Perceptual Difference Model (Case-PDM) for Evaluation of MR Images

J. Miao[1], D. Huo[2], and D. Wilson[1,3]

[1]Biomedical Engineering, Case Western Reserve University, Cleveland, Ohio, United States, [2]Keller Center for Imaging Renovation, Barrow Neurological Institute, Phoenix, Arizona, United States, [3]Radiology, Case Western Reserve University, Cleveland, Ohio, United States

INTRODUCTION

There is an extraordinary number of fast MR imaging techniques, especially for parallel imaging [1]. When one considers multiple reconstruction algorithms, reconstruction parameters, coil configurations, acceleration factors, noise levels, and multiple test images, one can easily create 1000's of test images for image quality evaluation. We have developed a Perceptual Difference Model (Case-PDM) which identifies perceptible differences between a "fast," possibly degraded image and a slow, high quality "gold standard" image (Fig. 1). Case-PDM objectively, quantitatively evaluates image quality, and we have found it to be quite useful in investigations of keyhole, spiral, SENSE, and GRAPPA applications [2-4]. In this new study, we have comprehensively compared human evaluation of MR image reconstructions to that from Case-PDM and other similar image quality models. To test the range of applicability, we compared results across multiple image types (brain, heart, etc.) and reconstruction algorithms. In some instances, it is desirable to obtain "fast" images imperceptibly different than the gold standard images. We investigated the possibility of determining a threshold PDM corresponding to "imperceptible difference".

METHODS

To compare PDM and human evaluation over a range of image qualities, we performed DSCQS (Double-Stimulus Continuous Quality-Scale) experiments. Three different image types and three different reconstruction algorithms were tested. To compare PDM scores to imperceptible differences under low-degradation conditions, we designed and performed 2AFC (Two-Alternative Forced Choice) experiments, where the GUI is shown in Figure 2a, with test images generated with two image types and three degradation patterns. Human subjects included both radiologists and engineers.

RESULTS

For the DSCQS experiment, Case-PDM was highly correlated ($r > 0.9$) with human subject ratings over 120 images and 3 reconstruction algorithms (Table 1). Case-PDM performed better than the widely used mean-squared-error (MSE) and NASA's DCtune, and performs similarly with the Image Difference Matrix (IDM, version 2.0, Sarnoff Corporation). For the AFC experiment, the threshold of imperceptible difference is obtained experimentally and its value varies with image types and degradation patterns (varies from 0.6 to 1.8 based on our experiment data and details can be seen in Table 2). Figure 2b shows result from a typical AFC experiment where d' is detectability index used in detection theory [5]. Base on the result data from the experiment, we found that the value of PDM score can be comparable only for the images with similar types and similar degradation patterns.

DISCUSSION AND CONCLUSSION

Thousands of MR images can easily be generated in studies aimed at optimizing fast MR imaging techniques. Case-PDM provides an efficient and reliable method for evaluating such images. In this study, we show that not only can images be ordered with regards to image quality, one can assess images which are "perceptually equal" to an original high quality, but slowly acquired image. This report is the most comprehensive evaluation to date of Case-PDM as applied to MRI. We conclude that for image quality ratings, Case-PDM could faithfully represent the human subject responses over a large range of image quality. Although Case-PDM is a very useful tool for comparing "similar images with similar degradation pattern," one should be careful when interpreting PDM scores across MR images.

Table 1. Comparison of PDM with other similar models

Dataset	SENSE	SPIRAL	GRAPPA
# of images	40	40	40
Size of image	256x256	128x128	209x256
	Pearson Correlation Coefficient (prediction accuracy)		
PDM	0.94447	0.97202	0.91383
IDM	0.95471	0.97069	0.95807
MSE	0.71223	0.82462	0.89063
DCTune	0.71519	0.19131	0.31936
	RMS (prediction accuracy)		
PDM	10.6432	5.9356	12.8692
IDM	9.6371	6.0730	9.0805
MSE	22.7353	14.2932	14.4099
DCTune	22.6377	24.8004	30.0297
	Spearman Rank-Order Correlation Coefficient (prediction monotonicity)		
PDM	0.9711	0.9740	0.9321
IDM	0.9780	0.9444	0.9634
MSE	0.9462	0.8678	0.9546
DCTune	0.8150	0.0914	0.4587
	Outlier Ratio (prediction consistency)		
PDM	0.0500	0.0000	0.0250
IDM	0.0000	0.0000	0.0000
MSE	0.0500	0.0000	0.0250
DCTune	0.0500	0.0000	0.0000

Table 2. Data Analysis for AFC Experiment

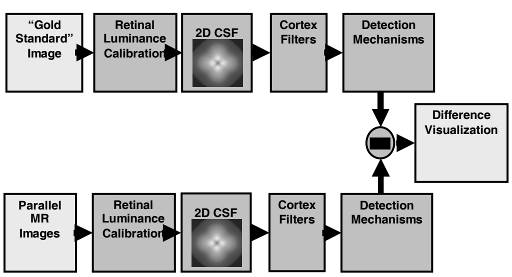
	Correlation	Threshold
Brain-Noise	0.77	0.7949
Brain-Blur	0.79	0.9024
Brain-Recon	0.80	0.5983
Abdomen-Noise	0.45	1.7790
Abdomen-Blur	0.79	1.4980
Abdomen-Recon	0.73	1.1177

ACKNOWLEDGEMENT

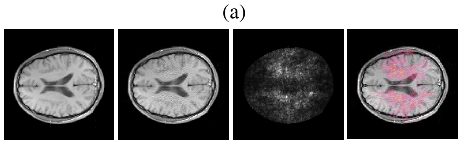
This work was supported under NIH grant R01 EB004070 and the Research Facilities Improvement Program Grant NIH C06RR12463-01. We thank Yuhao Jiang for helpful discussions and the subjects for participating in the experiments.

REFERENCES

[1] M. Blaimer, Top Magn Reson Imaging 2004 [2] KA Salem, Journal of Electronic Imaging, 2002 [3] D Huo, J Magn Reson 2006 [4] D Huo, ISMRM 2006 [5] NA MacMillan, Detection Theory, LEA publisher

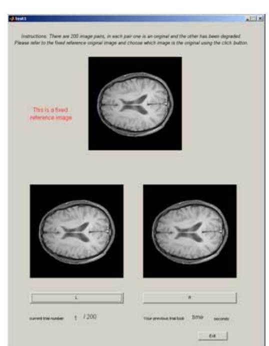


(a)

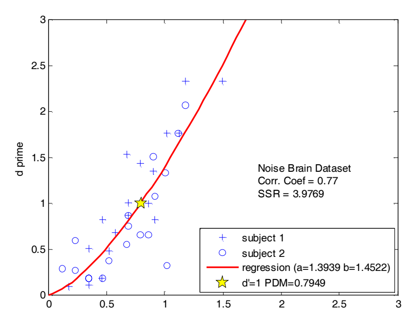


(b) (c) (d) (e)

Figure 1. Block diagram of the perceptual difference model (Case-PDM) is shown in (a). The inputs of the model are two images, a reference image (b) and a test image (c). The output is a spatial map (d) showing the perceived difference between two images. PDM could be used to tell the visual difference between two input images, as shown in the overlaid display in (e).



(a)



(b)

Figure 2. The GUI for a 2AFC task is shown in (a).The human subjects were instructed to choose the image from the two test images (at the bottom) that s/he thinks the same as the original image (on the top), by clicking on the left (L) or right (R) button. The results from a typical AFC experiment and the relationship between d' and PDM predictions were shown in (b).