

Studying the relative impact of ghosting and noise on the perceived quality of MR images

Hantao Liu^{*a}, Jos Koonen^b, Miha Fuderer^b and Ingrid Heynderickx^{a,c}

^aDepartment of Mediamatics, Delft University of Technology, Delft, The Netherlands

^bPhilips Healthcare, Best, The Netherlands

^cGroup Visual Experiences, Philips Research Laboratories, Eindhoven, The Netherlands

ABSTRACT

In current magnetic resonance (MR) imaging systems, design choices are confronted with a trade-off between structured (i.e. artifacts) and unstructured noise. The impact of both types of noise on perceived image quality, however, is so far unknown, while this knowledge would be highly beneficial for further improvement of MR imaging systems. In this paper, we investigate how ghosting artifacts (i.e. structured noise) and random noise, applied at the same energy level in the distortion, affect the perceived quality of MR images. To this end, a perception experiment is conducted with human observers rating the quality of a set of images, distorted with various levels of ghosting and noise. To also understand the influence of professional expertise on the image quality assessment task, two groups of observers with different levels of medical imaging experience participated in the experiment: one group contained fifteen clinical scientists or application specialists, and the other group contained eighteen naïve observers. Experimental results indicate that experts and naïve observers differently assess the quality of MR images degraded with ghosting/noise. Naïve observers consistently rate images degraded with ghosting higher than images degraded with noise, independent of the energy level of the distortion, and of the image content. For experts, the relative impact of ghosting and noise on perceived quality tends to depend on the energy level of the distortion and on the image content, but overall the energy of the distortion is a promising metric to predict perceived image quality.

Keywords: MRI, perceived image quality, ghosting, noise, human visual system

1. PURPOSE

In current magnetic resonance (MR) imaging systems, a variety of artifacts and noise, affecting the perceived quality of MR images is generated. Ghosting, which is a cross-talk type of artifact that generates a lower-intensity double image, spatially shifted with respect to the original content, is just one example of such an artifact. In an MR imaging system, there are situations where structured noise (e.g. ghosting) on one hand can be traded off with unstructured noise on the other hand. For example, one of the processing steps to avoid ghosting in the system is to redistribute its energy so that it appears as random noise spread over the entire image. Whether this way of processing changes the perceived quality of MR images, and if so, to what extent is so far unknown. Therefore, it is of fundamental importance to understand the relative impact of structured versus unstructured noise on the perceived quality of MR images, in order to improve MR imaging systems. In addition, it is worthwhile to investigate the influence of professional expertise on image quality perception. The impact of professionalism on *diagnostic* quality, of course, is known to be huge ([1]-[3]), but in case the impact of professionalism on *perceived image* quality is limited, larger-size studies could be undertaken without being too much of a burden for the limited number of MR experts.

2. METHOD

2.1 Stimuli

We selected two images of a brain and one image of a liver as the three original MR images to be used in this study. They are shown in Figure 1. Each source image was intentionally distorted with ghosting at five different energy levels. Subsequently, for each distorted image the added energy of ghosting was redistributed as random noise on top of the original image. Figure 2 illustrates the simulation process to add ghosting and noise to the original images. To simulate the ghosting, two new images were first generated: (1) a lower-intensity (LI) (20% of the maximal intensity in our

simulation) version of the original image, and (2) a mask image representing the area of the clinical object. The LI image was spatially shifted to the left (with negative intensity values to simulate a negative intensity ghost) and to the right (with positive intensity values to simulate a positive intensity ghost), with respect to the original content (i.e. the distance is 1/3 of the image width in our simulation). This operation resulted in a new LI image in which the clinical object was doubled. Combining this new LI image with the mask image generated a ghosting image. Adding this ghosting image to the original image yielded the test stimuli distorted with ghosting, as shown in Figure 3(b). To simulate the noise, first an image of white noise was generated, and then this image was combined with the mask image to yield a noise image. The intensity of the noise image was scaled such that the resulting total energy of noise was equal to the energy calculated for the ghosting image. The resulting stimulus distorted with noise was generated by adding the noise image to the original image, as shown in Figure 3(c). The added energy of ghosting and noise was then downscaled with factors of 4/5, 3/5, 2/5, 1/5, resulting in four new energy levels of ghosting and noise. By doing so, each original image was distorted with 5 levels of simulated ghosting and 5 levels of simulated noise. Hence, the test database of this study existed of 30 stimuli (i.e. 3 originals \times 5 energy levels \times 2 types of artifacts) in total.

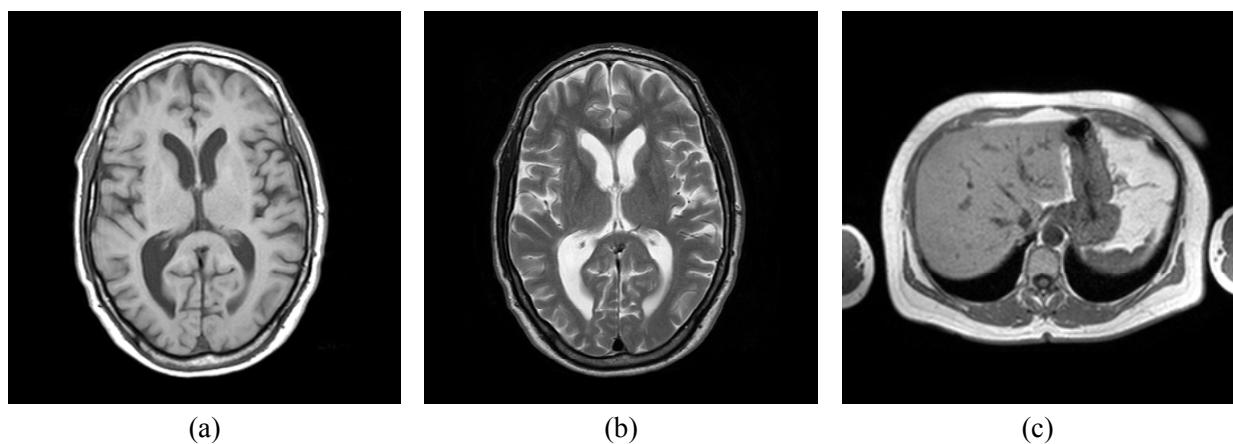


Figure 1. Source images, which are referred to as: (a) “brain_1”, (b) “brain_2”, and (c) “liver”.

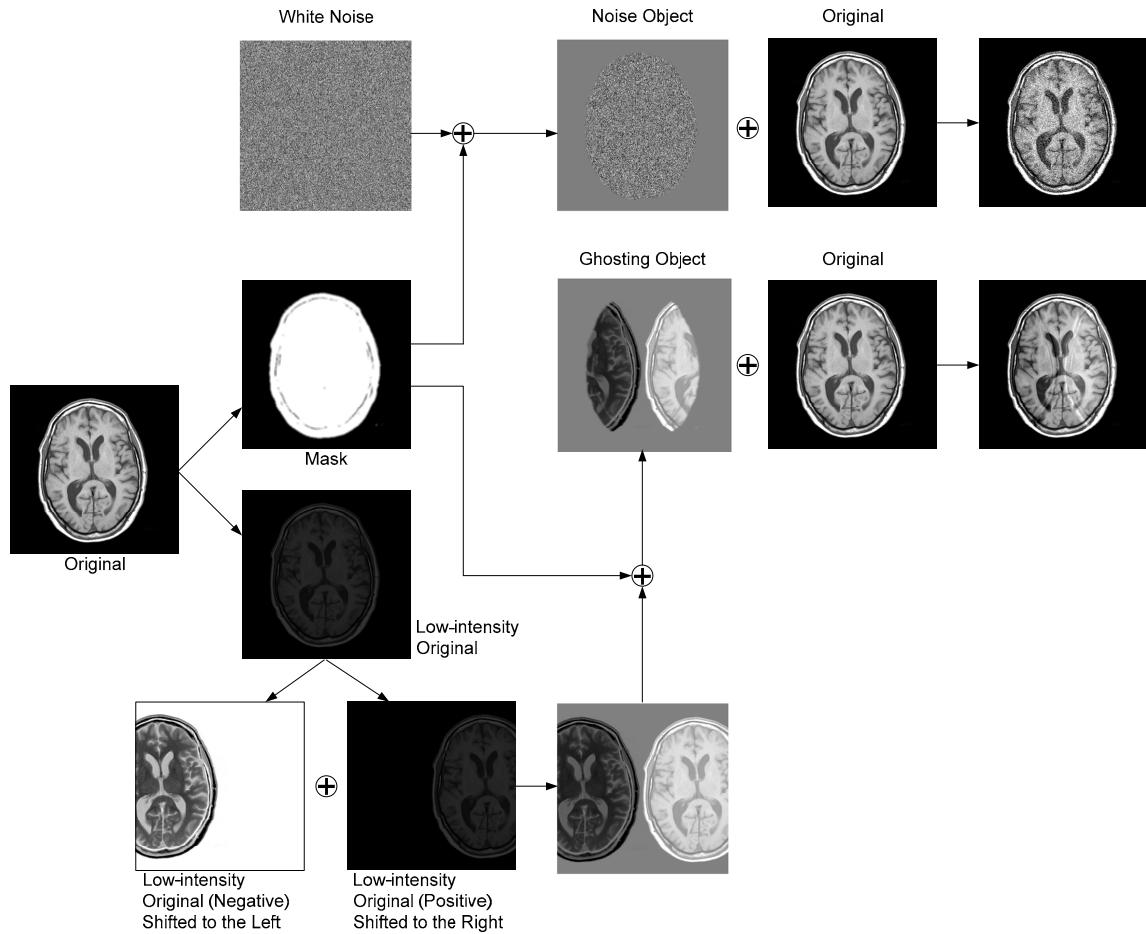


Figure 2. Illustration of the stimulation of ghosting and noise.

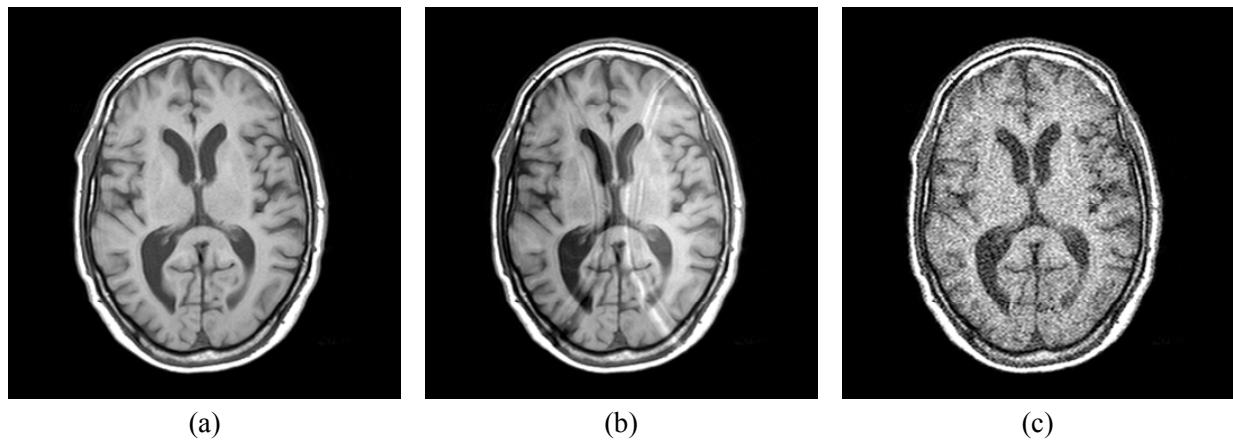


Figure 3. Examples of test stimuli: (a) source image of “brain_1”, (b) image (a) with ghosting, and (c) image (a) with noise. The energy of the added noise in Figure 3(c) is the same as the energy of added ghosting of Figure 3(b).

2.2 Experimental protocol

A simultaneous-double-stimulus (SDS) method [4], which means that subjects had to score the quality for each stimulus in the presence of a reference, was used in our experiment. The rating interface is illustrated in Figure 4; the two stimuli, i.e. the reference 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 as illustrated in Figure 4. 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 the 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 ghosting and noise 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 random order to each subject in a separate session.

The study was performed with two groups of participants, each having a different level of professional expertise on MR images. The 18 participants of the first group (i.e. G1) were recruited from the students or staff at the Delft University of Technology, representing naive subjects with no expertise on medical imaging at all. The 15 participants of the second group (i.e. G2) were clinical scientist or application specialists from Philips Healthcare in Best. The experiment was conducted with the subjects of G1 in an experimental lab at Delft University of Technology, Delft, The Netherlands, and with the subjects of G2 in an experimental lab at Philips Healthcare, Best, The Netherlands. At both locations, however, the same calibrated monitor and PC, and similar viewing conditions were used.

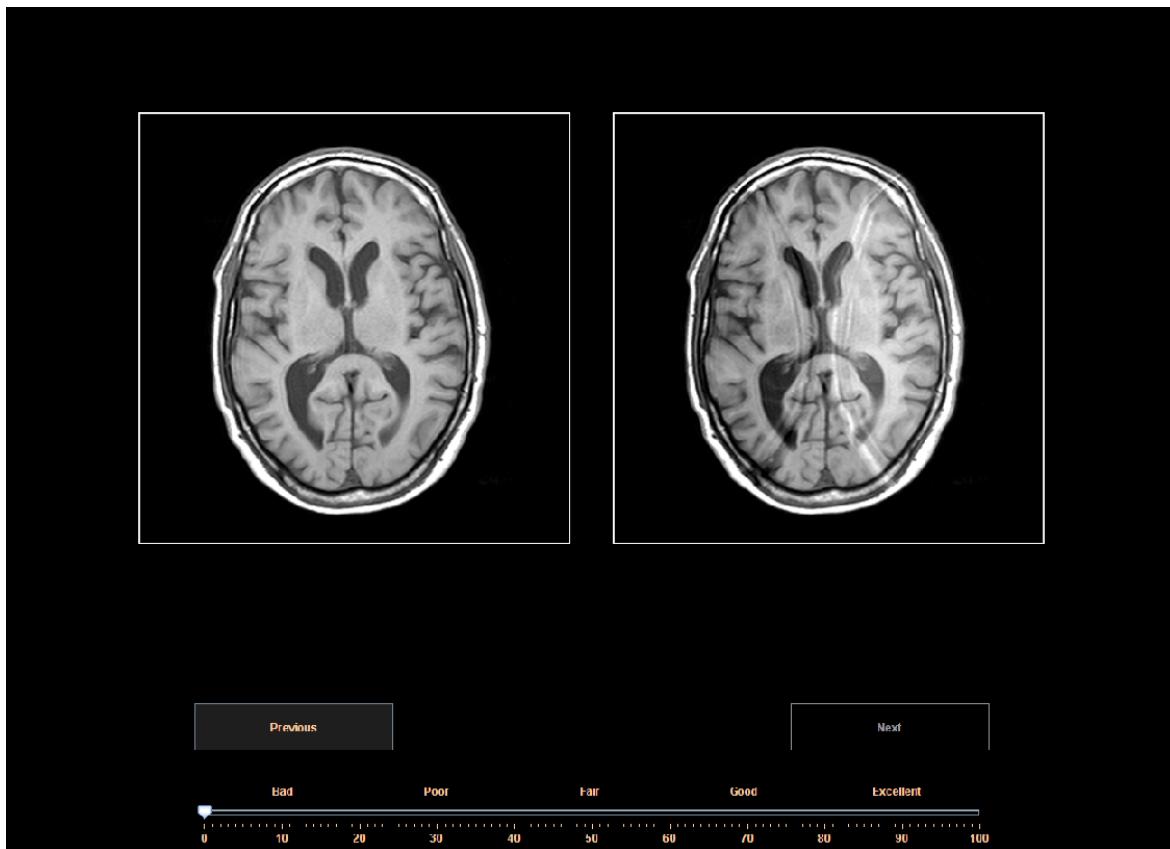


Figure 4. An illustration of the interface used during the experiment, including two stimuli, i.e. the reference at the left-hand side and the test stimulus at the right hand side, and the quality scale.

3. RESULTS

3.1 Processing of the raw data

First, a simple outlier detection and subject rejection procedure was applied to the raw scores. An individual score for an image was considered to be an outlier if it lay outside an interval of two standard deviations around the mean score for that image. All scores of a subject were rejected if more than six (20 percent) of his/her scores were outliers. Overall, two (out of 18) subjects in G1 and one (out of 15) subject in G2 were rejected, and only 2 of the remaining scores (both in G1 and G2) were rejected as additional outliers.

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

$$z_{ij} = \frac{r_{ij} - \mu_i}{\sigma_i}$$

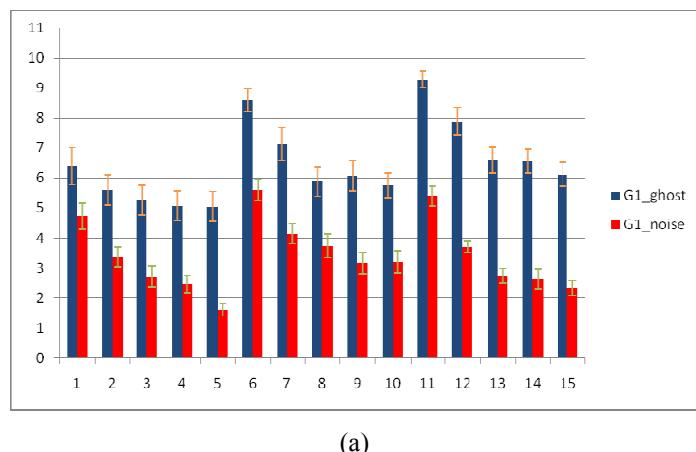
where r_{ij} and z_{ij} indicate the raw score and z -score for the i -th subject and j -th image, respectively. μ_i is the mean of the raw scores over all images scored by subject i , and σ_i is the corresponding standard deviation. The resulting z -scores were normalized to a scoring range of [1, 10], and then averaged across subjects to yield a mean opinion score (MOS) for the j -th image, i.e.

$$MOS_j = \frac{1}{S} \sum_{i=1}^S z_{ij}$$

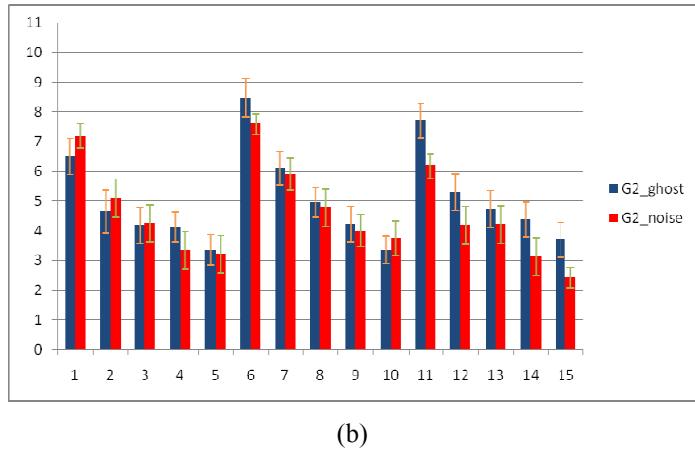
where S is the total number of subjects (after subject rejection).

3.2 Results

The MOSs and their corresponding error bars are illustrated in Figure 5. Figure 5(a) shows the results of the naïve observers (i.e. G1). The quality of an MR image degraded by ghosting is consistently scored higher than the quality of the corresponding image degraded by noise. It indicates that for the naïve observers, the perceived quality is largely reduced when changing the signal distortion from structured to unstructured noise, even for the same level of energy in the distortion. Figure 5(b) shows the results of the clinical scientists or application specialists (i.e. G2). It indicates that the difference between ghosting and noise in affecting the perceived quality is smaller than for the naïve viewers. In addition, whether at the same energy level either ghosting or random noise mostly affects the overall quality tends to depend on the distortion level and image content. For the source image “liver”, the added random noise consistently results in a lower image quality than the added ghosting (see stimuli referred to as 11-15 in Figure 5). A similar consistency, however, is not found for the two brain images, i.e. “brain_1” and “brain2”. For these stimuli the quality of images degraded by ghosting is comparable to the quality of images degraded by the same energy level of random noise. To really study the content dependency, a larger diversity of image content is needed, and this will be part of our future research.



(a)



(b)

Figure 5. The MOS resulting from the image quality assessment: (a) the experiment conducted with naïve observers (i.e. G1), and (b) the experiment conducted with clinical scientists or application specialists (i.e. G2). 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 and one for noise, with each the same energy in the signal distortion. The error bars indicate the 95% confidence interval.

4. CONCLUSIONS

In this paper, we investigate the relative impact of ghosting and noise on the perceived quality of MR images through an image quality scoring experiment. The experiment is performed with two groups of human observers, each having a different level of expertise in medical imaging; one group with naïve observers, and one group with clinical scientists or application specialists. The results indicate that naïve observers assess the quality of the images differently from experts. Naïve observers consistently score images degraded by ghosting higher than the corresponding images degraded by random noise. For experts, the relative impact of ghosting and noise on perceived quality tends to depend on the distortion level and the image content. Our plans for future research include an extension of the set of stimuli with a larger diversity in image content, and consequently a better understanding of the viewing behavior of experts on medical image quality assessment.

REFERENCES

- [1] Krupinski, EA., “Influence of experience on scanning strategies in mammography,” SPIE Medical Imaging Proceedings, 2712:95-10, (1996).
- [2] Krupinski, EA., Weinstein, RS., Rozek, LS., “Experience-related differences in diagnosis from medical images displayed on monitors,” Telemedicine Journal, 2:101-108, (1996).
- [3] Nodine, CF., Kundel, HL., Lauver, SC., Toto, LC., “The nature of expertise in searching mammograms for breast lesions,” SPIE Medical Imaging Proceedings, 2712:89-94, (1996).
- [4] ITU-R Recommendation BT.500-10, [Methodology for the subjective assessment of the quality of television pictures], ITU: Geneva (2000).
- [5] Wang, Z. and Bovik, A. C., [Modern Image Quality Assessment], Synthesis Lectures on Image, Video & Multimedia Processing, Morgan & Claypool Publishers (2006).
- [6] Liu, H., “Modeling Perceived Quality for Imaging Applications,” PhD Thesis, Delft University of Technology, The Netherlands (2011).
- [7] Zhang, L., Cavaro-Menard, C., Le Callet, P., Cooper, L.H.K., Hunault, G., Tanguy, J-Y., “The effects of anatomical information and observer expertise on abnormality detection task,” SPIE Medical Imaging, Orlando, (2011).
- [8] Fuderer, M., “The information content of MR images,” IEEE Transactions on Medical Imaging, vol. 7, no. 4, pp. 368-380, (1988).