as the object being scanned. "White" ghosting – in the rest of the paper referred to as edge ghosting – can be obtained by making the frequency spectrum flatter, e.g., by adding the gradient of the originally scanned object as a kind of double image to the original. Based on the analysis of the power spectral density of a thin-slice two-dimensional MR image [26], it can indeed by shown that the power spectrum of the gradient of any line of an MR image is rather flat, and so, for the purpose of our study is approximated to "white". Figure 1 illustrates the four types of artifacts on an exemplary MR image.

These four types of artifacts are of high practical relevance. White random noise is very common (or almost omnipresent) in MR imaging. Plain (colored) ghosting frequently occurs whenever a periodic disturbance in MR data acquisition (e.g., periodic object motion or a 60Hz main-field fluctuation) has a different frequency than the basic acquisition pace (i.e., the repetition time of the MR measurement). In real-world MR imaging, the source of colored random noise has been explained in [27] or [29]; real-world edge ghosting can occur with the Echo-Planar Imaging (EPI) sequence.

In current MR imaging systems, there are circumstances whereby one type of artifact may be traded off with another type of artifact, e.g. the trade-off between a structured artifact and an unstructured artifact [28]–[30]. A generally known trade-off in MR, among many, is the change of the receiver (or acquisition) bandwidth, which has a direct relationship to the signal to noise ratio (SNR). A smaller bandwidth improves SNR, but can cause spatial distortions (e.g. a fat tissue shift
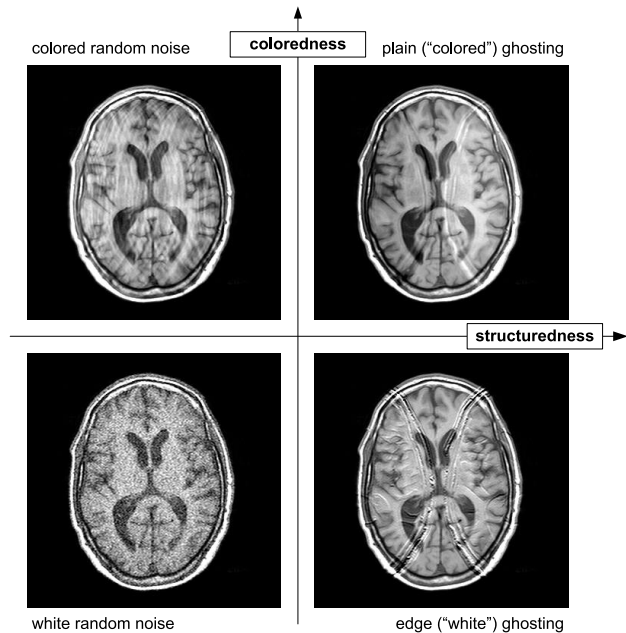


Fig. 1. Illustration of the four types of artifacts (addressed in this paper) in a typical MR image. The horizontal axis indicates the "structured-ness" of the artifact: the two left quadrants refer to the unstructured artifacts (i.e. random noise), and the two right quadrants refer to the structured artifacts (i.e. ghosting). The vertical axis indicates the colored-ness of the artifact: the two top quadrants refer to the colored artifacts, and the two bottom quadrants refer to the white artifacts.
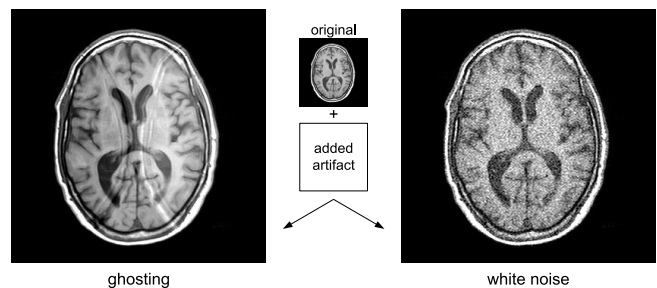


Fig. 2. Illustration of the perceived image quality for an MR image degraded by the same amount of energy in ghosting and white noise.

to a wrong position, acting like a ghosting artifact); a larger bandwidth reduces SNR, but allows faster imaging. In such a scenario, a balance between noise and spatial distortions needs to be taken into account towards a better perceived image quality. To optimize this trade-off, one needs to know the relative impact of the artifacts on diagnostic quality, and therefore, we focus in this paper on the relative impact of the various types of artifacts on perceived image quality. To what extent a given artifact present with a certain energy reduces the perceived image quality is unknown so far. It is often assumed that the energy of the artifact, being a physical measure of its signal strength, and typically defined as the mean squared value of the intensity of the artifact with respect to the background is a good measure for the perceived image quality. The latter may be the case when mutually comparing artifacts of one type, but not when comparing artifacts of different types. As illustrated in Figure 2, an MR image degraded to a certain energy amount of ghosting may have a different
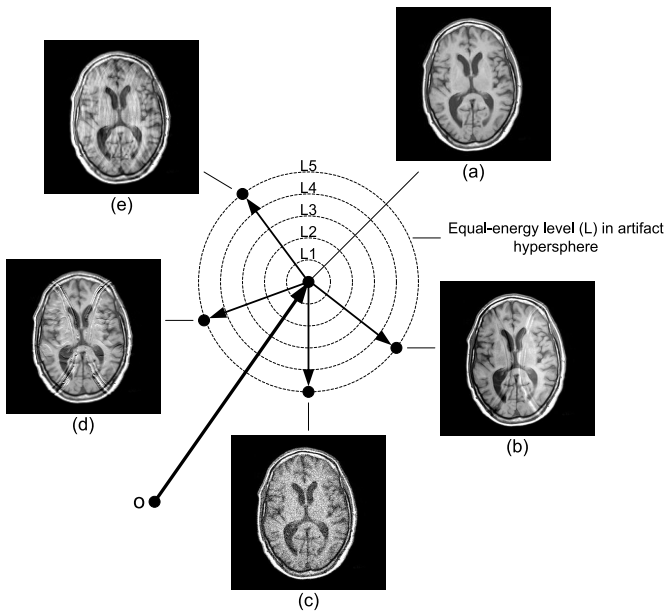
Fig. 3. Images with the same level of energy in the artifact added to the original image constitute a hypersphere in the image space: (a) original image, (b) image with ghosting, (c) image with white noise, (d) image with edge ghosting, and (e) image with colored noise. Five different levels of energy, i.e. L1, L2, L3, L4, L5, are used in our experiments.



Fig. 4. Illustration of the simulation of ghosting and white noise in an MR image.

perceived image quality compared to the same image degraded to the same energy amount of white noise. Hence, to optimize the trade-off between artifacts, it is necessary to understand the relative annoyance of the artifacts in terms of perceived image quality. This aim is addressed in this paper by measuring the relative impact of four types of artifacts: a white unstructured artifact (i.e., white noise), a colored unstructured artifact (i.e., colored noise), a white structured artifact (i.e., edge ghosting) and a colored structured artifact (i.e., ghosting). In order to have full control on the occurrence of the artifacts, they are simulated on top of original clean MR images.

## II. SIMULATION OF THE ARTIFACTS

To be able to vary the four types of artifacts – namely ghosting, edge ghosting, white noise, and colored noise – in a controlled way, they were simulated separately at different levels of energy, and then linearly added to the original image content, as illustrated in Figure 3. As a start, a benchmark energy level (BEL) was defined (illustrated as the energy level L5 in Figure 3). For an original image of size M × N (height × width) pixels with intensity of the simulated ghosting artifact $I_g(i, j)$ ($i \in [1, M]$, $j \in [1, N]$), we calculated BEL as:

$$BEL = \sum_{i=1}^{M} \sum_{j=1}^{N} I_g(i, j)^2 \tag{1}$$

The BEL was determined separately for each original image, and was defined by the amount of energy in a typical ghosting artifact for that particular content. As such, a ghosting artifact was always generated first in the simulation process. Based on the BEL defined for ghosting, the other levels of energy
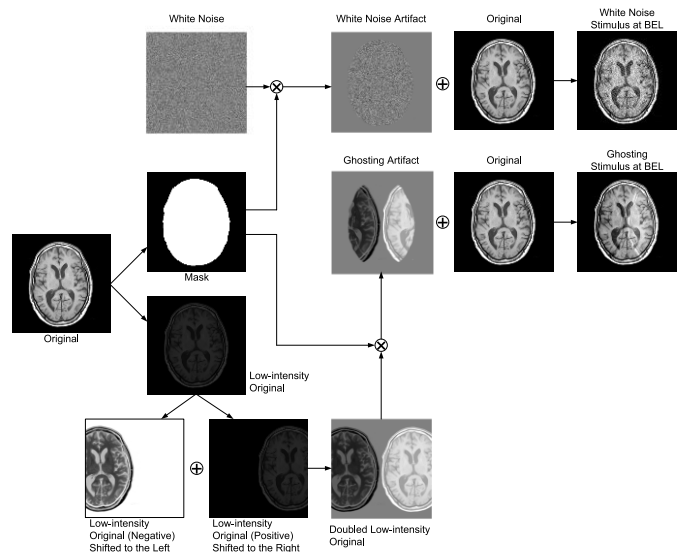
were determined by reducing the BEL successively with 20% (i.e., resulting in 0.8 × BEL for energy level L4, 0.6 × BEL for energy level L3, 0.4 × BEL for energy level L2 and 0.2 × BEL for energy level L1). It should be noted that all artifacts in our experiment are only applied to the area of the anatomical object (while in practice they most often extend over the whole image area). We intentionally applied a binary mask MI (obtained by thresholding the original image at the level of 5% of the maximal intensity, see also Figure 4) on the anatomical object in order to have the viewer's attention to that object, rather than that the presence of artifacts could be deduced from examining the background. The artifacts were simulated in a way to ensure as high realism as possible. The degree of simulation realism was subject to scrutiny by a few experts (i.e., MRI experts from Philips Healthcare and radiologists from a local hosptial), and judged to be appropriate.

### A. Ghosting

As illustrated in Figure 4, the simulation of ghosting was based on two new images generated from the original image: (1) the binary mask image MI representing the area of the clinical object and (2) a lower-intensity (LI) version of the original image, chosen to have 20% of the intensity of the original image when simulating the artifact at level L5. The LI image was then spatially shifted with respect to the original content, once to the left with negative intensity values (to simulate a negative intensity ghosting) and once to the right with positive intensity values (to simulate a positive intensity ghosting). The distance of this spatial shift was a constant in our simulations, and defined as 1/3 of the image width. Although we realize that a more common distance for ghosting in MR is half of the field of view, we here choose a distance of 1/3 in order to create a more substantial overlap with the object itself. The whole operation procedure resulted in a new (low-intensity) image in which the clinical object was doubled. Then the
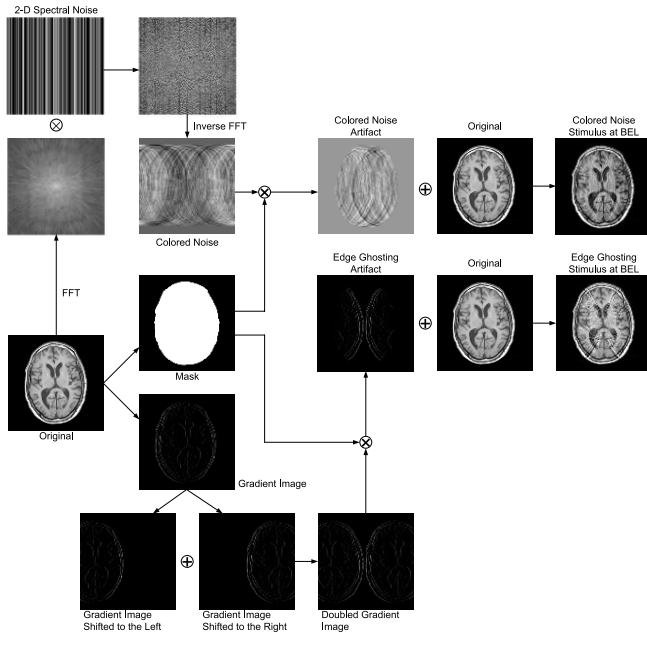
Fig. 5. Illustration of the simulation of edge ghosting and colored noise in an MR image.

so-called ghosting artifact image (i.e., $I_g$) was generated by a pixel-by-pixel multiplication of this new image with the binary mask. Adding this ghosting artifact image $I_g$ to the original image I yielded the test stimulus distorted with ghosting at the L5 energy level, which is shown in Figure 3(b). Subsequently, the amount of energy calculated for $I_g$ was considered as the BEL, and was used as the maximal energy level to simulate the other types of artifacts.

### B. White Noise

As illustrated in Figure 4, the simulation of white noise was also based on two new images: (1) the binary mask image MI and (2) an image containing additive white Gaussian noise (i.e., with a flat frequency spectrum) having the same size as the original image. Both images were multiplied pixel by pixel in order to generate the white noise artifact image (i.e., $I_{wn}$). The intensity of the white noise artifact image was scaled such that the resulting total energy of noise for the L5 energy level was equal to the BEL. The resulting stimulus distorted with white noise was generated by adding the white noise artifact image to the original image I, the result of which is shown in Figure 3(c).

### C. Edge Ghosting

Edge ghosting was simulated in basically the same way as normal ghosting, but now with a gradient image of the original spatially shifted with respect to the original content. As illustrated in Figure 5, the simulation of edge ghosting was also based on two new images generated from the original image: (1) again the binary mask image MI and (2) a gradient image (GI) of the original, calculated as:

$$GI(i, j) = |I(i, j + 1) - I(i, j)|, \ j \in [1, N - 1] \quad (2)$$

The gradient image GI was spatially shifted to the left and to the right with respect to the original content. Again the distance of this spatial shift was kept constant at a value of 1/3 of the image width. This procedure resulted in a new gradient image, in which the clinical object was doubled. Then the so-called edge ghosting artifact image (i.e., $I_{eg}$) was generated by a pixel by pixel multiplication of $I_{eg}$ with the mask image MI. The intensity of the edge ghosting artifact image $I_{eg}$ was then scaled such that its resulting total energy was equal to the BEL. Adding this $I_{eg}$ to the original image I yielded the test stimulus distorted with edge ghosting at the L5 energy level, which is shown in Figure 3(d).

### D. Colored Noise

Finally, also the simulation of colored noise was based on two images: (1) the binary mask image MI and (2) an image with colored noise. As illustrated in Figure 5, the colored noise was generated in the Fourier transform domain, to which a 2D spectrum with random values (i.e., complex white Gaussian) in the vertical direction and with constant values in the horizontal direction (i.e., the direction of (the Fourier-transform of) the left-right direction of the patient) was multiplied. The inverse Fourier transformation of the resulting spectrum yielded an image of "colored noise" in the sense that the power spectral density of the resulting noise pattern is identical to that of the MR image itself. Strictly speaking, the resulting pattern is not fully "unstructured", but it is more "unstructured" than plain ghosting, since any object structure is now spread out over the whole image field. The resulting image was multiplied pixel by pixel with the mask image MI to yield a colored noise artifact image (i.e., $I_{cn}$). The intensity of $I_{cn}$ was then scaled such that the resulting total energy of colored noise was equal to the BEL. The resulting stimulus distorted with colored noise at the energy level L5 was then generated by adding $I_{cn}$ to the original image I, and is shown in Figure 3(e).

## III. SUBJECTIVE EXPERIMENTS

The ultimate goal of this study is to quantitatively measure how the four types of artifacts, applied at the same energy level in the distortion affect the perceived quality of MR images. To this end, two perception experiments were performed with clinical application specialists (mainly qualified clinical medical physicists). The source MR images used in these two experiments were collected from Philips Achieva 1.5T images, and selected to be of high quality in terms of resolution, artifacts and signal-to-noise ratio. Although the images were from patients, we purposely selected images showing no apparent pathology, nor was the viewer informed about the indications for the MR scan. The rationale behind this choice was to make the viewer consider all plausible clinical uses of the images, rather than to concentrate on one specific pathology with one specific aim on contrast, lesion size, or conspicuity. Unless otherwise mentioned, the images were from axial multi-slice Fast Spin Echo scans with 90 degree RF excitation.
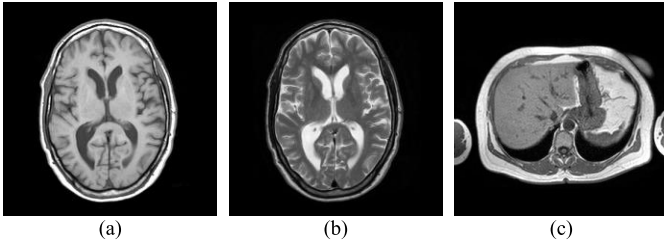
Fig. 6. Source images, which are referred to as: (a) "brain_1", (b) "brain_2", and (c) "liver".
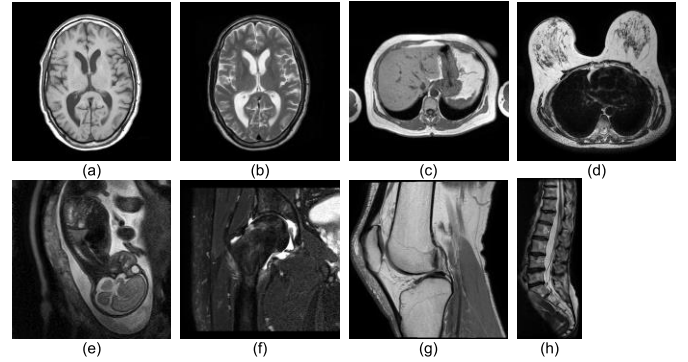


Fig. 7. Source MR images used in Experiment 2, which are referred to as: (a) "brain_1", (b) "brain_2", (c) "liver", (d) "breast", (e) "fetus", (f) "hip", (g) "knee", and (h) "spine".

## A. Experiment 1

The goal of Experiment 1 was to investigate the relative impact of structured versus unstructured artifacts on the perceived quality of MR images. The experiment consisted of two parts: one to compare images degraded with ghosting to images degraded with white noise, and one to compare images degraded with edge ghosting to images degraded with colored noise.

For this experiment we selected three original MR images: two images of a brain ("brain_1": T1-weighted brain, plain Spin-Echo, TR = 650, TE = 15, RF excitation = 69 degrees, 2 signals averaged, SENSE-head-8 coil, and nominal voxel size = 0.72*0.72*5mm; "brain_2": T2-weighted brain, TR = 4877, TE = 100, 3 signals averaged, Echo train length 15, SENSE-head-8 coil and nominal voxel size = 0.47*0.47*5mm) and one image of a liver (Field Echo liver, TR = 117, TE = 4.6, RF excitation = 80 degrees, 2 signals averaged, SENSE-torso-XL coil and nominal voxel size = 1.3*1.3*5mm). The three source images are shown in Figure 6. Each source image was first distorted with ghosting at the energy level BEL, and subsequently, edge ghosting, white noise and colored noise were applied at the same energy level. The related procedure is detailed in section II and Figures 4 and 5. The added energy (i.e. at BEL) of ghosting, edge ghosting, white noise and colored noise was then downscaled with factors of 4/5, 3/5, 2/5, 1/5, resulting in four new energy levels for each artifact type (as explained already in more detail in section II). By doing so, each original image was distorted with 5 levels of simulated ghosting, edge ghosting, white noise and colored noise, respectively. Hence, the test database of this experiment existed of 30 stimuli (i.e. 3 originals × 5 energy levels × 2 types of artifacts) per part, and so 60 stimuli in total.

## B. Experiment 2

The goal of Experiment 2 was to extend the findings of Experiment 1 for a larger diversity in image content, but then for just two energy level to limit the total amount of experimental time to reduce fatigue effects. To this end, we used (next to the 3 original images used in Experiment 1) 5 additional MR source images (as shown in Figure 7): "breast" (T2-weighted mammo, TR = 6107, TE = 120, 2 signals averaged, Echo train length 25, 4-elements SENSE-body coil, and nominal voxel size = 0.74*0.74*3mm), "fetus" (Proton-density-weighted single-shot image, TE = 140.

5-elements SENSE-cardiac coil, and nominal voxel size = 0.9*0.9*4mm), "hip" (T2-weighted coronal hip with SPAIR fat-suppression, TR = 2760, TE = 60, 4 signals averaged, Echo train length 20, 4-elements SENSE-body coil, and nominal voxel size = 0.31*0.31*3.5mm), "knee" (Proton-density-weighted image, TR = 5000, TE = 30, 2 signals averaged, Echo train length 11, 8-elements SENSE-knee coil, and nominal voxel size = 0.3*0.3*2.5mm), and "spine" (T2-weighted sagittal spine, TR = 3255, TE = 120, 6 signals averaged, Echo train length 22, 5-elements SENSE-spine coil, and nominal voxel size = 0.52*0.52*4mm). These images were then degraded with the four types of artifacts (following the same procedure as described in Section II and Figures 4 and 5), but just at the two most extreme energy levels (i.e., L1 and L5).

The findings of Experiment 1, however, demonstrated some problems with the simulation of the colored noise artifact. When simulating colored noise as illustrated in Figure 5, a 2D spectrum with random values in the vertical direction and constant values in the horizontal direction was generated. The resulting noise pattern, however, was affected by the randomization, and consequently, perceived quality depended on the specific pattern. To validate this hypothesis, we ran a small pilot experiment with five different versions of colored noise (i.e., five different random generations of the 2D spectrum) applied at the same energy level. The resulting stimuli degraded with colored noise, as shown for one particular original image content as an example in Figure 8, were judged to be significantly difference in perceived quality by a few clinical application specialists. Hence, the randomization procedure for simulating colored noise needed to be taken into account when evaluating perceived quality. To this end, we used in Experiment 2 four different versions of colored noise. As a consequence, this experiment contained 112 stimuli (i.e. 8 originals × 2 energy levels × 7 distortion versions) in total.

## C. Test Environment and Participants

The experiments were conducted in a standard office environment [31] (adjusted as close as possible to a typical radiology reading room environment [36]) at
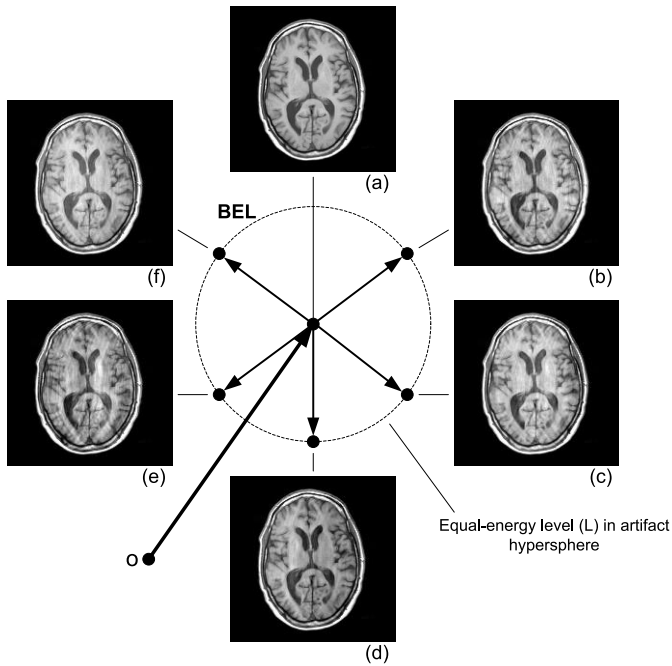
Fig. 8. The same original image (a) degraded by colored noise at the energy level BEL, generated five times with a different randomization procedure, resulting in five "different" stimuli (b)-(f).
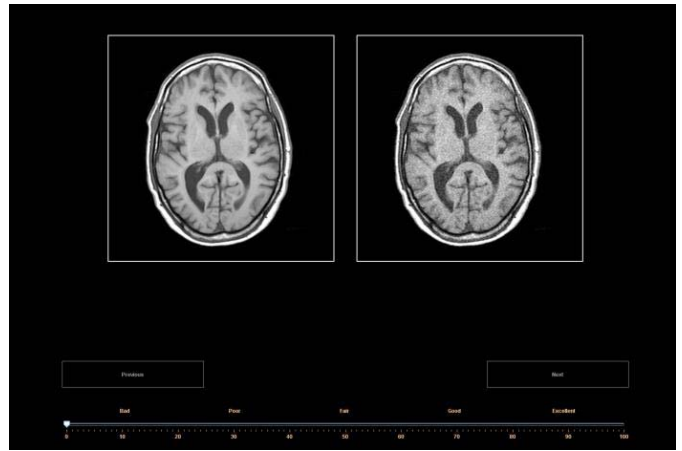


Fig. 9. Illustration of the rating interface used during the experiment, including two stimuli, i.e. the reference at the left-hand side and the test stimulus at the right hand side, and the quality scale.

Philips Healthcare, Best, in the Netherlands. The venue represented a controlled viewing environment to ensure consistent experimental conditions: low surface reflectance and approximately constant ambient light (i.e., with an indirect horizontal illumination of 100 lux). The stimuli were displayed on a Philips 24" wide-screen liquid-crystal display with a native resolution of $1920 \times 1200$ pixels, which was calibrated to the Digital Imaging and Communications in Medicine (DICOM): Grayscale Standard Display Function (GSDF) [37]–[39]. The viewing distance was approximately 60 cm. No image adjustment (zoom, window level) was allowed.

Since the impact of professionalism on the perceived quality of MR images was proved significant [32], the perception experiments were conducted with clinical application specialists. All participants were recruited from clinical scientists or application specialists at Philips Healthcare, Best, The Netherlands. The first part of Experiment 1 (i.e., scoring of images degraded with ghosting and white noise) was performed by 15 participants, being 10 males and 5 females. The second part of this first experiment (i.e., scoring of images degraded by edge ghosting and colored noise) was performed by 17 participants, being 11 males and 6 females. Finally, for Experiment 2 18 participants, being 10 males and 8 females, were recruited.

### D. Experimental Protocol

To score perceived quality we used a simultaneous-double-stimulus (SDS) method [31]; subjects were requested to score the quality for each stimulus in the presence of the original image as a reference. The rating interface is illustrated in Figure 9; the two stimuli, i.e. the original at the left-hand side and the test stimulus at the right-hand side were displayed side by side on the same screen. The scoring scale ranged from 0 to 100, and included additional semantic labels (i.e. "Bad", "Poor", "Fair", "Good" and "Excellent") at intermediate points. Subjects were requested to assess the quality of the test stimulus with respect to the quality of the reference by moving the slider on the scoring scale.

Before the start of each experiment, a written instruction about the procedure of the experiment (i.e. explaining the type of assessment, the scoring scale and the timing) was given to each individual subject. Subsequently, a set of ten images covering the same range of artifact annoyance as used in the actual experiment was presented to each subject in order to familiarize him or her with the impairments used and with how to use the range of the scoring scale. In a next step, six representative stimuli were shown one by one and the participant was asked to score their quality on the scoring scale. The images used in this training part of the experiment were different from those used in the actual experiment. After training, the test stimuli were shown one by one in a different random order to each participant in a separate session. Each stimulus was just shown once, and the participants could take as much time as they needed to assess the quality of each stimulus.

## IV. EXPERIMENTAL RESULTS

### A. Processing of the Raw Data

First, a simple outlier detection and subject exclusion procedure [33]–[35] was applied to the raw scores.[1] An individual score for an image was considered to be an outlier if it was outside an interval of two standard deviations around the mean score for that image. All scores of a subject were rejected if more than 20 percent of his/her scores were outliers. Overall, in Experiment 1 for the scoring of ghosting and white noise, one subject (out of 15) was excluded from further analysis, and only 2 of the remaining scores were rejected

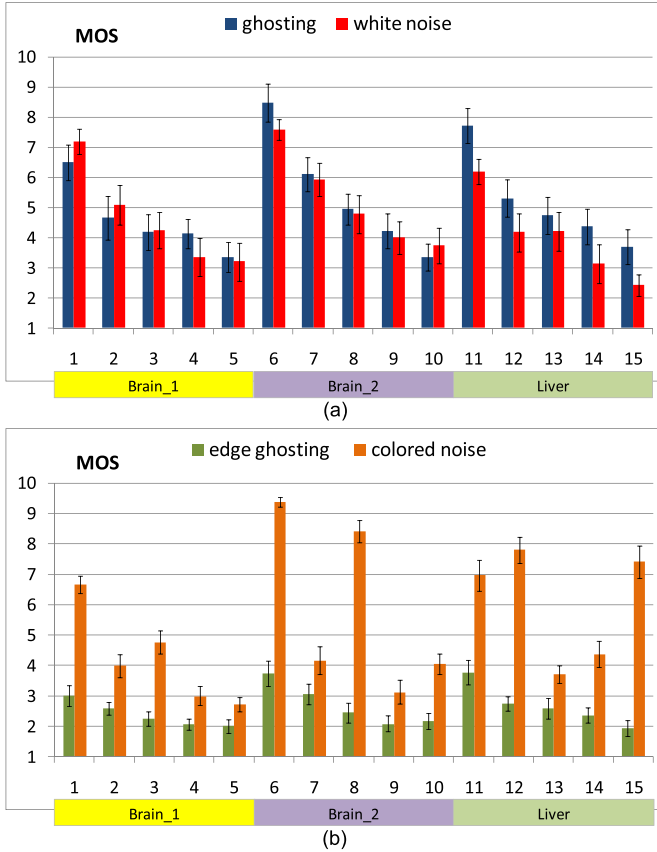[1]The data are made publicly available to the research community [43].

Fig. 10. The MOS resulting from the subjective image quality assessment: (a) images degraded with ghosting and white noise, and (b) images degraded with edge ghosting and colored noise. The numbers on the horizontal axis refer to the stimuli: numbers 1-5 for image "brain_1" with increasing level of distortion, numbers 6-10 for image "brain_2", and numbers 11-15 for image "liver". Each number corresponds to two bars; one for ghosting (or edge ghosting) and one for white noise (or colored noise), with each the same energy in the signal distortion. The error bars indicate the 95% confidence interval.

TABLE I
RESULTS OF THE ANOVA FOR EXPERIMENT 1 TO EVALUATE THE EFFECT OF GHOSTING AND WHITE NOISE ON THE DIFFERENT IMAGES

|  | F-value | df | P |
|---|---|---|---|
| Image content | 5.82 | 2 | 0.003 |
| Artifact type | 18.81 | 1 | <0.001 |
| Energy level | 50.50 | 4 | <0.001 |
| Participant | 41.71 | 13 | <0.001 |
| Image content x Artifact type | 3.00 | 2 | 0.051 |
| Artifact type x Energy level | 0.163 | 4 | 0.957 |
| Image content x Energy level | 0.566 | 8 | 0.806 |

where S is the total number of subjects (after subject exclusion). To make the final scores easier to interpret, the resulting MOSs were linearly remapped to the range of [1, 10].

*B. Results of Experiment 1*

The MOSs and their corresponding error bars are illustrated in Figure 10. Figure 10(a) indicates that the difference in perceived quality between degradations with ghosting and white noise is in general small. Whether at the same energy level either ghosting or white noise mostly affects the overall quality tends to depend on the distortion level and image content. For the source image "liver", the added white noise consistently results in a lower image quality than the added ghosting (see stimuli referred to as 11-15 in Figure 10(a)). A similar consistency, however, is not found for the two brain images, i.e. "brain_1" and "brain_2". For these stimuli the quality of images degraded by ghosting is comparable to the quality of images degraded by the same energy level of white noise.

Figure 10(b) shows that the quality of an MR image degraded by colored noise is consistently scored higher than the quality of the corresponding image degraded by edge ghosting. It implies that the perceived quality is largely reduced when changing the signal distortion from unstructured colored noise to structured edge ghosting, even for the same level of energy in the distortion. In addition, we can observe a trend from the comparison of the four types of artifacts that when either ghosting, white noise or edge ghosting is added to a source image, the perceived image quality monotonously decreases with the energy in the distortion; this, however, is not the case for colored noise, for which the resulting quality may jump up and down as a function of energy level in the distortion.

The observed tendencies are further statistically analyzed with an ANOVA (Analysis of Variance) per graph/part of the experiment separately (using the software package SPSS version 19). In each case, the perceived quality is selected as the dependent variable, the image content, artifact type and energy level as fixed independent variables and the participants as random independent variable. All 2-way interactions of the fixed variables are included in the analysis as well. The results for images degraded with ghosting and white noise are summarized in Table 1, and show that image content, artifact type and energy level have a significant effect on perceived quality. On average, images affected with ghosting

as additional outliers. For the scoring of edge ghosting and colored noise, two subjects (out of 17) were excluded, and again 2 of the remaining scores were rejected as additional outliers. Finally, in Experiment 2, one subject (out of 18) was excluded, but we also had to remove 27 of the remaining scores as additional outliers.

After having applied the outlier removal and subject exclusion procedure, the scores of the remaining subjects were calibrated towards the same mean and standard deviation using z-scores:

$$z_{ij} = \frac{r_{ij} - \mu_i}{\sigma_i} \quad (3)$$

where $r_{ij}$ and $z_{ij}$ indicate the raw score and z-score of the $i$-th subject and $j$-th image, respectively. $\mu_i$ is the mean of the raw scores over all images scored by subject $i$, and $\sigma_i$ is the corresponding standard deviation. These scores were averaged across subjects to yield a mean opinion score (MOS) for the $j$-th image, i.e.

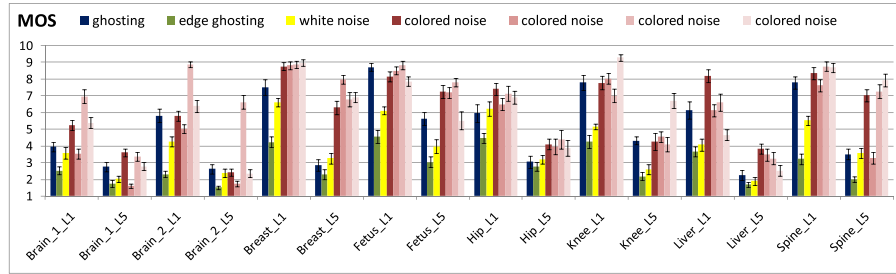$$MOS_j = \frac{1}{S} \sum_{i=1}^{S} z_{ij} \quad (4)$$

Fig. 11.   The MOS resulting from the image quality assessment of Experiment 2. The horizontal axis refers to 16 sets of stimuli, including 8 source images and 2 distortion levels. Each set corresponds to 7 distorted images: one for ghosting, one for edge ghosting, one for white noise, and 4 versions of colored noise. In each case the distortion is applied at the same energy level. The error bars indicate the 95% confidence interval.
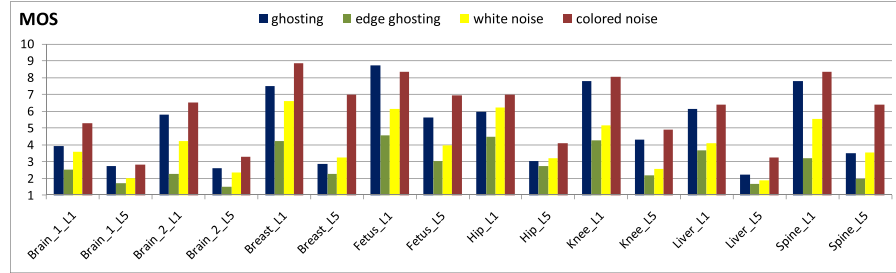


Fig. 12.   The MOS resulting from the image quality assessment of Experiment 2, in which the MOS values of the stimuli degraded with colored noise are averaged for each source image at a given energy level. The horizontal axis refers to 16 sets of stimuli, including 8 source images and 2 distortion levels. Each set corresponds to 4 distorted versions of the image, including degradation with ghosting, edge ghosting, white noise, and colored noise, each at the same energy in the signal distortion.

TABLE II

RESULTS OF THE ANOVA FOR EXPERIMENT 1 TO EVALUATE THE
EFFECT OF EDGE GHOSTING AND COLORED NOISE
ON THE DIFFERENT IMAGES

|  | F-value | df | p |
|---|---|---|---|
| Image content | 20.39 | 2 | <0.001 |
| Artifact type | 386.30 | 1 | <0.001 |
| Energy level | 38.50 | 4 | <0.001 |
| Participant | 38.13 | 14 | <0.001 |
| Image content x Artifact type | 12.02 | 2 | <0.001 |
| Image content x Energy level | 10.89 | 8 | <0.001 |
| Artifact type x Energy level | 10.74 | 4 | <0.001 |

are scored higher in quality than images affected with white noise ($<MOS>$ for ghosting $= 5.05$, $<MOS>$ for white noise $= 4.61$). The post-hoc analysis on image content shows that the viewers score the quality of the image "brain_2" ($<MOS> = 5.31$) on average statistically significantly higher than the quality of the images "brain_1" ($<MOS> = 4.59$) and "liver" ($<MOS> = 4.59$). Also the interaction between image content and artifact type is significant, which implies that the difference in quality between the two types of artifact is not the same for the three images.

The results for images degraded with edge ghosting and colored noise are summarized in Table 2. This table shows that all main effects and interactions are highly statistically significant. Overall images degraded with colored noise ($<MOS> = 5.37$) are scored higher in quality than images degraded with edge ghosting ($<MOS> = 2.59$). The post-hoc analysis on the image content indicates that the image "brain_1" ($<MOS> = 3.04$) received statistically significantly lower quality scores than the other two images ($<MOS> = 4.26$ for

"brain_2" and $<MOS> = 4.37$ for "liver"). The interaction between image content and artifact type is caused by the fact that the quality difference between images is much larger for the colored noise artifact than for the edge ghosting artifact. The interaction between artifact type and energy level is significant since the quality monotonically decreases with increasing energy level for the edge ghosting artifact, but not for the colored noise. In the latter case, the perceived quality fluctuates with increasing energy level. This phenomenon also explains the significant interaction between image content and energy level.

### C. Results of Experiment 2

The MOSs and their corresponding error bars are illustrated in Figure 11. Comparing only the four bars for colored noise in this figure confirms that the randomization procedure in the generation of colored noise indeed has a large impact on the perceived quality of MR images. The four stimuli degraded with colored noise vary in their perceived quality, even when applied on the same source image at the same energy level: in some cases (e.g. for stimulus "brain1_L1") the impact is large, while in other cases (e.g. for stimulus "breast_L1") the impact is less obvious. Therefore, in order to compare the effect of four specific types of artifacts on the perceived quality, the MOS values of the stimuli degraded with colored noise are averaged for each source image at a given energy level, the result of which is shown in Figure 12.

Figure 12 shows some general trends: (1) at the same energy level edge ghosting consistently results in the lowest image quality, independent of the image content, (2) in most cases, ghosting yields a higher image quality than white noise,

TABLE III

RESULTS OF THE ANOVA WITH ARTIFACT TYPE AS THE
INDEPENDENT VARIABLE FOR EXPERIMENT 2

| | F-value | df | p |
|---|---|---|---|
| Image content | 86.05 | 7 | <0.001 |
| Artifact type | 351.09 | 3 | <0.001 |
| Energy level | 1164.38 | 2 | <0.001 |
| Participant | 51.63 | 16 | <0.001 |
| Image content x Artifact type | 8.97 | 21 | <0.001 |
| Image content x Energy level | 4.71 | 7 | <0.001 |
| Artifact type x Energy level | 23.87 | 3 | <0.001 |

TABLE IV

MEAN AND STANDARD DEVIATION OF THE AVERAGED QUALITY
SCORES PER ARTIFACT TYPE FOR EXPERIMENT 2

| Artifact | mean | st. dev. |
|---|---|---|
| edge ghosting | **2.48** | 1.58 |
| white noise | **3.44** | 1.85 |
| ghosting | **4.31** | 2.35 |
| colored noise | **5.24** | 2.03 |

except for "breast_L5", "hip_L1", "hip_L5", and "spine_L5", in which the annoyance of ghosting is comparable to that of white noise, (3) for ghosting, edge ghosting, white noise and colored noise, the perceived quality at the low energy level is higher than at high energy level, independent of the image content. The latter means that the energy of the distortion seems to be a good metric to predict the perceived image quality induced by ghosting, edge ghosting, white noise and colored noise (when eliminating the effect of the randomness in the noise creation procedure by averaging).

To confirm these trends with a statistical analysis, an ANOVA is performed with the quality score as the dependent variable, the image content, artifact type and energy level as fixed independent variables, the participants as random independent variable, and all 2-way interactions of the fixed independent variables. The results, summarized in Table 3, show that all main effects and interactions are statistically significant. Clearly, the energy level has a statistically significant effect on the perceived quality, with a higher score for the lower energy level. Not all images have the same overall quality. The post-hoc test reveals the following order in quality (note that commonly underlined images are not significantly different from each other):

brain1 < liver < brain2 < hip < spine < knee < breast < fetus

There is also a significant difference in quality between the four types of artifacts, despite the fact that they are applied at exactly the same energy level. The post-hoc analysis reveals the following order in quality:

edge ghosting < white noise < ghosting < colored noise

In other words, edge ghosting deteriorates the image quality most, followed by white noise, standard ghosting and finally colored noise. Table 4 summarizes the impact of the four different artifacts on image quality, averaged over all source images. Assuming that quality is linearly related to the applied energy level (which is more or less the case for most of the artifacts as proven in Experiment 1), one can calculate to what

extent the energy level of the least annoying artifact may be increased in order to become as annoying as any of the other artifacts. In other words, starting from the mean quality level for ghosting, white noise and edge ghosting, one can calculate to what ratio the energy level of the colored noise artifact may be increased to obtain the same quality level. Doing so, one can conclude that the energy level of colored noise may be **1.6** times as high as the energy level of ghosting to be perceived as equally annoying. Similarly, the quality level of colored noise may be twice the energy level of white noise, and **2.7** times the energy level of edge ghosting.

Apart from the main effects, also the interactions are statistically significant. The impact the different artifacts have on image quality depends on the image content. The way the image quality changes with the energy level also depends on the image content and on the artifact type.

## V. COMPARING SPECTRAL COLORING AND STRUCTUREDNESS IN TERMS OF IMAGE QUALITY

The evaluation so far focused on the comparison of four individual types of artifacts, each of which statistically significantly impacted the perceived quality of MR images. As already illustrated in Figure 1, these artifacts can also be characterized at a different aggregation level, using "spectral coloring" or "structuredness" as the classification variable. The former variable refers to whether the artifacts have a spectral power density proportional to the spectral power density of the original image (i.e. colored noise and ghosting) or have a (more) flat power density (i.e. white noise and edge ghosting). The latter variable makes a distinction between artifacts that copy the structure of the original image (i.e. ghosting and edge ghosting) and artifacts that are spread over the whole image area (i.e. white noise and colored noise). Therefore, the question arises whether "spectral coloring" or "structuredness" has an impact on perceived quality, and if so, to what extent. To check such effect with a statistical analysis, an ANOVA was performed again on the results of Experiment 2 in a similar way as described in Section IV.C, but with two new independent variables to substitute the variable artifact type: i.e. spectral coloring and structuredness. In this case, the quality score is chosen as the dependent variable, the image content, spectral coloring (i.e. colored noise and ghosting are given a value of 1, and white noise and edge ghosting are given a value of 0), structuredness (i.e. ghosting and edge ghosting are given a value of 1, and white noise and colored noise are given a value of 0) and energy level as fixed independent variables, the participants as random independent variable, and all 2-way interactions of the fixed independent variables. The results are summarized in Table 5, and show that all main effects and most interactions are statistically significant. Since most of the conclusions we drew based on Table 3 are also valid here, we do not repeat them. Instead we mainly focus on the effect of the two new variables on image quality. As shown in Table 5, each of these variables has a statistically significant effect on the perceived quality, while their interaction is not significant. The corresponding mean values and standard deviations are given in Table 6, while the main effects and the

### TABLE V
RESULTS OF THE ANOVA WITH STRUCTUREDNESS AND SPECTRAL COLORING AS THE INDEPENDENT VARIABLES FOR EXPERIMENT 2

| | F-value | df | p |
|---|---|---|---|
| Image content | 82.294 | 7 | <0.001 |
| Spectral coloring | 791.104 | 1 | <0.001 |
| Structuredness | 216.175 | 1 | <0.001 |
| Energy level | 1113.594 | 1 | <0.001 |
| Participant | 49.373 | 16 | <0.001 |
| Image content x Energy level | 4.5 | 7 | <0.001 |
| Image content x Spectral coloring | 13.417 | 7 | <0.001 |
| Image content x Structuredness | 9.325 | 7 | <0.001 |
| Spectral coloring x Energy level | 36.517 | 1 | <0.001 |
| Structuredness x Energy level | 0.130 | 1 | 0.719 |
| Spectral coloring x Structuredness | 0.043 | 1 | 0.836 |

### TABLE VI
MEAN AND STANDARD DEVIATION OF THE AVERAGED QUALITY SCORES FOR STRUCTURED VS. UNSTRUCTURED ARTIFACTS AND FOR COLORED VS. WHITE ARTIFACTS

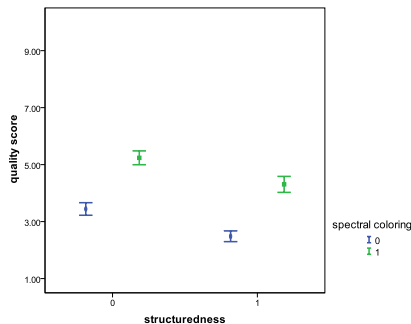| | mean | st. dev. |
|---|---|---|
| structured artifacts | **3.40** | 2.20 |
| unstructured artifacts | **4.34** | 2.14 |
| colored artifacts | **4.77** | 2.24 |
| white artifacts | **2.96** | 1.78 |



Fig. 13. Scatter plot of perceived quality for structuredness and spectral coloring of artifacts (averaged over all image content and the two energy levels). The "spectral coloring" refers to whether the artifacts were "colored" (i.e. colored noise and ghosting were given a value of 1) or "white" (i.e. white noise and edge ghosting were given a value of 0). The "structuredness" makes a distinction between structured artifacts (i.e. ghosting and edge ghosting were given a value of 1) and unstructured artifacts (i.e. white noise and colored noise were given a value of 0).

interaction are illustrated in Figure 13. Both the figure and table demonstrate that "colored" artifacts deteriorate the perceived quality less than "white" artifacts, and "unstructured" artifacts deteriorate the quality less than "structured" artifacts. The difference in quality between "colored" and "white" artifacts is bigger than the difference in quality between "structured" and "unstructured" artifacts. One can increase the energy of "colored" artifacts with a factor of about **2** to be perceived as annoying as "uncolored" artifacts, and similarly, one can increase the energy of "unstructured" artifacts with about **1.6** to be perceived as annoying as "structured" artifacts.

## VI. DISCUSSION

Our experimental results provide the general insight that at equal energy artifacts with a flat spectral power density in MR images are roughly twice as annoying in terms of image quality as artifacts with a spectral power density equal to that of the original image. Equivalently, artifacts that replicate the structure of the original content are 1.6 times as annoying as artifacts that are spread over the whole image area. One should realize though that these conclusions hold for the impact of the artifact on image quality averaged over all MR images and energy levels used in our experiments. The statistical analyses show significant interactions with content and energy level, implying that the conclusions, and especially the ratios in annoyance, are not necessarily true for all MR images and all energy levels.

The impact of image content on perceived image quality may have two possible causes. First, in our study we determined the BEL per original image, and as a consequence, its absolute value is different for different image content. Second, the image content itself may induce an intrinsic difference in sensitivity to artifacts, and thus, in the visibility of the artifacts. The first hypothesis can be tested by calculating the correlation between the BEL and the averaged MOS per original image. This value is very low ($\rho = 0.2$), and thus, differences in BEL are not the dominant factor explaining the image content dependency. More probably, differences in perceived quality per image content are caused by differences in sensitivity to artifact visibility between the different contents.

One should also realize that the artifacts used in this study were not actual artifacts captured during acquisition of the images, instead they were simulated on high-quality captured images as realistically as possible in order to have better control on the intensity level of the artifact. As such the annoyance of actual artifacts created during MR image acquisition may be slightly different. The latter may, for example, be the case for MR images degraded with ghosting, since we used in our simulations asymmetric ghosting, while MR acquisition usually result in more symmetric ghosting.

Also, it should be noted that the subjects involved in the study are mainly clinical medical physicists, who play a principal role in the review of diagnostic image quality, the development of systems and policies and the deployment of new imaging procedures. The difference in image quality perception between physicists and radiologists are, so far, not fully understood, but worth a further investigation.

It is also good to realise that our evaluation with a subjective experiment is intrinsically time-consuming, and therefore, limited with respect to the amount of test stimuli and the number of human subjects. Adding more experimental data to the evaluation would be highly beneficial, especially in terms of adding confidence to the generalizability of the conclusions.

Last, but not least, being able to measure the perceived image quality allows follow up research, where clinical tasks (e.g., lesion detection or characterization) could be included to study whether and to what extent the measured differences in image quality can impact the diagnostic performance (e.g., diagnostic accuracy or visibility of lesions). However, MRI is used for a wide variety of clinical tasks and could potentially display a very large gamut of pathology appearances. So, on one hand, inclusion of images with a specific pathology would make the study closer to reality, but on the

other hand, it would focus the viewer's criteria to the ability to read that very specific lesion. This would be no problem if we could present many different pathologies, but, as already discussed, the risk of viewer's fatigue limits the study to just a few cases. In that sense, the *absence* of a lesion forces the viewer to consider the readability of all lesions that *might* be present in that type of image.

## VII. CONCLUSIONS

In this study, we investigated the relative impact on perceived image quality of four distortion types (i.e. ghosting, edge ghosting, white noise and colored noise), which often occur in current MR imaging applications. We performed a series of image quality scoring experiments, in which MR images of different content affected with these types of artifacts at different levels of energy were assessed by clinical application specialists.

The results of these experiments showed, that in general, the energy in the frequency spectrum is a good measure to predict the perceived quality related to a given artifact. Gradually increasing the energy in the artifact decreases the perceived quality of the image. The exception to this observation is colored noise, where the randomization in the simulation of the artifact may have a bigger impact on quality than the energy.

The impact of the artifacts on image quality strongly depends on the specific content of the MR image. However, when neglecting this dependency (i.e. interactions with energy level and image content), we may conclude that in general "unstructured" artifacts deteriorate quality less than "structured" artifacts and "colored" artifacts deteriorate quality less than "white" artifacts. More specifically, we found that the energy of "unstructured" artifacts may be increased with a factor of 1.6 to become as annoying in perceived quality as "structured" artifacts. Similarly, the energy of "colored" artifacts may be doubled to become as annoying as "white" artifacts.

Observers consistently score images degraded by edge ghosting as having the lowest quality, independent of the energy level of the distortion, and of the image content. Assuming a linear relation of quality with energy, and taking the least annoying artifact (i.e., colored noise) as a reference, we may increase its energy level with a factor of 1.6, 2 and 2.7 to make its quality comparable to that of ghosting, white noise and edge ghosting, respectively.

This study provides new insights in the perception of image quality, and the findings are ready to be embedded in a real-world MR imaging system to adapt its parameters, and so to optimize the image rendering to the perception of users.

## REFERENCES

[1] S. C. Bushong, *Magnetic Resonance Imaging: Physical and Biological Principles*. St. Louis, MO, USA: Mosby, 1988.

[2] R. Freeman, *A Handbook of Nuclear Magnetic Resonance*. Essex, U.K.: Longman, 1997.

[3] E. Pusey *et al.*, "Magnetic resonance imaging artifacts: Mechanism and clinical significance," *Radiographics*, vol. 6, no. 5, pp. 891–911, 1986.

[4] J. A. Clark and W. M. Kelly, "Common artifacts encountered in magnetic resonance imaging," *Radiol. Clin. North Amer.*, vol. 26, no. 5, pp. 893–920, 1988.

[5] C. E. Willis, S. K. Thompson, and S. J. Shepard, "Artifacts and misadventures in digital radiography," *Appl. Radiol.*, vol. 33, no. 1, pp. 11–20, 2004.

[6] M. Hedley and H. Yan, "Motion artifact suppression: A review of post-processing techniques," *Magn. Reson. Imag.*, vol. 10, no. 4, pp. 627–635, 1992.

[7] S. B. Reeder, E. Atalar, B. D. Bolster, and E. R. McVeigh, "Quantification and reduction of ghosting artifacts in interleaved echo-planar imaging," *Magn. Reson. Med.*, vol. 38, no. 3, pp. 429–439, 1997.

[8] Y. C. Kim, J.-F. Nielsen, and K. S. Nayak, "Automatic correction of echo-planar imaging (EPI) ghosting artifacts in real-time interactive cardiac MRI using sensitivity encoding," *J. Magn. Reson. Imag.*, vol. 27, no. 1, pp. 239–245, 2008.

[9] M. A. Barish and H. Jara, "Motion artifact control in body MR imaging," *Magn. Reson. Imag. Clin. North Amer.*, vol. 7, no. 2, pp. 289–301, 1999.

[10] R. Archibald and A. Gelb, "A method to reduce the Gibbs ringing artifact in MRI scans while keeping tissue boundary integrity," *IEEE Trans. Med. Imag.*, vol. 21, no. 4, pp. 305–319, Apr. 2002.

[11] S. A. Mirowitz, "MR imaging artifacts. Challenges and solutions," *Magn. Reson. Imag. Clin. North Amer.*, vol. 7, no. 4, pp. 717–732, 1999.

[12] T. B. Smith and K. S. Nayak, "MRI artifacts and correction strategies," *Imag. Med.*, vol. 2, no. 4, pp. 445–457, Aug. 2010.

[13] I. M. C. J. van Overveld, "Contrast, noise, and blur affect performance and appreciation of digital radiographs," *J. Digit. Imag.*, vol. 8, no. 4, pp. 168–179, 1995.

[14] E. A. Krupinski, "The importance of perception research in medical imaging," *Radiat. Med.*, vol. 18, no. 6, pp. 329–334, 2000.

[15] E. A. Krupinski and H. L. Kundel, "Update on long-term goals for medical image perception research," *Acad. Radiol.*, vol. 5, no. 9, pp. 629–633, 1998.

[16] E. A. Krupinski, H. L. Kundel, P. F. Judy, and C. F. Nodine, "The medical image perception society. Key issues for image perception research," *Radiology*, vol. 209, no. 3, pp. 611–612, 1998.

[17] E. A. Krupinski, "The future of image perception in radiology: Synergy between humans and computers," *Acad. Radiol.*, vol. 10, no. 1, pp. 1–3, 2003.

[18] D. J. Manning, A. Gale, and E. A. Krupinski, "Perception research in medical imaging," *British J. Radiol.*, vol. 78, pp. 683–685, 2005.

[19] H. L. Kundel, "Images, image quality and observer performance," *Radiology*, vol. 132, no. 2, pp. 265–271, 1979.

[20] H. L. Kundel, "Visual perception and image display terminals," *Radiol. Clin. North Amer.*, vol. 24, no. 1, pp. 69–78, 1986.

[21] M. S. Chesters, "Human visual perception and ROC methodology in medical imaging," *Phys. Med. Biol.*, vol. 37, no. 7, pp. 1433–1476, 1992.

[22] A. Burgess, "Image quality, the ideal observer, and human performance of radiologic decision tasks," *Acad. Radiol.*, vol. 2, no. 6, pp. 522–526, 1995.

[23] M. Tapiovaara, "Image quality measurements in radiology," *Radiat. Protection Dosimetry*, vol. 117, nos. 1–3, pp. 116–119, 2005.

[24] C. E. Metz, "ROC analysis in medical imaging: A tutorial review of the literature," *Radiol. Phys. Technol.*, vol. 1, no. 1, pp. 2–12, 2008.

[25] E. A. Krupinski, H. Roehrig, and T. Furukawa, "Influence of film and monitor display luminance on observer performance and visual search," *Acad. Radiol.*, vol. 6, no. 7, pp. 411–418, 1999.

[26] M. Fuderer, "The information content of MR images," *IEEE Trans. Med. Imag.*, vol. 7, no. 4, pp. 368–380, Dec. 1988.

[27] H. Kanazawa, "Method and apparatus for magnetic resonance imaging," European Patent EP 0 640 843, May 23, 2001. [Online]. Available: https://www.google.com/patents/EP0640843B1?cl=en

[28] K. P. N. Forbes, J. G. Pipe, C. R. Bird, and J. E. Heiserman, "PROPELLER MRI: Clinical testing of a novel technique for quantification and compensation of head motion," *J. Magn. Reson. Imag.*, vol. 14, no. 3, pp. 215–222, 2001.

[29] J. Milles, M. J. Versluis, A. G. Webb, and J. H. C. Reiber, "Quantitative evaluation of Compressed Sensing in MRI: Application to 7T time-of-flight angiography," in *Proc. IEEE ITAB*, Nov. 2010, pp. 1–4.

[30] J. M. Peeters and M. Fuderer, "Minimal artifact factor SENSE," in *Proc. 20th Annu. Meeting ISMRM*, 2012, p. 3353.

[31] *Methodology for the Subjective Assessment of the Quality of Television Pictures*, document ITU-R Rec. BT.500-11, International Telecommunication Union, Geneva, Switzerland, 2002.

[32] H. Liu, J. Koonen, M. Fuderer, and I. Heynderickx, "Studying the relative impact of ghosting and noise on the perceived quality of MR images," *Proc. SPIE*, vol. 8318, p. 83181K, Feb. 2012.

[33] Z. Wang and A. C. Bovik, *Modern Image Quality Assessment: Synthesis Lectures on Image, Video, and Multimedia Processing*. San Rafael, CA, USA: Morgan & Claypool, 2006.

[34] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440–3451, Nov. 2006.

[35] H. Tao, N. Klomp, and I. Heynderickx, "A no-reference metric for perceived ringing artifacts in images," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 4, pp. 529–539, Apr. 2010.

[36] K. J. Robson and C. J. Kotre, "Pilot study into optimisation of viewing conditions for electronically displayed images," *Radiat. Protection Dosimetry*, vol. 117, nos. 1–3, pp. 298–303, 2005.

[37] E. Samei *et al.*, "Assessment of display performance for medical imaging systems," in *Report of the American Association of Physicists in Medicine (AAPM) Task Group 18*. Madison, WI, USA: Medical Physics Publishing, 2005.

[38] B. M. Hemminger, R. E. Johnston, J. P. Rolland, and K. E. Muller, "Introduction to perceptual linearization of video display systems for medical image presentation," *J. Digit. Imag.*, vol. 8, no. 1, pp. 21–34, 1995.

[39] National Electrical Manufacturers Association (NEMA), Rosslyn, VA, USA. (2006). *Digital Imaging and Communications in Medicine (DICOM) Part 14: Greyscale Standard Display Function*, accessed on Oct. 2013. [Online]. Available: http://medical.nema.org/dicom/2004/04_14PU.PDF

[40] A. E. Kumcu *et al.*, "Visual quality assessment of H.264/AVC compressed laparoscopic video," *Proc. SPIE*, vol. 9037, pp. 90370A-1–90370A-12, Mar. 2014.

[41] L. Platisa, "Image quality assessment: Utility, beauty, appearance," Ph.D. dissertation, Dept. Telecommun. Inf. Process., Ghent Univ., Ghent, Belgium, 2014.

[42] M. Razaak, M. G. Martini, and K. Savino, "A study on quality assessment for medical ultrasound video compressed via HEVC," *IEEE J. Biomed. Health Inform.*, vol. 18, no. 5, pp. 1552–1559, Sep. 2014.

[43] H. Liu, J. Koonen, M. Fuderer and, I. Heynderickx. (2016). *Subjective Data: The Relative Impact of Ghosting and Noise on the Perceived Quality of MR Images*. [Online]. Available: https://users.cs.cf.ac.uk/

**Jos Koonen** received the master's degree in physics from the RWTH-Aachen, in 1998. He joined Philips Healthcare in 2001, where he was involved in the development of MRI devices, in particular on system calibrations, and MR image quality.

**Miha Fuderer** received the M.Sc. degree in applied physics, Brussels, 1982. Since 1985, He has been with Philips Medical Systems on the development of MRI devices, most notably on reconstruction algorithms, and on metrics and specification of MR image quality.

**Hantao Liu** received the M.Sc. degree from the University of Edinburgh, Edinburgh, U.K., in 2005, and the Ph.D. degree from the Delft University of Technology, Delft, The Netherlands, in 2011. He is currently an Assistant Professor with the School of Computer Science and Informatics, Cardiff University, Cardiff, U.K. His research interests include visual media quality assessment, visual attention modeling and applications, visual scene understanding, and medical image perception. He is currently serving for the IEEE MMTC, as the Chair of the Interest Group on Quality of Experience for Multimedia Communications, and he is an Associate Editor of the IEEE TRANSACTIONS ON HUMAN-MACHINE SYSTEMS.

**Ingrid Heynderickx** received the Ph.D. degree in physics with the University of Antwerp. She joined the Philips Research in 1987. From 1999 to 2013, she was the Head of the research activities on visual perception of display and lighting systems with Philips Research. In 2005, she was appointed as a Guest Research Professor with the Southeast University of Nanjing, and a Part-Time Full Professor with the Delft University of Technology. Since 2013, she has been a Full Professor with the Eindhoven University of Technology, where she became the Dean of the Department of Industrial Engineering and Innovation Sciences in 2015.