

8-1-2016

Diagnostic quality assessment of compressed sensing accelerated magnetic resonance neuroimaging.

Mohammad Kayvanrad

Amy Lin

Rohit Joshi

Jack Chiu

Terry Peters

Follow this and additional works at: <https://ir.lib.uwo.ca/robertspub>



Part of the [Bioimaging and Biomedical Optics Commons](#)

Citation of this paper:

Kayvanrad, Mohammad; Lin, Amy; Joshi, Rohit; Chiu, Jack; and Peters, Terry, "Diagnostic quality assessment of compressed sensing accelerated magnetic resonance neuroimaging." (2016). *Robarts Imaging Publications*. 14.

<https://ir.lib.uwo.ca/robertspub/14>

Diagnostic quality assessment of compressed sensing accelerated magnetic resonance neuroimaging

M. Kayvanrad^{a,b,*}, A. Lin^c, R. Joshi^c, J. Chiu^c, T. Peters^{a,b,c}

^a *Imaging research laboratories, Robarts Research Institute, Western University, Canada*

^b *Biomedical Engineering program, Western University, Canada*

^c *Department of medical imaging, Western University, Canada*

* *Correspondence to: M. Kayvanrad, Robarts Research Institute, 100 Perth Drive, London, ON, Canada. Email: mkayvanrad@robarts.ca*

Abstract

Purpose: The aim of this study was to determine the efficacy of compressed sensing reconstructions for specific clinical neuroimaging applications of magnetic resonance imaging beyond more conventional k-space under-sampling approaches such as parallel imaging and simple low-resolution acquisitions.

Methods: Four routine clinical neuroimaging pulse sequences were chosen for this study due to their long acquisition time. In a series of blinded studies, three board-certified radiologists independently evaluated compressed sensing, parallel imaging, and low-resolution images at up to 5x accelerations. Experiments on synthetic brain images with artificial but realistic lesions were carried out to assess diagnostic accuracy for the detection of non-specific white matter lesions, permitting controlled evaluation of shift-variant compressed sensing reconstructions.

Results: Ringing and blurring were identified as the primary artifacts that hinder diagnostic quality of combined compressed sensing and parallel imaging reconstructions. The findings indicate that up to 5x acceleration is possible by a combined compressed sensing and parallel imaging reconstruction. However, efficacy of compressed sensing reconstructions as well as the improvement in image quality over the more conventional parallel imaging and low-resolution acquisitions appear to vary with pulse sequence.

Conclusion: Mild to moderate accelerations are possible for those sequences by a combined compressed sensing and parallel imaging reconstruction while maintaining diagnostic quality of reconstructions. Nevertheless, for certain sequences/applications one might mildly reduce the acquisition time by appropriately reducing the imaging resolution while maintaining diagnostic quality/accuracy, rather than the more complicated compressed sensing reconstruction.

Keywords- Accelerated magnetic resonance imaging; Parallel imaging; Compressed sensing MRI; Magnetic resonance neuroimaging; Diagnostic quality assessment; Diagnostic accuracy.

Introduction

The appearance of an MR image is the result of trade-offs between imaging speed, spatial resolution, signal-to-noise ratio (SNR), and image artifacts [1]. Due to considerations such as cost efficacy, patient comfort, and motion sensitivity of acquisitions, acceleration of MR imaging

has been an active field of research since the introduction of this modality and there remains a critical need for continuous improvement of the trade-off between speed and “image quality.” [2]

In addition to advancements in hardware and pulse sequence design, MRI acceleration techniques based on under-sampled k-space acquisitions are finding widespread applications. Two major categories of such techniques include parallel imaging (PI) [3]–[5] and compressed sensing (CS) [6]. While with PI missing k-space data are interpolated based on a priori knowledge of the coil sensitivity profiles, CS interpolates the missing data by imposing an a priori transform domain sparsity constraint to regularize the reconstruction problem [7]. Due to the independence of CS and PI reconstruction constraints, a combined CS and PI reconstruction allows for further acceleration [8]. In this work we explored the latter approach to accelerated MRI acquisitions (where multi-channel data were available).

The objective of this work was to study the efficacy of CS reconstructions for specific clinical neuroimaging applications of MRI. In particular this study was initiated in an attempt to answer the question of to what degree current CS techniques are capable of accelerating such clinical neuroimaging sequences beyond more conventional k-space under-sampling approaches such as PI or simple low-resolution acquisitions.

Defining image quality remains a challenging and controversial aspect of clinical MRI, particularly with the emergence of complex acquisition strategies such as those aimed at accelerating MR imaging by partial k-space acquisitions [9]. A variety of quality metrics for medical images are used [10]. Nevertheless, a consideration of medical image quality is currently dominated by a subjective and observational perspective in the clinical practice of radiology [9], [10]. Accordingly, we based our study on blinded clinical assessment of under-sampled MRI reconstructions.

In this paper we present the results of blinded clinical assessment of four routine neuroimaging pulse sequences at up to 5 acceleration factors. The study provides insight on the clinical efficacy of CS reconstructions in comparison with that of PI reconstructions and low-resolution acquisitions (zero-padding), which was used as a baseline since the simplest approach to reducing the acquisition time is to appropriately reduce the imaging resolution.

Furthermore, since compressed sensing reconstructions are inherently shift-variant and location-dependent, further experiments were designed based on synthetic brain images with artificial but realistic non-specific lesions embodied in random locations in the cerebral white matter. Design and creation of such synthetic data permit controlled evaluation of location-dependent reconstructions. Since the location of each lesion is known, it can be used as the reference for the assessment of diagnostic accuracy for the reconstructed images.

Methods

Data acquisition

Raw k-space data were acquired on a 3T GE scanner (Discovery 750, software revision 22M32, General Electric Healthcare, Waukesha, WI) with a 32-channel head coil using a protocol approved by the institutional Office of Research Ethics. Four pulse sequences were chosen for this study: T2-FLAIR, multiacquisition SSFP (also known as CISS, or FIESTA-C), time of flight (TOF) MR angiography, and spoiled gradient echo (SPGR). These sequences were chosen due to

their long acquisition time and widespread use in clinical neuroimaging and epilepsy protocols. T2-FLAIR is the workhorse of neuroimaging protocols and a T2-FLAIR sequence was also chosen to generate synthetic brain images. The acquisition parameters for these pulse sequences are summarized in Table 1.

Blinded studies	T2-FLAIR	TR/TE = 8000/120.8 ms, TI = 2250 ms, flip angle = 90°, matrix = 256 x 256 (Pixel spacing = 0.859 mm isotropic), Number of slices = 28, slice thickness = 3 mm, slice spacing = 5 mm, BW = ±31.3 kHz
	FIESTA-C	TR/TE = 5.51/2.42 ms, flip angle = 55°, matrix: 288 x 288 (pixel spacing = 0.625 mm isotropic), Number of slices = 56, slice thickness = 1 mm, slice spacing = 1 mm, BW = ±46.9 kHz
	TOF	TR/TE = 20/2.6 ms, flip angle = 15°, matrix = 320 x 224 (pixel spacing = 0.750 x 1.07 mm), Number of slices = 70, slice thickness = 1.4 mm, slice spacing = 1.4 mm, BW = ±19.5 kHz, number of slabs = 3
	SPGR	TR/TE = 3.71/8.36 ms, flip angle = 18°, matrix: 220x220x160, resolution = 1 mm isotropic, BW = ±16.5 kHz
Synthetic images	T2-FLAIR	TR/TE = 8000/120.9 ms, TI = 2250 ms, flip angle = 90°, matrix = 256 x 256 (Pixel spacing = 0.859 mm isotropic), Number of slices = 56, slice thickness = 2 mm, slice spacing = 2.5 mm, BW = ±31.3 kHz

Table 1- Prescribed acquisition parameters for the pulse sequences used in this study

Image reconstruction

Compressed sensing reconstructions were performed using an iterative stationary wavelet transform (SWT) thresholding algorithm [11]. Where multiple-channel data were available, the multiple channel version of the algorithm was used, which would amount to a combined CS and PI reconstruction, hereafter simply referred to as the CS reconstruction. The iterative SWT reconstruction was performed using the second Daubechies wavelet (D4) and soft thresholding with a practically optimal threshold obtained automatically based on the Birgé-Massart strategy [11], [12]. A fixed 50 iterations were set as the point of convergence.

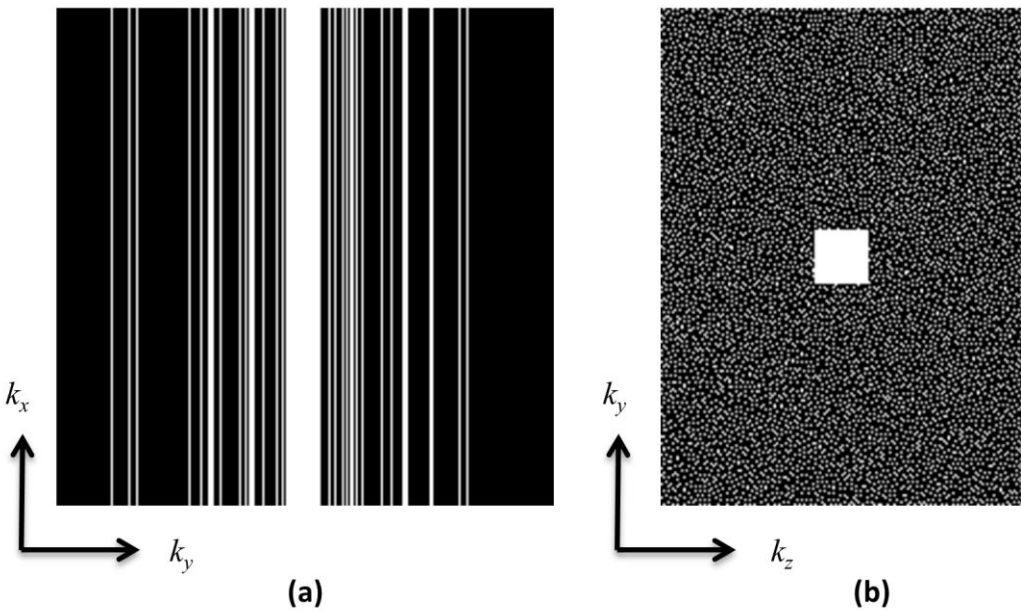


Figure 1- Illustration of variable density and Poisson disk sampling for compressed sensing. (a) Variable density sampling at 5x acceleration with 6% of the central phase-encode lines fully sampled amounting to autocalibration data. (b) Poisson disk sampling at 5x acceleration with autocalibration data amounting to the central 2% of k-space.

For CS reconstructions, k-space data were retrospectively under-sampled using a variable density sampling scheme with 6% of the central phase-encode lines fully sampled amounting to autocalibration data (Figure 1(a)). For 3D sequences, i.e., sequences for which k-space is phase-encoded in both k_y and k_z directions, k-space data were also under-sampled using a Poisson disk sampling scheme with autocalibration data amounting to the central 2% of k-space in the two phase-encode directions (Figure 1(b)). Of interest was to observe how the choice of the sampling scheme would affect the perception of the reconstruction quality.

Parallel imaging was performed using the traditional GRAPPA reconstruction [3]. K-space data were appropriately under-sampled for GRAPPA by sampling phase-encode lines in regular intervals with 12.5% of the central phase-encode lines fully sampled amounting to autocalibration data.

Low-resolution reconstructions were obtained by sampling the central phase-encode lines at the desired acceleration, providing a baseline for accelerated imaging. Low-resolution data were zero-padded to the desired matrix size with Butterworth apodization to eliminate ringing artifacts.

Study design

Blinded studies

The goal of the blinded studies was to assess the quality of accelerated reconstructions on a number of commonly used neuroimaging pulse sequences (Table 1). Whole brain data sets of healthy volunteers were acquired and reconstructed as previously described for 2x, 3x, 4x, and 5x accelerations. CS with variable density sampling (hereafter referred to as VDCS), GRAPPA, and low-resolution (hereafter referred to as lowres) reconstructions were performed on all pulse sequences with an additional CS reconstruction with Poisson disk sampling [8] appropriate for 3D sequences (hereafter referred to as PDCS) performed on the SPGR data set due to the availability of 3D data, yielding a total of 52 whole brain reconstructed data sets.

Three board-certified radiologists independently assessed the diagnostic quality of the images. First, reconstructed images were randomly permuted and presented to the readers for blinded review. The readers scored image quality based on a 5-point scale: 1 = Not interpretable, 2 = Severely degraded, 3 = Moderately degraded, 4 = Mildly degraded, 5 = No significant artifacts. For consistency, all three readers based their assessment of diagnostic quality on the criteria summarized in Table 2 for each pulse sequence. Wilcoxon signed rank tests were performed in R (R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria) to test the null hypothesis that there was no significant difference in diagnostic quality between VDCS and PI, PDCS (where available) and PI, VDCS and lowres, PDCS (where available) and lowres, PI and lowres, and PDCS (where available) and VDCS reconstructions. Inter-rater reliability (IRR) analysis was performed using a two-way mixed, consistency, average-measures ICC [13] to assess the degree that readers provided consistency in their ratings of diagnostic quality.

For T2 FLAIR, FIESTA-C, and TOF sequences, the reconstructions were also directly compared in addition to the individual assessments. The images were presented to the reader in blinded randomized order for side-by-side comparison at each acceleration. The readers, who were blinded to the type of reconstruction and the acceleration factor, were permitted to simultaneously scroll through all the three image volumes side-by-side and ranked the images in the order of diagnostic quality preference, again based on the criteria shown in Table 2. Wilcoxon signed rank tests were performed on numerical preference values (with 1 being the most preferred and 3 being the least proffered) to test the null hypothesis that there was no significance difference in quality preference between VDCS and PI, VDCS and lowres, and PI and lowres reconstructions. IRR was again assessed using a two-way mixed, consistency, average-measures ICC. We were unable to perform side-by-side comparison for SPGR data due to software limitations.

In all the aforementioned analysis, the significance of findings was evaluated at a significance level of 0.05. The ICCs were interpreted based on the commonly-cited Cicchetti cutoffs [14], with IRR being poor for ICC values less than .40, fair for values between .40 and .59, good for values between .60 and .74, and excellent for values between .75 and 1.0.

Pulse sequence	Criteria for diagnostic quality assessment
T2 FLAIR	Delineation of basal ganglia, thalamus, grey-white matter junction, subarachnoid space, ventricles
FIESTA-C	Delineation of cranial nerves
TOF	Delineation of cerebral arteries
SPGR	Delineation of basal ganglia, thalamus, grey-white matter junction, subarachnoid space, ventricles

Table 2- Criteria for diagnostic quality assessment of under-sampled reconstructions for each pulse sequence

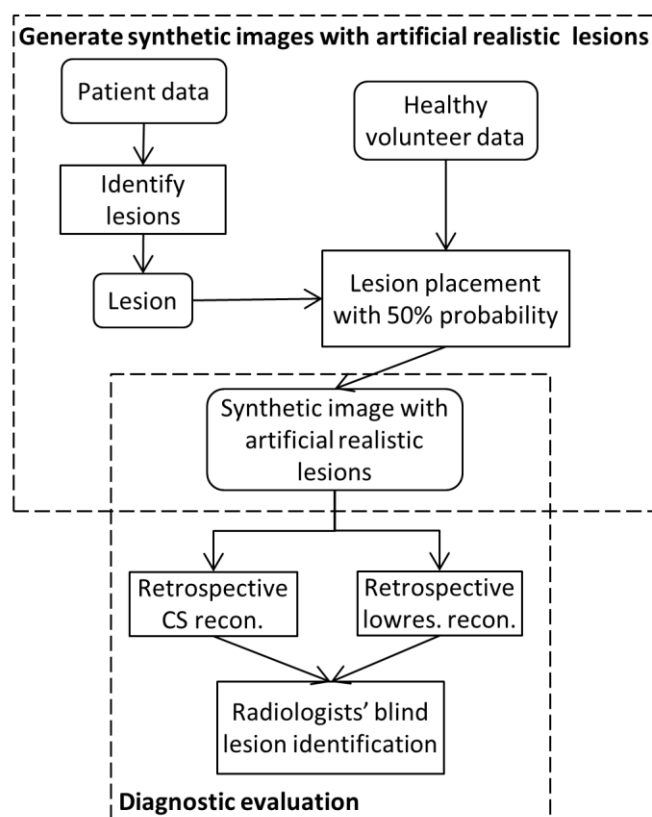


Figure 2- Study design for synthetic images: First synthetic images with realistic lesions are generated by artificially incorporating a lesion into random locations on the cerebral white matter based on a uniform probability density function (pdf) on ROIs a priori defined on axial T2-FLAIR images of a healthy volunteer. The synthetic images are then retrospectively reconstructed for radiologists' blinded diagnostic evaluation.

Synthetic images

The goal of these experiments was to assess diagnostic accuracy for CS reconstructed images for the detection of non-specific white matter lesions. As noted, use of synthetic images

appropriately generated for the study permits controlled evaluation of shift-variant reconstructions such as CS. Figure 2 shows a flowchart of the study design, which consists of two parts: (1) generation of synthetic images with realistic lesions, and (2) diagnostic evaluation of the reconstructions.

In order to generate synthetic images with realistic lesions, white matter lesions were identified on T2-FLAIR brain images of multiple sclerosis patients. A 2 mm small round lesion was selected as a “typical” lesion and synthetic images were generated by artificially incorporating the lesion into random locations on the cerebral white matter based on a uniform probability density function (pdf) on ROIs a priori defined on axial T2-FLAIR images of a healthy volunteer. The readers were blinded to both the lesion and the ROIs. In order to preserve the SNR, lesion placement was carried out by manipulating the intensity levels on the destination image to match those of the lesion relative to its background on the source image. An example of synthetic images with a lesion is shown in Figure 10(a). Lesion placement was randomized with a probability of 50% such that 50% of the synthetic images had a lesion incorporated in them and the rest were lesion-free. Synthetic k-space data were generated by 2D fast Fourier transformation (2DFFT) and the data were reconstructed at 1x, 2x, 3x, 4x, and 5x accelerations for VDCS and lowres, where 1x acceleration amounts to a fully-sampled reconstruction. A data set consisting of 30 images per reconstruction per acceleration was generated amounting to a total of 300 images.

Diagnostic evaluation of the reconstructions was performed by three board-certified radiologists who independently and blindly viewed the images in randomized orders. The readers either identified a lesion on the image or declared there to be none. The readers also declared their level of confidence using a 4-point scoring scheme (1: non-diagnostic; 2: low confidence; 3: moderately confident; 4: high confidence) for each image. ROC analysis of the results was performed based on the confidence levels declared by the readers following the methodology of Metz [15]. Reconstruction error was computed in terms of the normalized root mean square error (NRMSE) with respect to the fully-sampled reference, providing a quantitative measure of reconstruction quality.

Results

Blinded studies

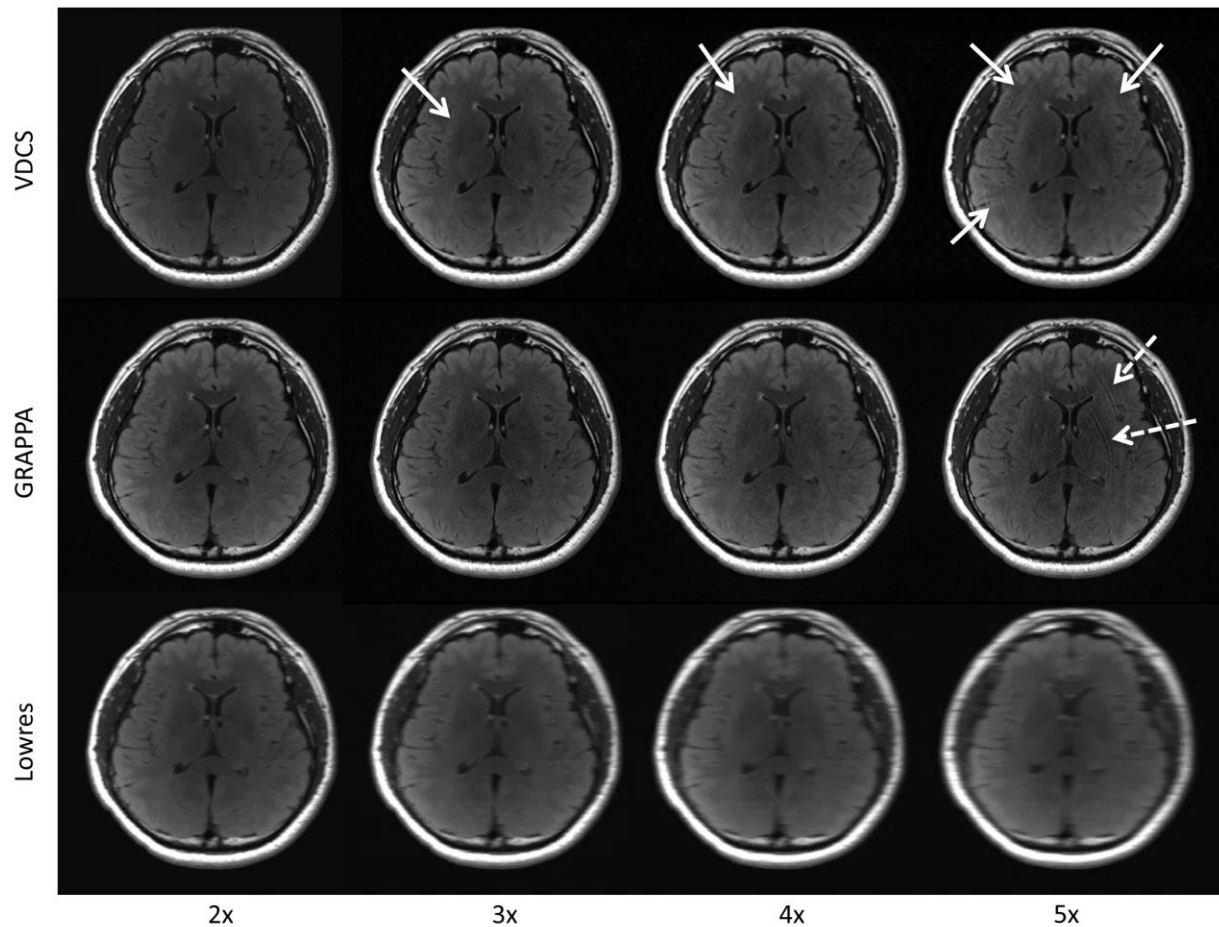


Figure 3- T2-FLAIR images reconstructed using variable density compressed sensing (VDCS) and GRAPPA, as well as low-resolution images at 2x, 3x, 4x, and 5x acceleration. The solid arrows highlight areas on which ringing artifacts were most apparent for the VDCS reconstructions. There is also slight blurring on VDCS reconstructions as acceleration increases. The dashed arrows highlight areas with prominent artifacts for the GRAPPA reconstructions. There are global “grainy” artifacts on the GRAPPA reconstructions and severe blurring on lowres images with increasing acceleration.

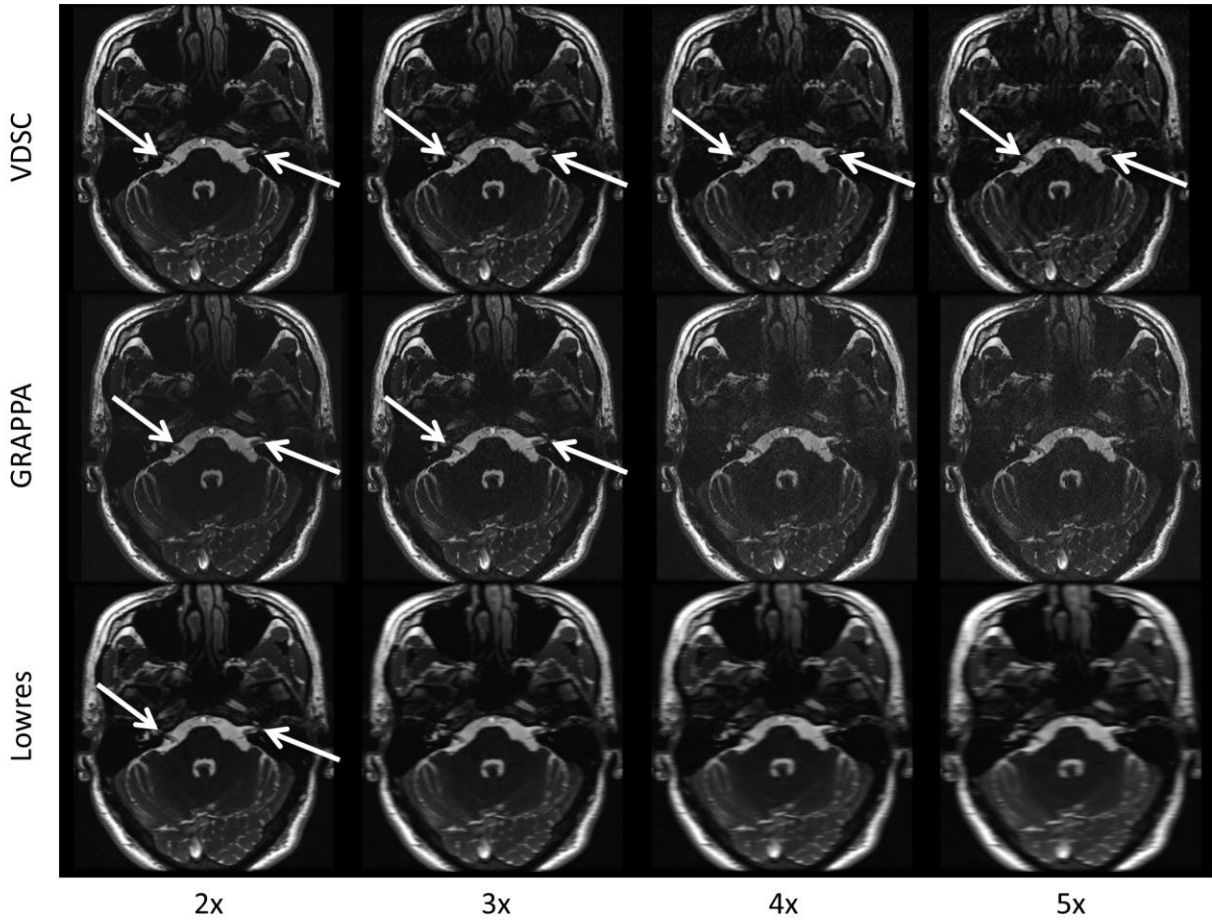


Figure 4- FIESTA-C images reconstructed using variable density compressed sensing (VDCS) and GRAPPA, as well as low-resolution images at 2x, 3x, 4x, and 5x acceleration. Note delineation of CN VII- facial and CN VIII- vestibulocochlear nerves (white arrows) on 2x, 3x, 4x, and 5x VDCS, as well as on 2x, and 3x GRAPPA, and 2x lowres images. There is global ringing artifacts and slight blurring on VDCS reconstructions, global “grainy” artifacts on GRAPPA reconstructions, and severe blurring on lowres images, with increasing acceleration.

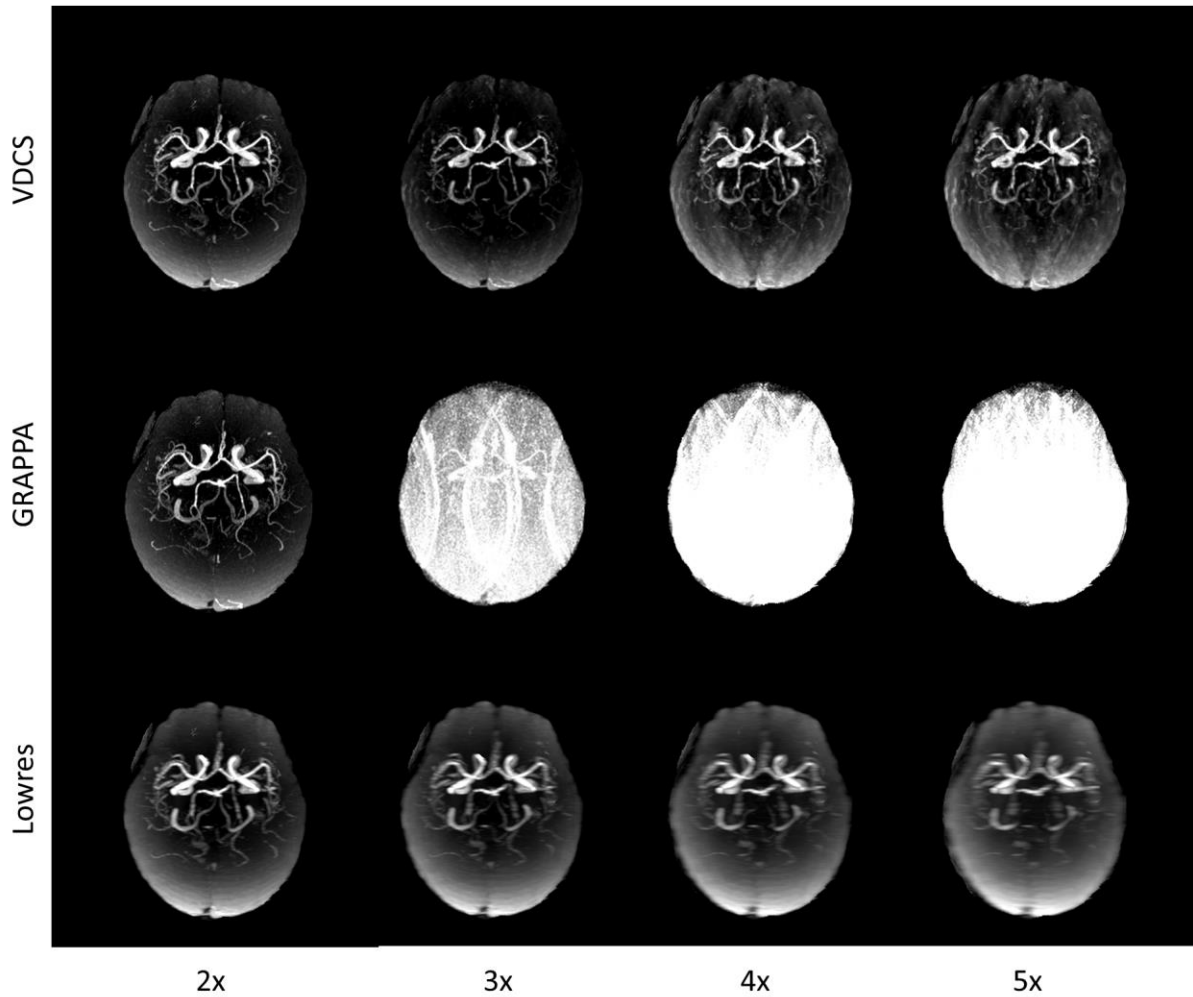


Figure 5- Maximum intensity projection (MIP) images of TOF data reconstructed using variable density compressed sensing (VDCS), GRAPPA, and low-resolution at 2x, 3x, 4x, and 5x acceleration. There are global ringing artifacts on VDCS images, and severe blurring on lowres images, with increasing acceleration. GRAPPA failed to produce acceptable images at 3x, 4x, and 5x accelerations due to limited SNR of the original TOF acquisitions used in this study.

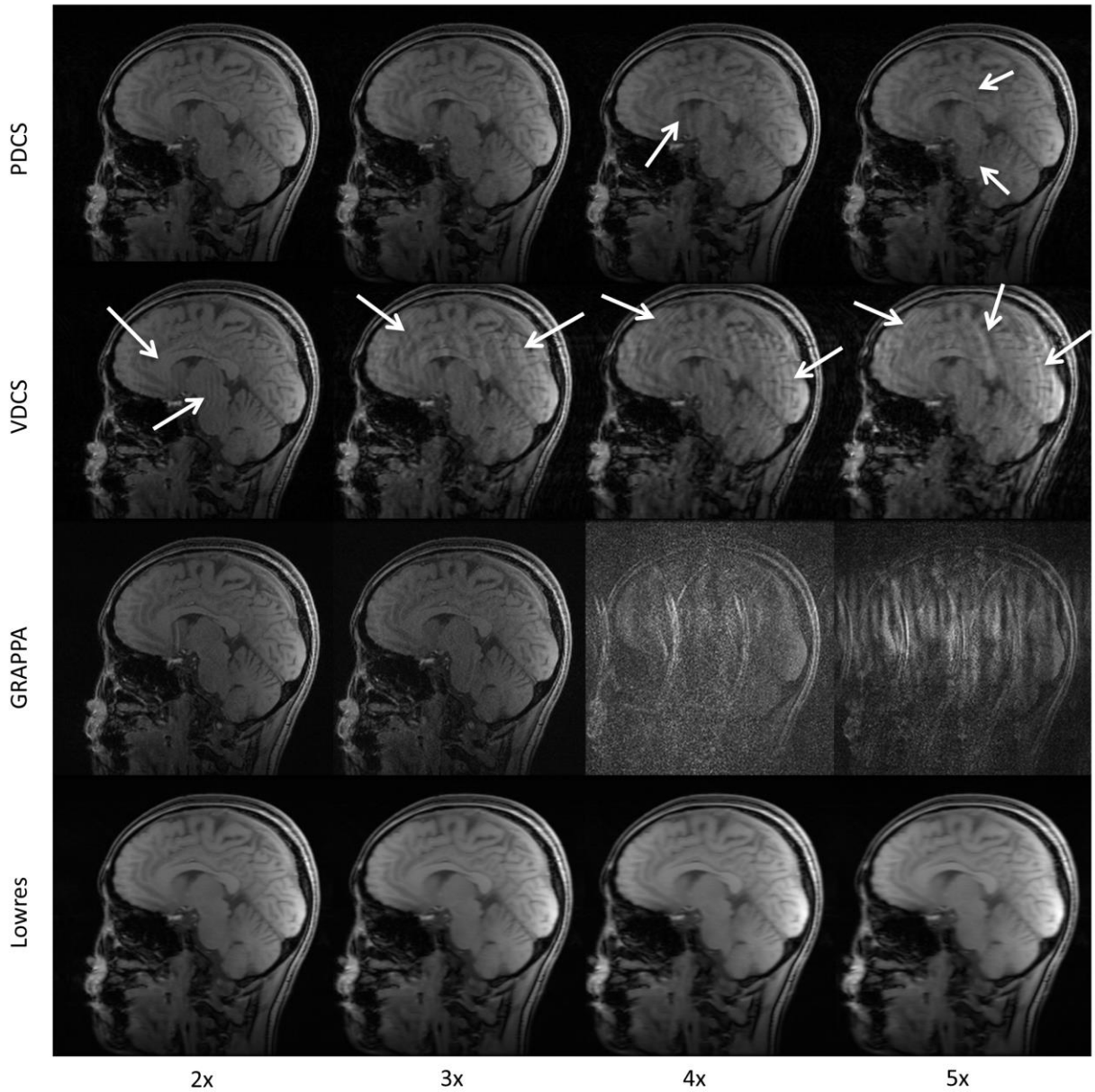


Figure 6- SPGR images reconstructed using variable density compressed sensing (VDCS), Poisson disk sampling compressed sensing (PDCS), and GRAPPA, as well as low-resolution images at 2x, 3x, 4x, and 5x acceleration. The solid arrows highlight areas on which ringing artifacts were most apparent for the CS reconstructions. There is also slight blurring on PDCS reconstructions, and severe blurring on VDCS reconstructions as acceleration increases. There are global “grainy” artifacts on the GRAPPA reconstructions and severe blurring on lowres images with increasing acceleration. GRAPPA failed to produce acceptable images at 4x and 5x accelerations due to limited SNR of the original SPGR acquisitions used in this study.

Figures

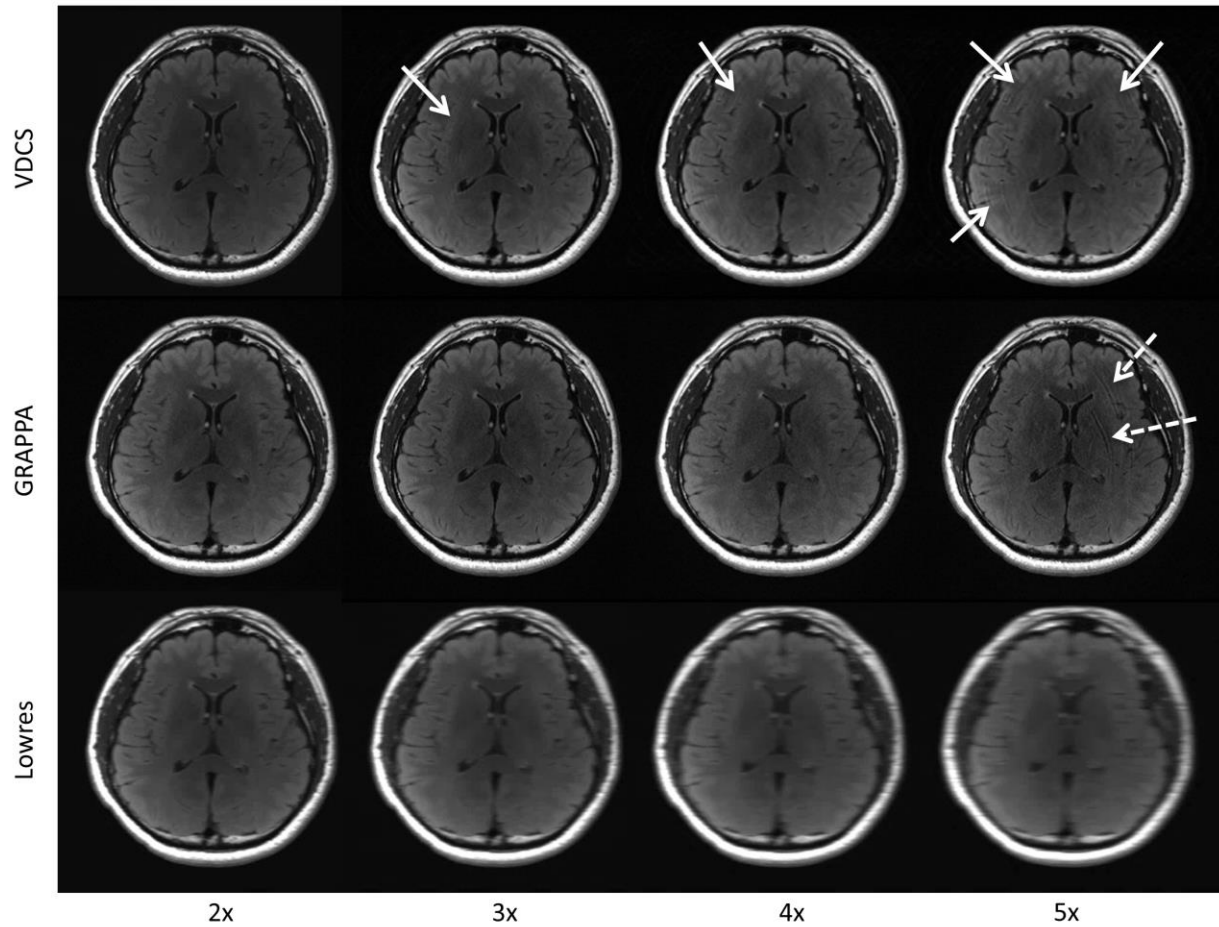


Figure 3, Figure 4, Figure 5, and Figure 6 show examples of under-sampled reconstructions for T2-FLAIR, FIESTA-C, TOF, and SPGR data, respectively. The primary artifact identified on PDCS reconstructions was ringing. The artifacts on VDCS images were characterized as ringing and blurring. Blurring was identified as the primary artifact on the lowres images. GRAPPA artifacts were characterized as “noisy” and “grainy”.

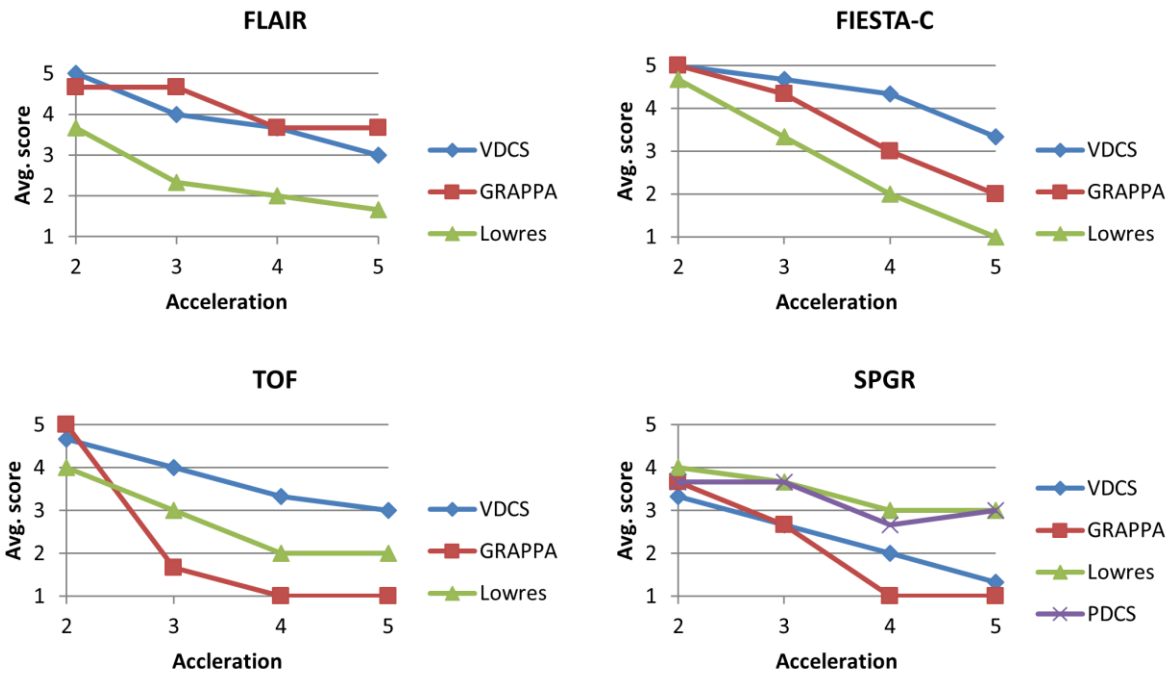


Figure 7- Overall qualitative scores, averaged over the readers, for each pulse sequence. Scoring was based on a 5-point scale: 1 = Not interpretable, 2 = Severely degraded, 3 = Moderately degraded, 4 = Mildly degraded, 5 = No significant artifacts

	T2-FLAIR		FIESTA-C		TOF		SPGR
	Individual assessment	Side-by-side comparison	Individual assessment	Side-by-side comparison	Individual assessment	Side-by-side comparison	Individual assessment
VDCS vs. GRAPPA	p = 0.2986	p = 0.145	p = 0.05676	p = 0.1831	p < 0.05	p < 0.05	p = 0.3741
VDCS vs. lowres	p < 0.05	p < 0.05	p < 0.05	p < 0.05	p < 0.05	p < 0.05	p < 0.05
PDCS vs. GRAPPA	N/A	N/A	N/A	N/A	N/A	N/A	p < 0.05
PDCS vs. lowres	N/A	N/A	N/A	N/A	N/A	N/A	p = 0.6675
PDCS vs VDCS	N/A	N/A	N/A	N/A	N/A	N/A	p < 0.05
GRAPPA vs. lowres	p < 0.05	p < 0.05	p < 0.05	p < 0.05	p = 0.1449	p = 0.6501	p < 0.05

Table 3- Comparison of individual quality scores as well as numerical preference values based on side-by-side comparison of under-sampled reconstructions. Wilcoxon signed rank tests were performed to test the null hypothesis that there was no significant difference in diagnostic quality scores or in numerical preference values between each pair of reconstructions.

Figure 7 shows plots of average quality scores for individual assessment of under-sampled reconstructions. Wilcoxon signed rank test results for the comparison of individual quality scores as well as side-by-side comparison of under-sampled reconstructions are detailed in Table 3. For FLAIR and FIESTA-C data, image quality was superior with VDCS and GRAPPA reconstructions compared to lowres (Wilcoxon signed rank test, $p < 0.05$). However, no statistically significant difference was found between qualitative scores given to VDCS and GRAPPA reconstructions. For TOF data, VDCS was superior to GRAPPA and lowres reconstructions (Wilcoxon signed rank test, $p < 0.05$). No statistically significant difference was found between GRAPPA and lowres reconstructions. For SPGR data, PDCS and lowres were judged to be superior compared to VDCS and GRAPPA reconstructions (Wilcoxon signed rank test, $p < 0.05$). No statistically significant difference was found between PDCS and lowres, and between VDCS and GRAPPA reconstructions.

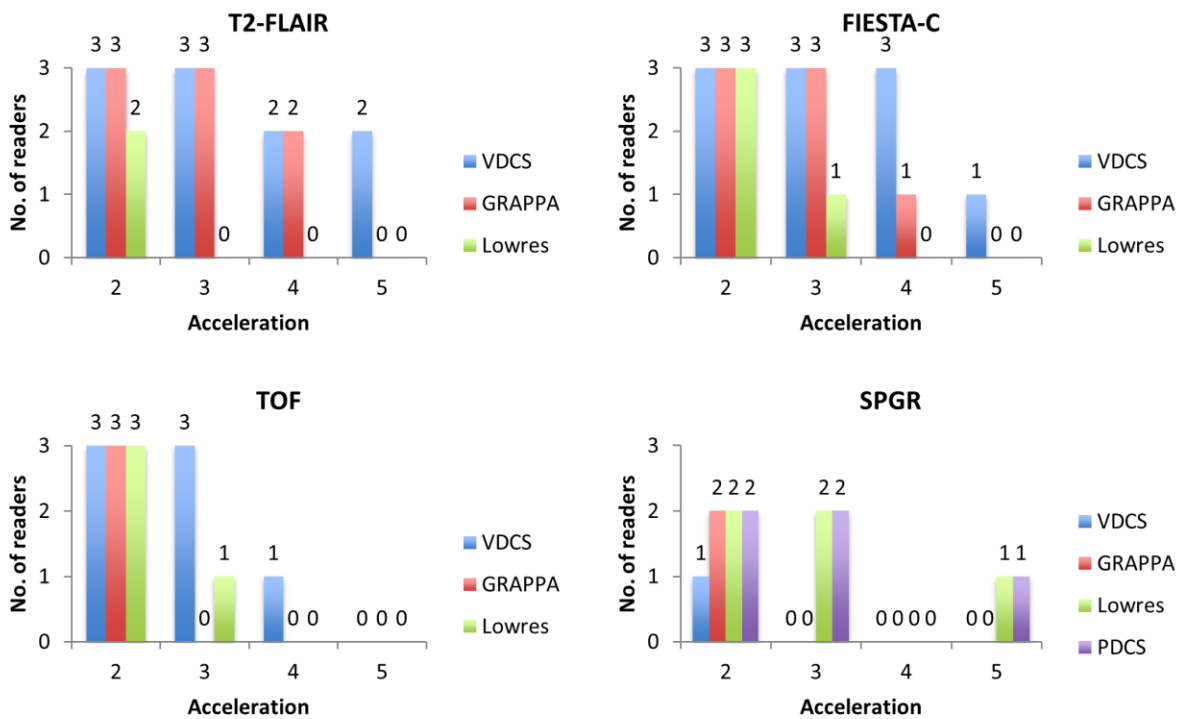


Figure 8- Subjective assessment of the reconstruction quality for each pulse sequence: number of readers who considered the reconstruction to be of clinical diagnostic quality is shown for each reconstruction and acceleration.

Figure 8 shows a bar plot of the number of readers who considered the reconstructions to be of clinical diagnostic quality. VDCS reconstructions were considered to be of diagnostic quality by all the readers at 2x and 3x acceleration for T2-FLAIR, FIESTA-C, and TOF image sets, while for FIESTA-C all the readers also considered VDCS reconstructions at 4x acceleration of

diagnostically acceptable quality, and 2 out of the 3 readers considered the VDCS reconstruction at 4x and 5x acceleration diagnostically acceptable for T2-FLAIR data, suggesting that VDCS reconstructions are of some clinical diagnostic quality at mild (2x and 3x) accelerations for TOF data, and at mild (2x and 3x) to moderate (4x and 5x) accelerations for FIESTA-C and T2-FLAIR data, with little to no diagnostic quality at higher accelerations. The results in Figure 8 generally suggest that VDCS is comparable with or superior to GRAPPA and lowres in terms of producing diagnostic quality images for these sequences. For SPGR data, 2 out of the 3 readers considered PDCS reconstructed images to be of diagnostic quality at 2x and 3x acceleration, suggesting that PDCS reconstruction is of some diagnostic quality at mild (2x and 3x) accelerations and of little or no diagnostic quality at higher accelerations for this sequence. The results are comparable with those of the lowres image sets and superior to those of VDCS and GRAPPA reconstructions.

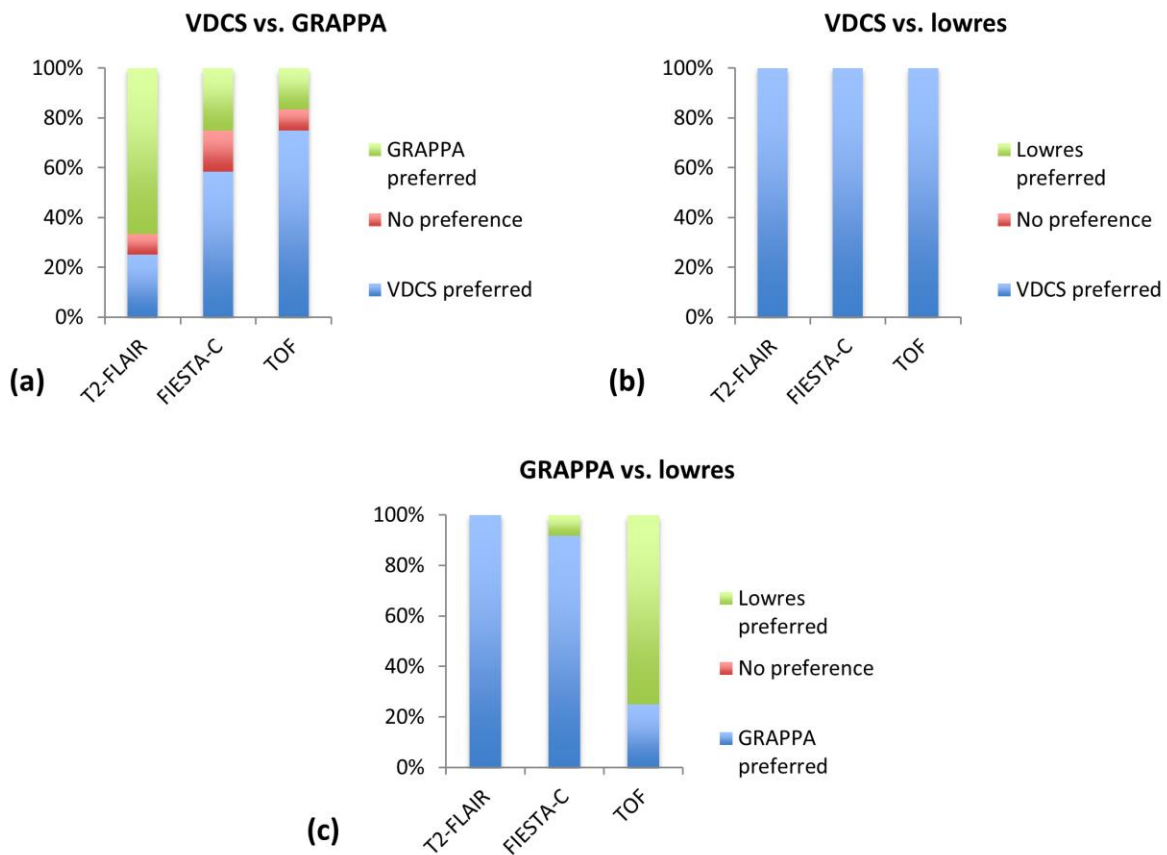


Figure 9- Side-by-side comparison of under-sampled reconstructions for T2-FLAIR, FIESTA-C, and TOF data sets: (a) VDCS vs. GRAPPA; (b) VDCS vs. lowres; (c) GRAPPA vs. lowres. Bar plots were calculated based on comparisons at all accelerations for each sequence.

Figure 9 shows bar plots summarizing readers' preference for side-by-side comparison of the reconstructions. VDCS reconstructed image sets were preferred over lowres image sets for T2-FLAIR, FIESTA-C, and TOF data (Wilcoxon signed rank test, $p < 0.05$). VDCS was preferred

over GRAPPA for TOF data (Wilcoxon signed rank test, $p < 0.05$). No statistically significant difference in image quality preference was found between VDCS and GRAPPA for T2-FLAIR and FIESTA-C data. GRAPPA reconstructed image sets were preferred over lowres image sets for T2-FLAIR and FIESTA-C data (Wilcoxon signed rank test, $p < 0.05$). However, no statistically significant difference in quality preference was found between GRAPPA and lowres image sets for TOF data. The results of side-by-side comparisons are generally consistent with those obtained based on individual quality assessment of the reconstructed image sets.

	T2-FLAIR	FIESTA-C	TOF	SPGR
Individual assessment	0.931	0.938	0.938	0.806
Side-by-side comparison	0.958	0.914	0.987	N/A

Table 4- Inter-rater reliability assessment using intra-class correlation (ICC) between the readers for individual quality assessment and side-by-side comparison of the reconstruction. The IRR was in the excellent range ($ICC > 0.75$) for individual assessment of the quality of reconstructions as well as for side-by-side comparison of reconstructions for all the pulse sequences.

Inter-rater reliability for individual quality assessments and side-by-side comparisons are shown in

	T2-FLAIR	FIESTA-C	TOF	SPGR
Individual assessment	0.931	0.938	0.938	0.806
Side-by-side comparison	0.958	0.914	0.987	N/A

Table 4. The IRR was in the excellent range ($ICC > 0.75$) for the individual assessment of the quality of reconstructions as well as for side-by-side comparison of reconstructions for all the pulse sequences, indicating that readers had a high degree of agreement and suggesting that reconstruction quality was rated similarly across readers. The high ICC suggests that a minimal amount of measurement error was introduced by the independent readers, and therefore statistical power for subsequent analyses is not substantially reduced.

Synthetic images

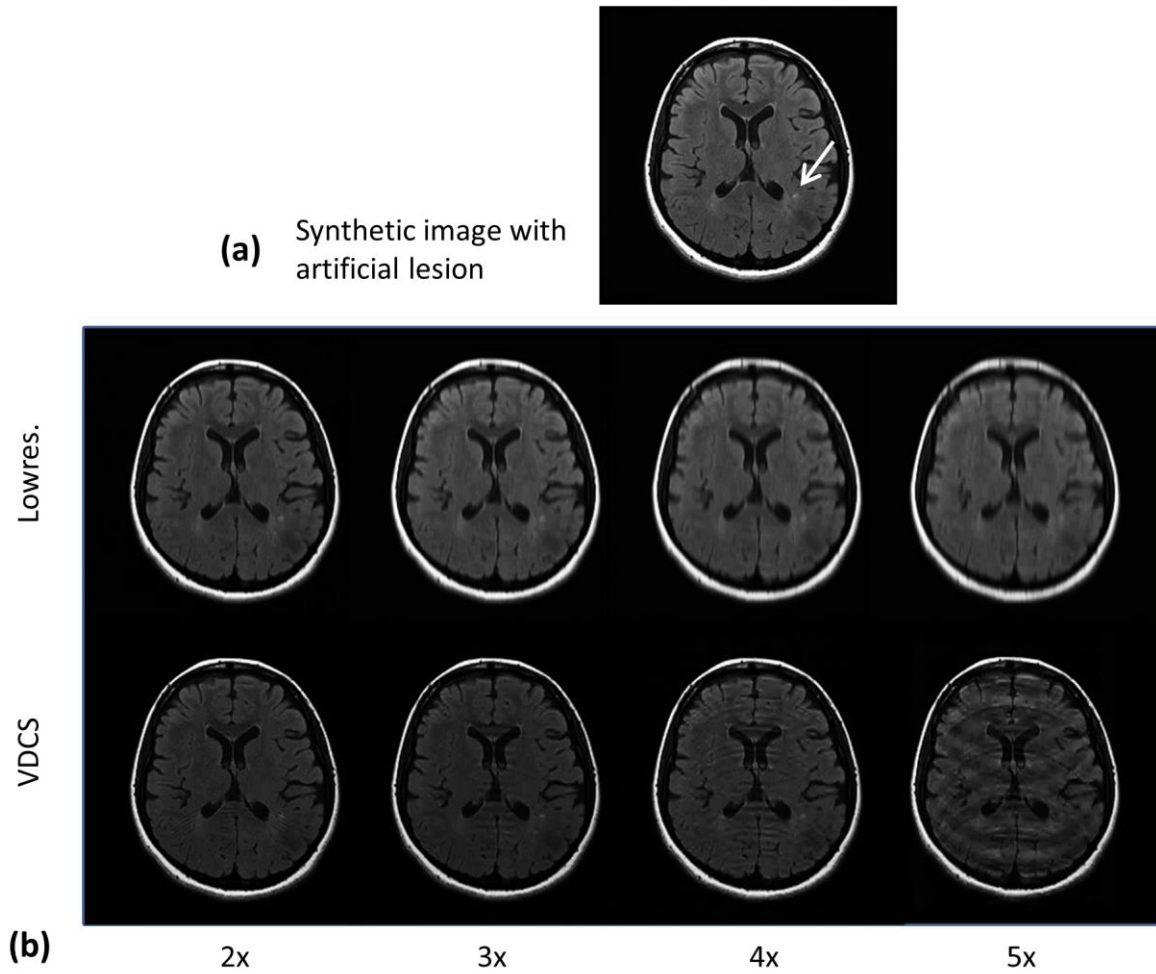


Figure 10- Example of synthetic images with an artificial but realistic lesion: (a) Synthetic image with a lesion- the arrow points to the artificially incorporated lesion. (b) The synthetic image retrospectively reconstructed using variable density compressed sensing (VDCS) and low-resolution sampling at 2x, 3x, 4x, and 5x acceleration.

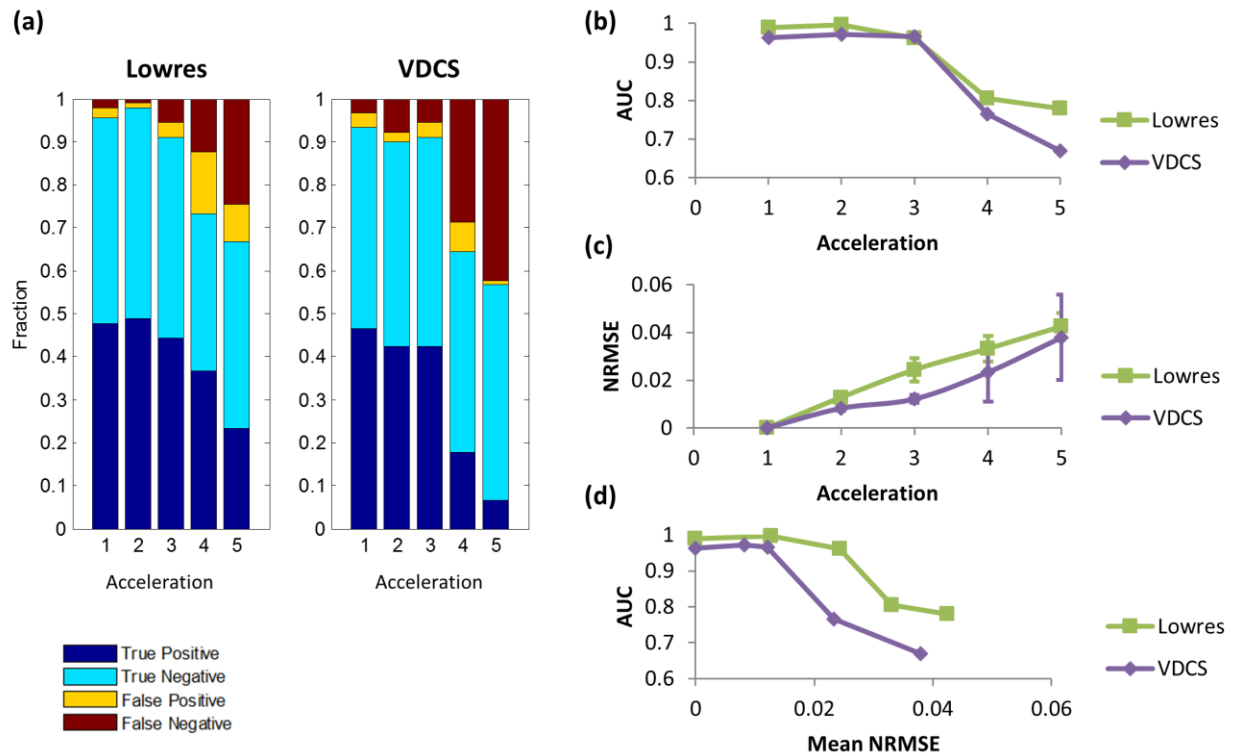


Figure 11- Diagnostic accuracy for variable density compressed sensing (VDCS) and low-resolution (lowres) images for the detection of non-specific white matter lesions on synthetic images: (a) Pooled results shown in terms of the true positive, true negative, false positive, and false negative fractions at each acceleration factor for VDCS and lowres images, (b) area under the ROC curve (AUC), and (c) mean normalized root mean square error (NRMSE) with respect to the fully-sampled image. The error bars on the NRMSE plot denote one standard deviation. (d) AUC plotted versus mean NRMSE at the corresponding acceleration factors.

Figure 10 shows an example of synthetic images with a lesion and its lowres and VDCS reconstructions. Figure 11 summarizes diagnostic accuracy for the reconstructed images. For mild (2x and 3x) accelerations, true positive rate (TPR), i.e., sensitivity, slightly decreased from 93% and 96% at 1x acceleration to 88% and 89% at 3x acceleration for the VDCS and lowres images, respectively. The decrease in TPR was more rapid for moderate (4x and 5x) accelerations to 14% and 49% at 5x acceleration for the VDCS and lowres images, respectively. No particular trend was observed in the true negative rate, i.e., specificity. As image quality degraded with increasing acceleration, the readers made more negative decisions with lower confidence, since the lesions had become less apparent due to blurring and ringing artifacts. This, consequently, resulted in a decrease of the true positive fraction, while the true negative fraction did not meaningfully change.

Area under the ROC curve (AUC) and the reconstruction error are shown in Figure 11 (b-d). There was little to no difference in AUC between 1x to 3x accelerations, while the data showed a decrease in AUC of 23% and 36% at for 4x and 5x CS images respectively, and of 21% and 24%

for 4x and 5x lowres images respectively, compared to the 1x images. Based on readers' consensus, diagnostic accuracy was acceptable for mild (i.e., 2x and 3x) accelerations and unacceptable for higher accelerations. In terms of AUC, diagnostic accuracy for VDCS reconstructions was comparable with that of lowres images at mild accelerations. However, for moderate accelerations, diagnostic accuracy was superior for lowres images compared to the VDCS images. Nevertheless, VDCS images had significantly lower reconstruction error than lowres images for 2x, 3x, and 4x accelerations (two sample student's t-test under the null hypothesis that there was no significant difference in mean NRMSE between VDCS and lowres images, $p < 0.001$). No meaningful correlation was found between AUC and NRMSE.

Discussion

Efficacy of CS reconstructions appears to vary with pulse sequence. For example, while the results suggest that for T2-FLAIR and FIESTA-C data sets CS reconstructions are of diagnostic quality at mild to moderate accelerations, for TOF and SPGR data sets CS reconstructions are of some diagnostic quality only at mild accelerations, with little or no diagnostic quality at higher accelerations. Furthermore, the improvement in image quality with a CS reconstruction over PI and lowres also varies with pulse sequence. For example, while the results show significant differences in median quality scores between CS images and PI and lowres images for TOF data, no statistically significant difference was found between CS and GRAPPA for T2-FLAIR and FIESTA-C data, and between CS and lowres for SPGR data.

The data show significant disparity between the quality of PDCS and VDCS reconstructions for SPGR images. While the results show that PDCS reconstructions are of some diagnostic quality at mild (2x and 3x) accelerations, VDCS reconstructions are of little or no diagnostic quality. It is generally accepted that a 2D under-sampling scheme (such as Poisson disk sampling) over both phase-encode directions (k_y and k_z) is more appropriate for 3D sequences, such as the SPGR sequence used in this study, than a 1D under-sampling scheme (such as variable density sampling) [8], [11].

An interesting perspective emerges from SPGR and synthetic data. SPGR data show no statistically significant difference in quality scores between PDCS and lowres images. In terms of synthetic data, no improvement in diagnostic accuracy was found for CS over lowres images. The results generally suggest that for these sequences/applications one can mildly reduce the acquisition time by appropriately reducing the imaging resolution rather than the more complicated CS reconstruction while maintaining diagnostic quality/accuracy.

In this study k-space data were retrospectively under-sampled with sampling schemes appropriate for each of the reconstructions, which permitted implementation and comparison of different reconstructions on each data set. Clearly, however, clinical implementations of CS require proper modifications of the pulse sequences for data acquisition based on the desired sampling pattern (see, for example, [16]–[18]).

For TOF and SPGR data, GRAPPA failed to produce acceptable images at accelerations higher than 2x (Figure 5) and 3x (Figure 6), respectively, which is attributed to the limited SNR of the TOF and SPGR sequences used in this study. Clearly, this limit can be improved by increasing the SNR, e.g., by increasing voxel volume. In this study, however, imaging parameters were chosen based on those of typical clinical sequences. In general, it is expected that the prescribed

pulse sequence parameters affect the quality of the reconstructions. A thorough investigation of the effect of SNR and imaging parameters on the reconstructions calls for additional studies and is beyond the scope of the current paper.

One of the limitations of the experiments based on synthetic images was lack of multiple-channel data. Although synthetic images permit controlled evaluation of location-dependent reconstructions, the conclusions are limited to a single-channel reconstruction. Since coil data are not available for synthetic images, CS reconstructions suffer from more prominent visual artifacts compared to their multiple-channel, i.e., combined CS and PI reconstruction, counterparts. (Compare Figure 10 and Figure 3.) Another limitation of these experiments was that the size of the lesions was fixed. While the experiments provide a framework for the assessment of diagnostic accuracy of the reconstructions, further studies should be performed to investigate the effect of the size of the lesions on the diagnostic accuracy for under-sampled reconstructions.

Conclusion

This work provided an investigation into the efficacy of conventional CS techniques for specific clinical neuroimaging applications of magnetic resonance imaging. Qualitative assessments showed that mild, i.e., 2x and 3x, to moderate, i.e., 4x and 5x, accelerations are possible for those sequences by a combined CS and PI reconstruction while maintaining the diagnostic quality of the reconstructions. Nevertheless, for certain sequences/applications one might mildly reduce the acquisition time by appropriately reducing the imaging resolution, rather than the more complicated CS reconstruction, while maintaining diagnostic quality/accuracy. Use of synthetic images with artificial but realistic lesions permitted controlled evaluation of diagnostic accuracy for location-dependent compressed sensing reconstructions.

Reference

- [1] A. Macovski, "Noise in MRI," *Magn. Reson. Med.*, vol. 36, no. 3, pp. 494–497, 1996.
- [2] D. J. Larkman and R. G. Nunes, "Parallel magnetic resonance imaging," *Phys. Med. Biol.*, vol. 52, no. 7, p. R15, Apr. 2007.
- [3] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase, "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Magn. Reson. Med.*, vol. 47, no. 6, pp. 1202–1210, 2002.
- [4] D. K. Sodickson and W. J. Manning, "Simultaneous acquisition of spatial harmonics (SMASH): Fast imaging with radiofrequency coil arrays," *Magn. Reson. Med.*, vol. 38, no. 4, pp. 591–603, 1997.
- [5] A. A. Samsonov, E. G. Kholmovski, D. L. Parker, and C. R. Johnson, "POCSense: POCS-based reconstruction for sensitivity encoded magnetic resonance imaging," *Magn. Reson. Med.*, vol. 52, no. 6, pp. 1397–1406, 2004.
- [6] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, 2007.
- [7] I. Daubechies, M. DeFrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Commun. Pure Appl. Math.*, vol. 57, no. 11, pp. 1413–1457, 2004.

- [8] M. Lustig and J. M. Pauly, "SPIRiT: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space," *Magn. Reson. Med.*, vol. 64, no. 2, pp. 457–471, 2010.
- [9] N. M. Rofsky, "The importance of image quality: In the eyes of the beholder?," *J. Magn. Reson. Imaging*, vol. 41, no. 4, pp. 861–865, Apr. 2015.
- [10] N. Sinha and A. G. Ramakrishnan, "Quality assessment in magnetic resonance images," *Crit. Rev. Biomed. Eng.*, vol. 38, no. 2, pp. 127–141, 2010.
- [11] M. H. Kayvanrad, A. J. McLeod, J. S. H. Baxter, C. A. McKenzie, and T. M. Peters, "Stationary wavelet transform for under-sampled MRI reconstruction," *Magn. Reson. Imaging*, vol. 32, no. 10, pp. 1353–1364, Dec. 2014.
- [12] L. Birgé and P. Massart, "From Model Selection to Adaptive Estimation," in *Festschrift for Lucien Le Cam*, D. Pollard, E. Torgersen, and G. L. Yang, Eds. Springer New York, 1997, pp. 55–87.
- [13] K. O. McGraw and S. P. Wong, "Forming inferences about some intraclass correlation coefficients.," *Psychol. Methods*, vol. 1, no. 1, p. 30, 1996.
- [14] D. V. Cicchetti, "Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology.," *Psychol. Assess.*, vol. 6, no. 4, p. 284, 1994.
- [15] C. E. Metz, "Applications of Roc Analysis in Diagnostic Image Evaluation," Chicago Univ., IL (USA); Franklin McLean Memorial Research Inst., Chicago, IL (USA), CONF-790783-1, Jan. 1979.
- [16] S. D. Sharma, C. L. Fong, B. S. Tzung, M. Law, and K. S. Nayak, "Clinical Image Quality Assessment of Accelerated Magnetic Resonance Neuroimaging Using Compressed Sensing," *Invest. Radiol.*, Mar. 2013.
- [17] T. Zhang, S. Chowdhury, M. Lustig, R. A. Barth, M. T. Alley, T. Grafendorfer, P. D. Calderon, F. J. L. Robb, J. M. Pauly, and S. S. Vasanawala, "Clinical performance of contrast enhanced abdominal pediatric MRI with fast combined parallel imaging compressed sensing reconstruction," *J. Magn. Reson. Imaging*, vol. 40, no. 1, pp. 13–25, Jul. 2014.
- [18] S. S. Vasanawala, M. J. Murphy, M. T. Alley, P. Lai, K. Keutzer, J. M. Pauly, and M. Lustig, "Practical parallel imaging compressed sensing MRI: Summary of two years of experience in accelerating body MRI of pediatric patients," in *Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on*, 2011, pp. 1039 –1043.