Quantitative Testing of fMRI-compatibility of an Electrically Active Mechatronic Device for Robot-Assisted Sensorimotor Protocols

Andria J. Farrens *Student Member, IEEE*, Andrea Zonnino *Student Member, IEEE*, Andrew Erwin, *Student Member, IEEE*, Marcia K. O'Malley, *Senior Member, IEEE* Curtis L. Johnson, *Member, IEEE*, David Ress, *Senior Member, IEEE*, and Fabrizio Sergi, *Member, IEEE*,

Abstract— Objective: To develop a quantitative set of methods for testing the fMRI compatibility of an electrically-active mechatronic device developed to support sensorimotor protocols during fMRI. Methods: The set of methods includes phantom and in vivo experiments to measure the effect of a progressively broader set of noise sources potentially introduced by the device. Phantom experiments measure the radio-frequency (RF) noise and temporal noise-to-signal ratio (tNSR) introduced by the device. The in vivo experiment assesses the effect of the device on measured brain activation for a human subject performing a representative sensorimotor task. The proposed protocol was validated via experiments using a 3T MRI scanner operated under nominal conditions and with the inclusion of an electrically-active mechatronic device — the MR-SoftWrist — as the equipment under test (EUT). Results: Quantitative analysis of RF noise data allows detection of active RF noise sources both in controlled RF noise conditions, and in conditions resembling improper filtering of the EUT's electrical signals. In conditions where no RF noise was detectable, the presence and operation of the EUT did not introduce any significant increase in tNSR. A quantitative analysis conducted on in vivo measurements of the number of active voxels in visual and motor areas further showed no significant difference between EUT and baseline conditions. Conclusion and significance: The proposed set of quantitative methods supports the development and troubleshooting of electrically-active mechatronic devices for use in sensorimotor protocols with fMRI, and may be used for future testing of such devices.

Keywords – MR-compatible mechatronics; Signal-to-Noise Ratio; fMRI; Robotics; Sensorimotor learning; RF interference

I. INTRODUCTION

Functional Magnetic Resonance Imaging (fMRI) has been used extensively to study functional brain processes [1], [2], enabling localization of task-related brain activation [3], the analysis of intrinsic and/or task-related fluctuations [4], [5] and

causal coupling among brain signals [6] or, more recently, as a tool to provide on-line feedback on neural activity to improve task learning [7]. Tasks studied through fMRI include primary sensory functions, attention and recognition, word processing, and motor tasks [8].

Every brain process that is compatible with a subject laying with their head very still within a 60-70 cm diameter scanner bore can be studied through fMRI. However, only some motor tasks belong to this category. The space constraints of the MRI scanner, coupled with the stringent requirements for tolerable head movements and movement-generated scanner magnetic field distortion, make only very simplified motor protocols amenable to investigation through fMRI. Previously studied protocols include finger tapping [9], wrist pointing [10], grip force control [11], shoulder and elbow movement [12], and stepping [13].

In motor control, an advancement in the analysis of the neural control of movements has been provided by the combination of accurate motion tracking techniques with robotic devices capable of systematic kinesthetic or tactile feedback to assist, resist, augment, or perturb human movements. Such studies, coupled with computational models of motor control and motor learning, have led to an improved understanding of the criteria for motion planning and execution [14], adaptation to a new dynamic environment [15], modulation of mechanical impedance to stabilize unstable tasks [16], and processing of motor error information [17]. In addition to basic motor control research, the analysis of motor learning during human-robot interaction is very important for robot-aided neurorehabilitation [18], performance augmentation [19], and surgery [20].

In the pursuit of combining accurate kinesthetic feedback during sensorimotor protocols with simultaneous observation of brain activity via fMRI, researchers have endeavored to build MR-compatible haptic devices to support unconstrained hand movements [12], [21]–[23], hand/finger grasp [7], [24], [25], and foot movements [13], [26]. Such devices need to adhere to strict requirements for MR-compatibility, such as *i)* avoiding the use of magnetic materials to prevent distortion of the static magnetic field, *ii)* minimizing electromagnetic signals in both the radio-frequency (RF) and audible range to avoid interference with the RF-transceiver and gradients subsystems used in MRI, and *iii)* limiting use of in-bore electrical conductors to avoid coupling of eddy currents and
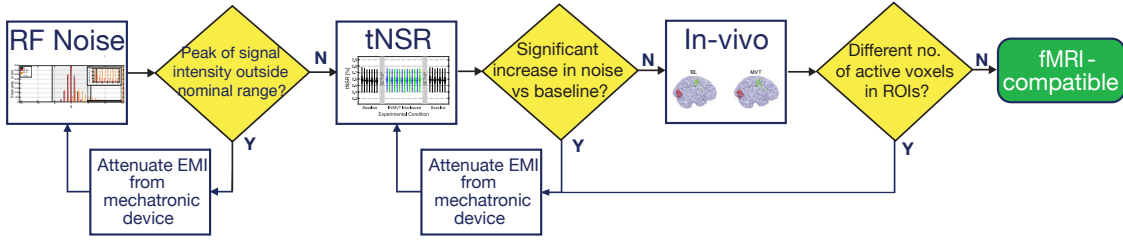
Fig. 1. Overview of the developed procedure for testing the fMRI compatibility of a mechatronic device, the MR-SoftWrist. The schematic emphasizes the role of the three sequential tests described in this paper, and of the decision blocks based on comparisons with "normal" signal and noise values, measured in the same scanner/coil configuration, and in a baseline condition which does include the mechatronic device under testing.

consequent bi-directional compatibility issues, as extensively discussed in [27].

MR-compatibility of a mechatronic device can be demonstrated experimentally, by measuring noise levels in different conditions of operation of the device. Experimental validation of the MR-compatibility of mechatronic devices has been performed by adapting techniques used for diagnostic scan purposes, and extensively used in surgical robotics [28], [29]. Standard tests for MR-compatibility of mechatronic devices quantify possible degradations in *image quality* introduced by the devices, such as the image signal-to-noise ratio, geometric distortion, slice thickness, position accuracy, and image uniformity, and use visual inspection to check for interference patterns (e.g., corduroy artifacts) [30], [31]. All these tests are designed to check for the different types of image degradation that can be introduced by magnetic materials, eddy currents in conductive materials, or electromagnetic interference in the radio-frequency or audible range. Such tests are routinely used for MRI image quality assurance; a device is qualified as MR-compatible if there is no detectable degradation in any measure of image quality.

However, when using mechatronic devices for fMRI, it is necessary to consider also the *temporal stability* of the measured signal, which could be affected by dynamic noise sources that do not necessarily result in image degradation. Since fMRI experiments measure signal changes on the order of 1%, it is important to ascertain that the temporal fluctuations of the fMRI signal are not altered by the presence of an electrically active mechatronic device. This aspect is usually quantified by the temporal signal-to-noise ratio (tSNR) [32]. Routine fMRI image quality assurance protocols address this aspect by analyzing the temporal structure of noise [32]–[35], but are not appropriate for conducting statistical inference to determine the effect of a given experimental condition. In the MR-compatible robotics literature, tSNR analysis has been previously performed in either a representative subset of the entire scanning volume [13], [27], or in repeated measurements in multiple slices [36]. However, previous work did not provide a method to establish fMRI-compatibility based on rigorous inference testing, capable of controlling for the false-positive rate. Moreover, a limitation of the tSNR measure is that it is an aggregate measure of noise, which makes it hard to understand which aspect of the device is causing interference problems in case of a positive result of a tSNR test.

In this paper, we present a set of methods that evolved during the development and testing of an fMRI-compatible mechatronic device, the MR-SoftWrist, designed to study wrist pointing under force feedback during fMRI. In distinction from longitudinal and multi-center quality assurance protocols developed for fMRI [32]–[35], the described approach is composed of a series of tests that gradually increase in experimental complexity to measure an increasing range of noise sources potentially introduced by the device (see Fig. 1). As such, the pursued approach is practically useful for identifying and troubleshooting potential sources of noise and interference caused by electrically-active mechatronic devices operating in MRI scanners. Moreover, we present a quantitative analysis on the set of developed metrics, which makes it possible to apply rigorous statistical inference to test whether the introduction of a mechatronic device has a significant effect on image quality and fMRI statistical parametric maps.

The developed set of methods includes three tests: RF noise, temporal noise-to-signal ratio (tNSR) and an *in vivo* validation. These metrics are ordered by decreasing specificity to noise sources, and by increasing experimental complexity and duration. The first test, RF noise, is a test for RF-interference (RFI) in the useful bandwidth around the nuclear magnetic resonance frequency, which tests whether the Equipment Under Test (EUT) emits any RFI, and whether its use couples RFI into the shield enclosure of the MR scanner. The second test, tNSR, is an aggregate measure of noise effects such as RFI and gradient coupling to conductors in the EUT. Finally, the third test is an *in vivo* measurement of fMRI contrast during a simple blocked sensorimotor task, conceived to provide a final validation of the compatibility of the device for a sensorimotor protocol.
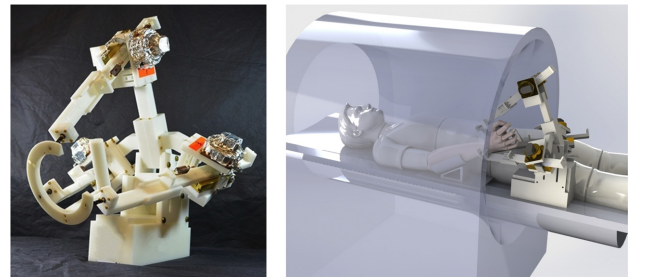

Fig. 2. Left: Picture of the MR-SoftWrist and Right: 3D CAD rendering showing its operational location during robot-assisted wrist pointing movements during fMRI.

II. MATERIALS

We conducted experiments using a 60 channel head coil on the Siemens Prisma 3T scanner at the University of Delaware, Center for Biomedical and Brain Imaging. The RF Noise and tNSR tests were conducted using a Siemens D165 spherical phantom. The phantom, composed of distilled H_2O doped with 1.25g $NiSO_4x6H_2O$ per 1kg H_2O, had relaxation properties that resemble those of the cerebral cortex. The *in vivo* experiment was conducted on a healthy right-handed subject, and was regulated by the University of Delaware Institutional Review Board (protocol no. 906215).

The MR-SoftWrist, shown in Fig. 2, is a robot developed to study wrist pointing movements under force-feedback during fMRI [37], [38]. The robot, used in this study as the EUT, includes several electrically active components: three rotary piezoelectric ultrasonic motors (Shinsei Corp. USR60-E3NT, powered by voltages on the order of 110 Vrms at 40-45 kHz), and six optical encoders. Through a parallel manipulator design, the MR-SoftWrist supports wrist flexion-extension and radial-ulnar deviation movements while locating its active components outside the scanner bore, at a distance (along the scanner z-axis) slightly longer than 1 m from the scanner isocenter for a subject of standard height. Tripolar twisted-pair cables, with an additional outer shield, were used for encoder lines. The cable shield, as well as metallic components, were connected to the electrical ground provided by the scanner penetration panel to the control room. To ensure decoupling of the signal references and to avoid bidirectional noise issues, encoder and motor power lines were low-pass filtered using 5600 pF and 1300 pF capacitive filters, respectively, with the filter frames connected to the penetration panel. To artificially introduce RFI with the EUT for one of the experimental conditions reported below, the encoder lines were not low-pass filtered, but connected through the penetration panel.

To generate controllable RF noise in the scanner, we used a custom Voltage Controlled Oscillator (VCO), with a variable output frequency in the range of 70-200 MHz (controlled by a voltage input), and +10 dBm output power. To generate controllable RFI, the VCO output was attenuated by 50 Ω RF attenuators (AIM-Cambridge RF, 27-9300-20), cascaded to provide a desired attenuation level down to −80 dB. The generated signal was a sinusoid with 3 dB bandwidth below 100 kHz. The VCO output was connected to the scanner room through a BNC cable input on the penetration panel on the control room side. A 40 cm antenna (gain: 2.15 dBi, impedance: 50 Ω) was connected to the VCO via a BNC connector on the scanner room side. During MRI operation, we observed fluctuations in the frequency of the generated signal up to ± 100 kHz. This setup allowed generation of controllable sources of RFI at frequencies close to the MR resonant frequency for the 3T Siemens scanner, and assessment of the ability of the developed techniques to detect the presence of RFI.

III. RF NOISE TEST

The RF noise test is a standard quality assurance test provided as a utility in commercial MRI scanners, used here to check whether the presence or operation of the EUT has an effect on the RF noise level in the scanner room, quantified directly by the MRI receiver. The RF noise test uses the MRI receiver as a spectrometer to capture any RF noise present when no RF excitation is generated by the scanner transmitter; gradient systems are also quiescent. In this test, the receiver is set to record the power of RF signal detected by the RF head coil at its center frequency. The receiver coil's center frequency is gradually increased, with a step Δf, across the full intermediate-frequency band centered around the scanner's proton nuclear magnetic resonance frequency. During a 6-minute RF noise test, the receiver takes M repeated measures of the RF signal intensity at n linearly sampled frequencies. We will refer to $s_{i,j}$ as the ith measurement ($i = 1,...,M$) at the jth frequency ($j = 1,...,n$).In the conditions used for this work, the receiver acquired multiple ($M = 256$) signal intensities at $n = 12750$ linearly increasing frequencies centered around the 3 T resonant frequency f_{res} = 123.25MHz and spanning a range $f_{res} \pm 250$ kHz, with a spacing of Δf = 39.2 Hz.

The RF signal intensities $s_{i,j}$, measured in arbitrary digital units, are expected to have a Rayleigh distribution, because RF noise should be characterized as a random process with positive-definite values. However, due to the frequency-dependency of the RF coils used, this fit is only approximate (Fig. 3A), requiring careful analysis to extract peaks of RF signal intensity that would indicate narrowband RF noise. For each frequency sampled, the M measurements of RF signal intensity are averaged together to produce a spectrum of mean signal intensities $\bar{s}_j$. Analysis of $\bar{s}_j$ shows a dependency of the measured RF signal on frequency, which is influenced by the shape and electrical parameters of the specific coil used, resulting in a variable frequency response (Fig. 3B). For our system, the dependency was approximated by a second-order polynomial model $\hat{s}_j = s_0 + s_1 f_j + s_2 f_j^2$. When this quadratic trend is removed, the detrended average signal intensities, $\check{s}_j = \bar{s}_j - \hat{s}_j$, have a symmetric distribution (Fig. 3C). The standard deviation of the distribution of $\check{s}_j$ is a measure of the intrinsic noise levels of the scanner receiver and coil combination. However, the large number of samples in $\check{s}_j$ reduce the sensitivity of the distribution parameters (mean and standard deviation) to narrowband RF noise sources, limiting their efficacy in identifying noise introduced by an electrically active mechatronic device. Additionally, it is not possible to assume that measurements of $\check{s}_j$ are all independent. This complicates the definition of a threshold of detrended signal intensity values that separates a region of normal variability of RF signal measurement from a region of "excessive" RF signal, such as RF noise introduced by an EUT, while controlling for the false positive rate of such a decision. To define such a threshold, we conducted repeated measurements in a baseline condition to empirically derive a 95% confidence interval for the measurement of peak $\check{s}_j$ values that served to characterize "excessive" RF noise.

We validated the RF noise metric in three separate experiments. Experiment 0 established the distribution of the maximum detrended signal intensity of the 3T scanner in normal operating conditions, which controlled for the false positive rate of identification of RF noise introduced by an
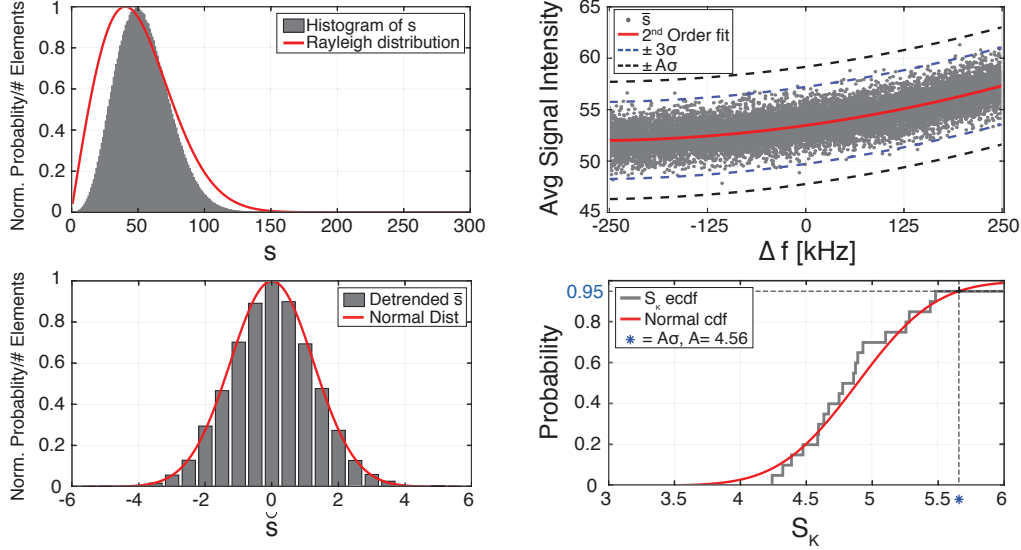
Fig. 3. Top Left: Distribution of signal intensities $s_{i,j}$ for the RF noise test, described by the intensity data histogram and fitted Rayleigh distribution (estimated distribution mode is 40.61 ± 0.02). Both functions are scaled to have unitary maximum. Top Right: Averaged signal intensity as a function of frequency for the RF noise test. The increasing trend is modeled by a second-order polynomial model, with overlaid parametric bounds corresponding to $p < 0.003$ uncorrected (region half width equal to three times the standard deviation of the detrended signal intensities), and the empirical 95% confidence interval ($\pm A\sigma$). No intensity value is outside the 95% confidence interval range. Bottom Left: Distribution of detrended signal intensities $\check{s}$. The bar plot reports the measured data, plotted as histogram of bins with unitary width, while the red line reports the probability density function of the fitted normal distribution (estimated parameters $\mu = 0$, $\sigma = 1.25$). Bottom Right: Empirical cumulative distribution function of maximum signal intensity values (S_k) for 20 repeated baseline experiments, with the cumulative distribution function of the normal distribution fitting S_k ($\mu = 4.9$, $\sigma = 0.5$) overlaid in red.

EUT. Experiment 1 established the validity of the test by demonstrating its capability to detect RF noise peaks in two cases where they were artificially introduced, using our voltage controlled RF signal generator. Experiment 2 established the sensitivity of the test with respect to conditions of operation of a mechatronic device, by quantifying the RF noise measured in different conditions of signal filtering for the electrically active components of the EUT.

A. Experiment 0 – Characterization of nominal RF noise

We conducted experiment 0 to characterize the nominal range of RF signal measured by the scanner, and to calculate a threshold of RF noise that controls for false positive rates when testing for RFI introduced by the EUT. Instances of RFI generated by an EUT can only increase the RF signal intensity recorded by the receiver and are thus captured as increases in the mean signal intensity. Characterization of the maximal signal intensity under baseline conditions allows for identification of instances of RFI introduced by the EUT that would fall above this threshold. Such RFI analyses therefore assume positive-definite noise statistics.

1) Methods: To determine a threshold of maximal RF noise in baseline conditions, we conducted a quantitative analysis of RF noise levels measured in 20 repeated scans performed at baseline (BL) conditions. The maximum detrended signal intensity was calculated for each experiment k, producing the set $S_k = max(\check{s}_j)$, $k = 1,...,20$. Given the normality of the measurements S_k (a Kolmogrov-Smirnov test failed to reject the null hypothesis that they are normally distributed with $p=0.60$), we fit a normal distribution to the measured data. Based on the estimated normal distribution, we calculated

a threshold of nominal peak signal intensity corresponding to a 5% false positive rate using the cumulative distribution function of the normal distribution. We used this threshold to create a 95% confidence interval for baseline RF signal intensity, which we then applied to distributions of detrended signal intensity, $\check{s}_j$.

2) Results: The fitted normal distribution $\mathcal{N}_{\mu,\sigma}$ had parameters $\mu = 4.9$ for the mean, and $\sigma = 0.5$ for the standard deviation (Fig. 3D). We determined the threshold of detrended signal intensity corresponding to a 5% false positive rate by finding the value of X such that $P(S_k \leq X) = 0.95$, which yielded the threshold of detrended signal intensity $X = 5.66$ (Fig. 3D). To calculate the coefficient A, such that $\pm A\sigma$ defines the 95% confidence interval of maximum nominal signal intensity for any RF distribution measured on the scanner, we divided X by the average standard deviation of $\check{s}_j$, resulting in $A = 4.56$. The coefficient A was used in all following RF noise tests to define the threshold at which values of greater signal intensity could be deemed to be introduced by an outside noise source with 5% chance of a false positive result.

A graphical example of the confidence interval used for RF noise analysis is shown in Fig. 3B, where the averaged signal intensities are plotted as a function of frequency, with the second-order polynomial regression line overlaid, and the $\pm 3\sigma$ and the $\pm A\sigma$ 95% confidence intervals. The presence of one voxel outside the 95% confidence interval would imply the presence of a peak of signal intensity that is unlikely to occur by chance, thereby demonstrating an RF noise problem at that specific frequency. The establishment of the value of coefficient A that achieves a false positive rate (FPR) of 5% allows for sensitive detection of narrowband RF noise, as
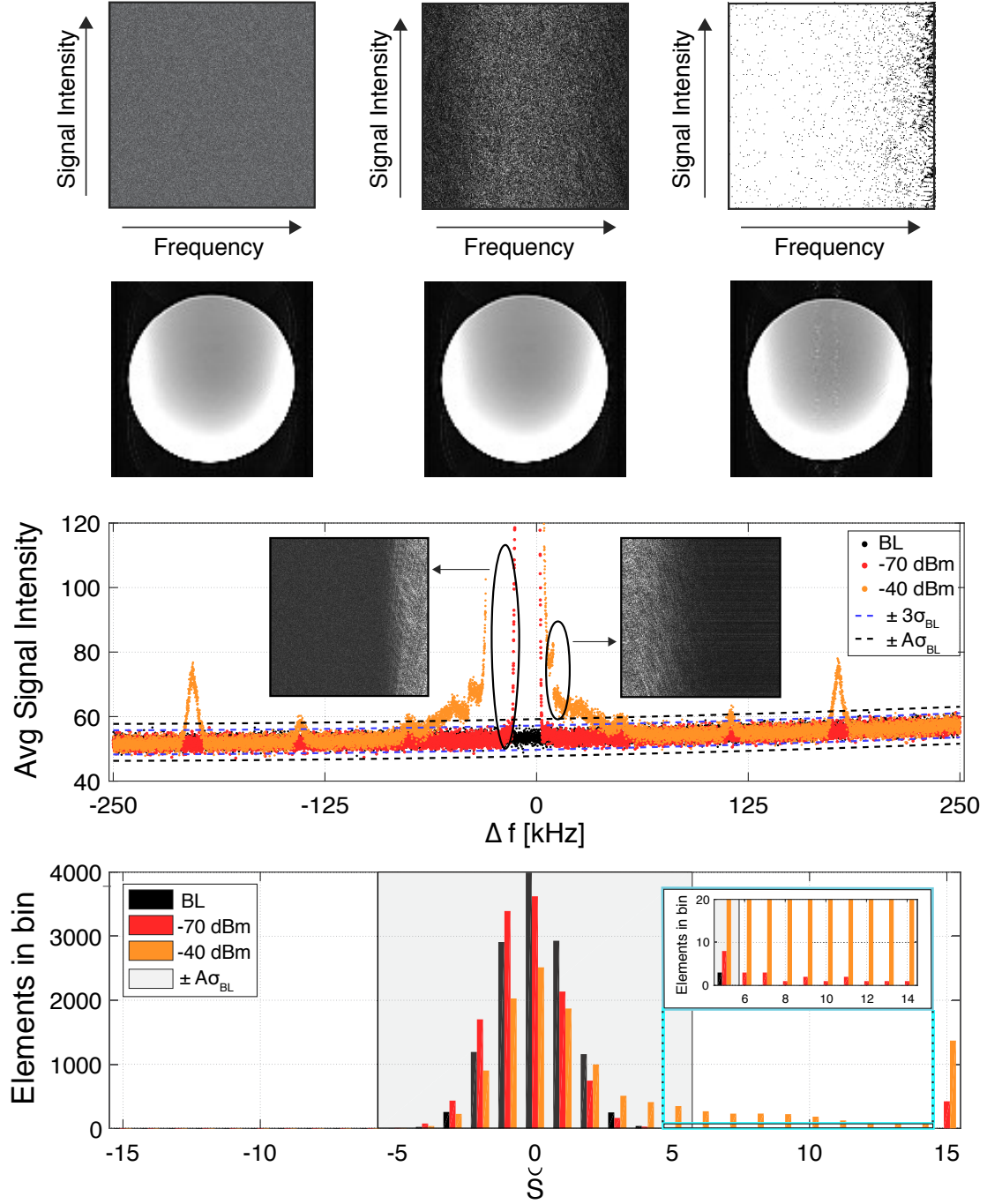
Fig. 4. RF noise images (first row) and corresponding EPI scans (second row) in the conditions studied in Experiment 1 (left column: baseline condition; center column: -70 dBm condition; right column: -40 dBm condition). First row: In the RF noise images, columns represent signal intensities measured at different frequencies, while rows represent repeated measurements of signal intensity at a given frequency. Second row: Below each RF noise images is a slice of the EPI volume for the spherical phantom imaged in the same experimental condition. The presence of RF interference in the phantom image is visible through white lines as are shown in the right phantom image. Third row: The spectrum of the averaged signal intensity over frequency shows the points of RF intensity values outside the $\pm A\sigma_{BL}$ for the RFI conditions. Black ellipses highlight representative peaks in the RFI conditions, with corresponding RF noise images. The central peaks of signal intensity for both RFI conditions are out of range. Bottom row: A histogram of detrended signal intensities, $\check{s}_j$. The insert shows the number of intensity values detected outside the $\pm A\sigma_{BL}$ in the -70 dBm and -40 dBm condition.

detailed in the following experiments 1 and 2.

B. Experiment 1 – Validation of the RF noise test

We conducted experiment 1 to establish validity of the RF noise test, i.e. its ability to detect the presence of RFI in conditions where it was purposely introduced by the Voltage Controlled Oscillator (VCO) described in section II.

1) Methods: We conducted an analysis of RF noise levels in the following experimental conditions:
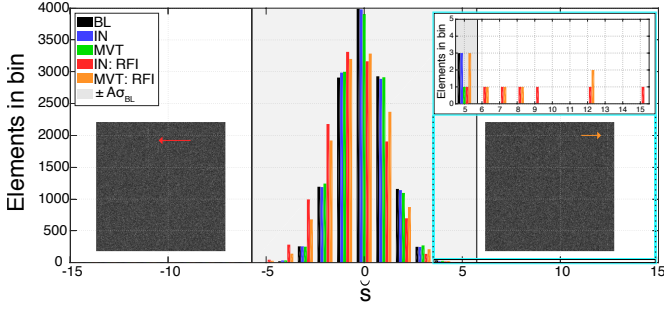
1) Baseline (BL), without the robot;

Fig. 5. Distributions of the detrended signal intensities ($\check{s}$) of RF noise for all conditions tested in Experiment 2. None of the experimental conditions involving the use of the MR-SoftWrist introduce significant noise at RF frequencies when the signal lines are properly shielded and grounded, as shown from the absence of peaks outside the $\pm A\sigma_{BL}$ range. In the two conditions where the manipulator signal cables were incorrectly shielded, peaks of signal intensity outside this range were measured at specific frequencies (shown in RF noise image inserts).

C. Experiment 2 – Effect of the EUT on RF noise

We conducted an additional experiment to establish sensitivity of the RF noise test with respect to operating conditions of an electrically active mechatronic device.

1) Methods: We conducted the RF noise test in the following experimental conditions:

1) BL, i.e. without the robot;
2) IN, with the EUT in its operational condition, powered on and properly connected for operation;
3) MVT, with the EUT position controlled to move its end effector through a sinusoidal trajectory (peak-to-peak amplitude: 1 cm, frequency: 0.5 Hz);
4) IN-RFI, with the EUT in the same position as in condition 2), but with the encoder lines directly connected to the penetration panel - i.e. without using capacitive low-pass filters on the EUT's encoder lines;
5) MVT-RFI, with the EUT performing the same movement as in condition 3), but with the encoder lines directly connected to the penetration panel.

As done in Experiment 1, we proceeded to calculate the detrended signal intensity spectra in conditions 2) – 5), and assessed whether peaks of detrended signal intensities lie within or outside the 95% confidence interval defined by the BL condition, using the A coefficient defined in Sec. IIIA.2.

2) Results: Based on the previously determined 95% confidence interval, none of the conditions in which the robot was properly connected and operated (IN and MVT) show elements outside the nominal range (Fig. 5). Instead, in the two conditions with improper signal filtering (i.e. conditions IN-RFI and MVT-RFI), there are several elements with intensity values outside the nominal range, suggesting that there might be elements inducing abnormal RF noise at those specific frequencies. The identification of these abnormal peaks of RF signal intensity demonstrate a high sensitivity of this analysis, as it detects patterns of RF noise hardly visible in the RF noise images (see inserts in Fig. 5).

This analysis demonstrates that there are configurations for the MR-SoftWrist that do not result in significantly different distributions of signal intensities relative to the BL condition. These configurations are considered not to have introduced corrupting RF noise, and, as such, are used for further experimental characterization and validation with the tNSR and in vivo experiments.

2) RFI -70 dBm (-70 dBm), where artificial RFI at 70 dBm (100 pW) was generated by connecting the VCO (output amplitude +10 dBm) to the antenna through an 80 dBm attenuator network. The VCO was set to generate a signal with center frequency within the scanner receiver bandwidth (123.5 ± 0.25 MHz).
3) RFI -40 dBm (-40 dBm), where artificial RFI at 40 dBm (100 nW) was generated by connecting the VCO to the antenna through a 50 dBm attenuator network. The VCO was set to generate a signal with center frequency within the scanner receiver bandwidth (123.5 ± 0.25 MHz)

We calculated the detrended signal intensity spectra for conditions 2) and 3), to assess whether peaks of detrended signal intensities fell within or outside the 95% confidence interval defined by the distribution of the BL condition, using the A coefficient defined in Sec. IIIA.2, and the standard deviation σ_{BL} measured in the BL condition.

2) Results: Visual inspection of the averaged signal intensity spectrum reveals a central peak of high intensity RF noise in the -70 dBm condition, and multiple peaks in the -40 dBm condition (Fig. 4). The histogram of detrended signal intensities shows that the previous determination of the 95% confidence interval for "nominal" maximum detrended signal intensity avoids any false positive results for the baseline condition, while still providing sensitivity to detect RFI introduced in both the -70 dBm and the -40 dBm RFI conditions.

Additionally, these results demonstrate the greater sensitivity of the RF noise metric in detecting noisy conditions compared to visual inspection of EPI images, which are shown in the second row of Fig. 4. In fact, from qualitative analysis of the phantom images, we can see that while the RF noise test captured RF interference for both -70 dBm and -40 dBm input power to the antenna, no visibly detectable image corruption could be observed in EPI images in the -70 dBm condition. Consequently, the RF noise metric can be seen to provide a strict check on the presence of RF noise sources.

IV. TEMPORAL NOISE-TO-SIGNAL RATIO (tNSR)

We calculated tNSR using procedures similar to those previously described in [32], with a phantom providing temporally and spatially constant signal when imaged using standard fMRI acquisition sequences. To extract the tNSR, a Region Of Interest (ROI) is defined near the center of the phantom, and signal intensity is measured from each voxel i in the ROI and concatenated in a time series s_i, with mean signal intensity $\bar{s}_i$. To increase specificity to rapid signal fluctuations, such as those generated by intermittent noise sources, the time-series s_i is detrended with a second-order polynomial to obtain the timeseries $\check{s}_i$. Noise n_i is calculated as the standard deviation

of the detrended signal intensity $\check{s}_i$. Finally, tNSR is calculated in percentage terms as:

$$\text{tNSR}_i = \frac{n_i}{\bar{s}_i} \times 100. \qquad (1)$$

Via analysis of tNSR values, we quantify the possible relation between an experimental condition (i.e. the presence or operation of a mechatronic device inside the scanner) and the increase in fluctuations of the fMRI signal compared to a baseline condition. Low tNSR values indicate that the signal is temporally stable in the absence of task-related changes in brain activity. A low tNSR thus enables detection of signal changes that arise in task-related fMRI protocols with full statistical power. This is a crucial aspect of fMRI since the effects of interest measured through increase in blood oxygen level-dependent (BOLD) signal are within 1% in magnitude [2].

Previous uses of tNSR (or its reciprocal, tSNR) in the field of MR-compatible mechatronics involved inference testing to compare the distributions of tNSR in two or more experimental conditions by using an ANOVA or a Kruskal-Wallis test. Use of these statistical tests to establish confidence intervals and determine *significant* change in tNSR due to a given experimental condition assumes the independence of the measured data (i.e. tNSR values of different voxels). Unfortunately, it is well known that such an assumption is strongly violated in MRI and fMRI [32], [34], due to spatial and temporal correlations in the signal and noise levels measured in different voxels. Moreover, it is unclear how many voxels should be considered in the analysis of tNSR values, with previous studies using from 49 voxels [13], to 5376 voxels [37], or simply implementing slice-wise comparisons [35], [36].

We have conducted two experiments to develop our quantitative approach in determining fMRI-compatibility of mechatronic devices based on tNSR measurements; one (Experiment 3) that aimed to evaluate the validity of whole-volume inference analysis based on tNSR data, and one (Experiment 4) that aimed to quantify the effect of the EUT on tNSR levels.

A. Experiment 3 – Volume inference based on tNSR values

We conducted Experiment 3 to test the effect of number of voxels in the analysis of whole-volume tNSR analysis.

1) Methods: We conducted seven repeated scans to estimate the mean tNSR in a spherical phantom (see Sec. II), with the scanner in normal operating condition. Imaging parameters included: single-shot Echo Planar Imaging (EPI); 2.04x2.04x3.00 mm voxel size; 34 slices with 0.75mm gaps; image size: 94x94 px; 192x192x102 mm^3 total imaging volume; 90 degree flip angle, TE=30 ms, TR=2030 ms, bandwidth per pixel=2315 Hz/pixel, receiver gain: high. For each scan, $i = 1, ..., 7$, we selected a cuboid ROI in the center of the phantom including $V = 18 \times 22 \times 22 = 8712$ voxels, and calculated the tNSR for all voxels within the ROI. The resulting tNSR values were ordered by the voxel x, then y, then z coordinates, from which subsets of tNSR data were obtained through spatial (re)sampling to produce the sets $A_{i,\Delta}$:

$$A_{i,\Delta} = \{tNSR_1, tNSR_{1+\Delta}, tNSR_{1+2\Delta}, ...\}, \qquad (2)$$

where i is the scan repetition and Δ is the spatial sampling period. Sets of $A_{i,\Delta}$ were created using the sampling periods $\Delta = [1, 2, 3, 5, 10, 50, 100, 200]$ voxels, which correspond to the inclusion of tNSR values from N=[8712, 4356, 2904, 1742, 871, 174, 87, 43] voxels. 95% confidence intervals for the mean tNSR were estimated for each set as $\overline{tNSR_i} \pm t_{0.05,N-1}\sigma_i/\sqrt{N}$, with $\overline{tNSR_i}$ as the sample mean and σ_i the standard deviation of the sample.

As we assume that the scanner is ideally stable during the seven repeated experiments, any pairwise comparison between the means estimated in two sets with the same Δ (i.e. between $A_{1,\Delta}$ and $A_{2,\Delta}$), which provides a statistically significant difference at the 95% level is to be considered a false positive. A FPR can be defined as the number of comparisons that produce a statistically significant difference, divided by the total number of possible pairwise comparisons, $c = \binom{7}{2} = 21$. In ideal, independent conditions, the FPR should be close to 0.05 for multiple comparisons of distributions with no significant difference between repeated runs.

2) Results: Analysis of the sets of tNSR data showed that the measured FPR depends on the number of voxels, N, used to determine the 95% confidence intervals of mean tNSR (where $N \propto \frac{1}{\Delta}$). Pairwise comparisons between the confidence intervals generated from the sets of tNSR data identified several conditions in which the FPR is much higher than the desired 5%. Comparisons between sets with $\Delta < 10$ ($N > 871$) resulted in FPRs that were 28% or higher. In these conditions, the 95% confidence intervals are very narrow due to an excessive number of voxels considered in the whole-volume analysis, which violates the assumption of independence of samples due to the spatial proximity of the voxels included. Conversely, in cases where a lower number of voxels in the volume are considered (when $\Delta > 100$, $N < 87$), the FPR dropped to zero. In these cases the 95% confidence intervals were expanded to widths of 4% tNSR. To determine the sensitivity of these large confidence intervals in detecting meaningful changes of signal to noise (SNR) measured during an activation experiment, we used the parameters $\bar{s}_i$ and n_i calculated from the tNSR set, and the definition of SNR

$$SNR_{\text{act}} = \frac{\Delta S}{n_i} = \frac{\Delta S}{\bar{s}_i} \cdot \frac{\bar{s}_i}{n_i} \qquad (3)$$

with ΔS equal to the change in BOLD signal. Considering an optimistic 2% increase in BOLD signal due to neural activation, and using the $N = 43$ tNSR results, the SNR_{act} values at the extremes of the interval range between 4.25 and 3.77, or $4.01 \pm 6\%$. As a result, in the worst-possible condition considered in this example (i.e. including only 43 voxels), pairwise comparison is not capable of discriminating between experimental conditions that would result in a 6% drop of SNR_{act} during fMRI activation tests. In this case, it is likely that the power of the test to detect a true positive is reduced, leading to the false conclusion that a condition was not significantly altering tNSR data, when in fact this could have been the case.

B. Experiment 4 – Effect of the EUT on tNSR

Based on the observation that the FPR of inference tests applied to whole-volume tNSR measurements depends on arbitrary parameters such as the number of voxels, we sought to develop a paradigm to quantify the fMRI-compatibility of a mechatronic device based on tNSR values, such that inference testing can be performed with strict control on FPR and significance levels. In our new approach, we extract the mean tNSR for the whole volume ROI as measured in multiple repeated scans. We then use the non-parametric method of bootstrapping to create 95% confidence intervals of mean tNSR from the resulting mean tNSR values. By repeating the measurement multiple times for each experimental condition and using bootstrapping of the obtained measurements, we calculate the distribution of the mean tNSR in a given experimental condition and its confidence interval at a specific significance level without violating the statistical assumptions detailed in Experiment 3. We can then conduct statistical inference to test the effect of a given experimental condition on tNSR (i.e. the presence or movement of a mechatronic device during fMRI) by comparing the estimated mean tNSR confidence intervals.

1) Methods: In Experiment 4, we used the same phantom and EPI sequence used in Experiment 3 to acquire seven repeated scans for each of the following experimental conditions related to the status of the EUT:

1) BL1, first baseline condition without the EUT;
2) IN, with the EUT in its operational condition, powered on and properly connected for operation;
3) MVT, with the EUT position-controlled to move its end effector through a sinusoidal trajectory (peak-to-peak amplitude: 1 cm, frequency: 0.5 Hz);
4) BL2, a second baseline condition, after removing the EUT from the scanner, to evaluate the test-retest error of the procedure.

For each condition, we extracted the 18 central slices of the phantom and defined a 22x22 voxels square ROI to obtain a cubiod ROI spanning a volume of 136 cm^3, comprised of 8,712 voxels. We computed the tNSR volume mean, and resampled through bootstrapping ($n = 1 \cdot 10^4$) the 7 repetitions for each experimental mode to compute confidence intervals and perform subsequent pairwise hypothesis testing. Through pairwise comparisons between *i)* the ensemble of repeated baseline conditions and *ii)* the IN condition and MVT condition respectively, we sought to determine whether the distribution of tNSR values was significantly altered in the IN and MVT condition relative to baseline.

2) Experiment 4 - Results: Pairwise comparisons between both the IN and MVT conditions with the combined baseline conditions fail to reject the null hypothesis at the $p < 0.05$ level, demonstrating that the presence of the MR-SoftWrist does not have a significant effect on noise levels during fMRI. The distributions of tNSR values measured in the seven repetitions for each experimental mode are shown in Fig. 6. Additional validation of the method presented in Experiment 4 is detailed in the supplementary material.
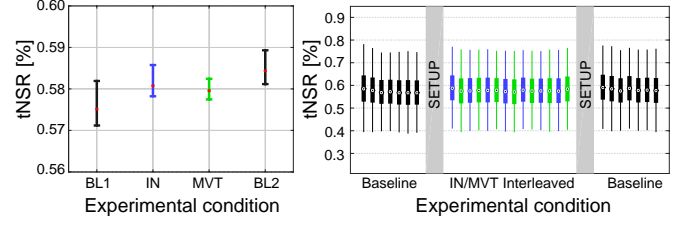

Fig. 6. Left: Bootstrapped confidence intervals for mean tNSR in the ROI volume for each condition at the $p < 0.05$ level, and Right: The distribution of tNSR for all repeated measures (seven for each experimental condition) ordered in terms of acquisition time.
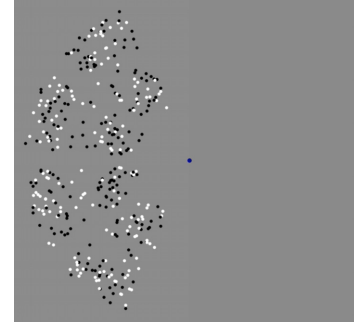

Fig. 7. Screenshot describing the visual stimulus, with black and white moving dots over a gray background, with the blue fixation dot in the center.

V. IN VIVO VALIDATION

The quality of statistical parametric maps obtained during an fMRI experiment is a function of both the noise intensities measured in the absence of activation, as well as the signal change measured in the transition between two or more experimental conditions. Although experiments in homogeneous phantoms do enable rigorous analysis of the spatial and temporal noise characteristics of the measured signal, tNSR alone does not account for the many noise and nuisance processes associated with the signal derived from task-related activation. Instead, the effect of any experimental condition on such a measure can be quantified by measuring concurrently the task-related signal change and its fluctuation, which can be done in a human experiment.

A. Experiment 5 – Effect of the EUT on task-related activations

We conducted a block design experiment to check whether the presence and operation of the EUT significantly affects the contrast of fMRI activation measured during a representative sensorimotor task.

1) Methods: We developed a visuomotor task, where a subject laying in a standard supine position in the MRI scanner is asked to visually fixate on a static dot in the center of a screen, while radially moving dots appear on the left side of the visual field (see Fig. 7). Dot direction changes repeatedly from centripetal to centrifugal during the task, and the subject is asked to tap their ipsilateral hand every time that a direction change is detected. The task alternates between active blocks, described above, and rest blocks, where the subject is asked to

fixate on a single center dot with a stationary background. The change of dot movement direction occurs every ΔT seconds, with ΔT values pre-computed from a bimodal distribution, obtained through superposition of two normal distributions with means at 0.5 s and 1 s and standard deviation of 0.25 s, and adjusted so as to result in a prescribed number of dot movement direction changes, and thus cued finger taps, for each constant-duration block. Through this pseudo-random stimulus, the effects of task learning or habituation are minimized, enabling comparisons between multiple runs of the same stimulus. By extracting the distributions of measured brain activation from the statistical parametric maps calculated for each repeated measurement, and comparing the distributions deriving from different experimental conditions (such as the operation of the EUT in the scanner room), it is possible to quantify the possible signal corruption introduced by a given experimental condition.

In this block design task, the subject was cued to execute 16 finger taps with his left hand in each of the four 12-second long active blocks, which separated five 12-second long rest blocks, for a total task duration of 114 s. The task was repeated seven times for each of the following experimental conditions:

1) BL1, a first baseline condition without the EUT;
2) MVT, with the EUT position controlled to move its end effector through a sinusoidal trajectory (peak-to-peak amplitude: 1 cm, frequency: 0.5 Hz). The robot was placed next to the subject in its normal operating position, but did not physically interact with the subject.
3) BL2, a second baseline condition, after removing the EUT from the scanner, to enable evaluation of the test-retest reliability of the procedure.

The total duration of the experiment was approximately 90 minutes, including 20 minutes for system setup and removal from the scanner room, and the seven scans per condition were executed consecutively.

Functional images were acquired using the same EPI fMRI sequence used in Experiment 3. The measured images were processed in the native fMRI space using only realignment to the first image of the BL2 session, using SPM12 realignment function, with options quality = 95%, separation 2.5 mm. A general linear model was constructed, using the block variable (stimulus on/off), convolved with SPM12 canonical hemodynamic response function, as a regressor. Model estimation yielded one t-map for each repeated run for the contrast (active-rest > 0). We defined a task-specific functional mask by averaging the activation maps from two random scans in each condition (i.e. two for BL1, two for MVT, two for BL2), and thresholding at the $p < 0.05$ uncorrected significance level. The resulting functional mask was restricted to anatomically defined visual and motor areas contralateral to the presentation of the visual stimulus and the hand performing the task. Anatomical masking was accomplished using the Juelich Anatomical Atlas, with thresholds set at 50%. The anatomic visual mask was defined as the union of contralateral primary and secondary visual cortices V1 and V2, and the motor mask was defined as the union of primary motor cortex BA4 anterior and posterior and premotor cortex BA6.

The resulting intersection between the task-specific functional mask and the anatomic visual and motor masks were used as the functionally-defined visual ROI ("Right V1") and motor ROI ("Right M1") in all subsequent analysis. The functionally-defined visual ROI contained 419 voxels and the motor ROI contained 204 voxels.

To estimate the effect of different experimental conditions on measured activation *at the ROI level*, we calculated the mean t-score in the functionally defined ROIs for each scan, and used bootstrapping ($n = 1 \cdot 10^4$) to estimate 95% confidence intervals for the mean t-score from the resulting sample (seven runs for BL1, MVT, and BL2). To estimate the effect of different experimental conditions on activation *at the voxel level*, we calculated the mean t-score for all voxels within the functionally-defined ROIs across the seven repeated measures taken in each experimental condition, and computed the difference in mean t-score distributions relative to BL1. One distribution (MVT−BL1) quantified the effect caused by the EUT, and the other (BL2−BL1) quantified the test-retest error of the task. To compare the effect of the EUT on voxel activation to the test-retest error of the measurement, we used a Kolmogorov-Smirnov test to test the null hypothesis that the two samples, MVT−BL1 and BL2−BL1, are taken from the same distribution.

2) Results: Analysis of the seven repeated runs of condition BL1, thresholded at the $p < 0.05$ uncorrected significance level, showed that the number of active voxels within the ROIs was (mean±standard deviation) 320±61 voxels (5.0±0.96 cm^3) in the visual ROI, and 155±31 voxels (2.4±0.49 cm^3) in the motor ROI. Robustness of activation, a feature depending only on the selected task, is determined by the relatively low variability of the number of active voxels.

Pairwise comparisons of the estimated 95% confidence intervals of mean t-scores show that no experimental condition results in statistically significant difference at the $p < 0.05$ significance level in either ROI (Fig. 8, top panels). Through this analysis, we demonstrate that the experimental condition (MVT) does not introduce significant degradation of activation contrast in visual or motor areas for the task considered.

A Kolmogorov-Smirnov test failed to reject the null hypothesis in the Visual ROI ($p = 0.43$), but not in the Motor ROI ($p = 0.003$). The resulting distributions are displayed in Fig. 8, bottom panels. Surprisingly, in the Motor ROI, the MVT−BL1 t-scores distribution was more closely clustered around zero than the BL2−BL1 distribution, suggesting that the variation in activation between the EUT and BL1 condition was no greater than the in test-retest error of the measurement (mean of MVT−BL1 t-scores distribution: 0.074, mean of the BL2−BL1 t-scores distribution: -0.279). The effect size of the difference in means between conditions is characterized as "small", with a Cohen's d of .25.

These results are qualitatively reflected in the similarity of the two activation maps presented in Fig. 9. The thresholded surface reconstructions show the average activation measured in the repeated runs of both BL conditions and in the MVT condition, together with the anatomical ROIs used to extract the distribution of region-specific statistical parameters.
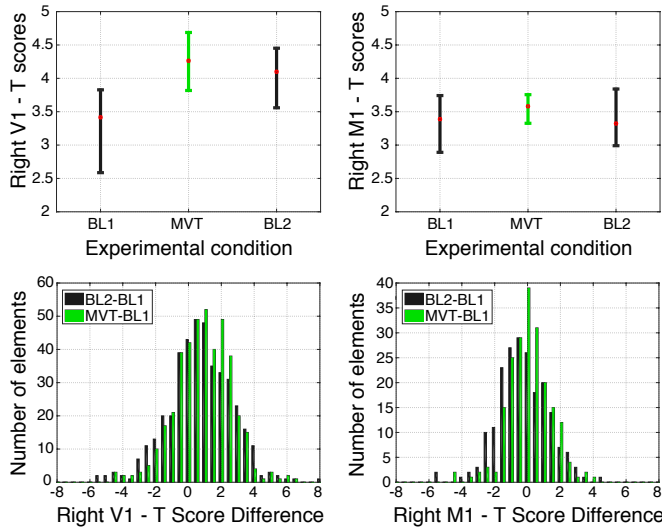
Fig. 8. Top Row: Distribution of the mean t-scores in the two functionally defined ROIs, represented by the distribution mean (red dot) and whiskers indicating the 95% confidence interval. Pairwise comparisons for both ROIs are not statistically significant at the $p < 0.05$ level, indicating no significant effect for the experimental condition. Bottom: Histogram of the between-condition change in t-score measured for each voxel in the ROI.
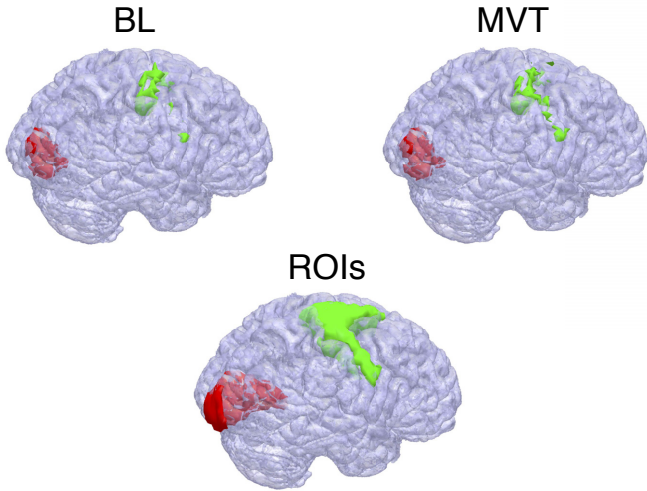

Fig. 9. Top: Thresholded ($p < 0.05$, uncorrected) activation regions for all BL conditions (left) and MVT (right), overlaid on the subjects cortical surface reconstruction. Bottom: ROIs describing the two anatomical Juelich masks used to compute the distribution of statistical parameters in the different experimental conditions.

VI. DISCUSSION

We presented a set of quantitative methods, which include experimental and analytical techniques used to quantify the fMRI-compatibility of a mechatronic device designed for operation in sensorimotor protocols during fMRI. Although several quantitative metrics have been proposed for characterizing scanner performance for fMRI studies [21], [32], [33], and quality assurance protocols have been developed for multicenter studies [32], [34], [35], [39], the presented methods have been conceived to support identification and troubleshooting of possible sources of noise and interference caused by mechatronic devices operating in MRI scanners. The developed approach utilizes experiments that are progressively less specific with respect to noise sources, and involve a gradually increasing level of experimental sophistication, and adherence to real-world application conditions. The developed set of methods can be conducted in parallel to standard MR quality assurance protocols that are required for anatomical MRI images, and focus on assessing image quality [30], [31].

Our approach, developed specifically for fMRI-compatibility analysis, combines three metrics and their related experimental procedures and analysis techniques. None of the metrics included in the presented approach are absolutely new in the field of MRI quality assurance; however, previous metrics were introduced without explicit reference to a numerical criterion to quantify the effect of the measured values on image or contrast degradation. With our work, we have provided a framework for performing statistical inference testing to determine if the introduction of a mechatronic device has a significant effect on image quality and fMRI statistical parametric maps. This aspect of our protocol is crucial in providing quantitative assessment of the devices being developed for use with fMRI and to understand their effect on experimental results. We discuss the significance of the presented methods and results below.

A. RF noise metric

The RF noise metric is a conservative test for RF noise present in the useful bandwidth of the scanner. For this metric, we have presented a quantitative method for automatic detection of RF interference, which results in a rapid test for a source of possible fMRI image degradation. This is an improvement over previous methods of RF noise detection, which were conducted by qualitative visual inspection of the RF noise spectra generated using standard quality assurance protocols. To the best of our knowledge, a qualitative analysis of RF noise results has been reported in only one case in the MR-compatible mechatronics literature [40]. Our quantitative method is particularly useful for troubleshooting possible sources of RF interference due to an EUT and its operation, both during the development phase and for regular quality assurance of the device. Our analysis demonstrated that the RF noise test, coupled with an empirically determined 95% confidence interval for the distribution of detrended signal intensities, resulted in a sensitive test for RF interference, as demonstrated via visual comparison with phantom images. Furthermore, while this metric tests specifically for RFI, we found that for the MR-SoftWrist, filtering and shielding configurations identified by the RF noise test to introduce negligible levels of RF interference also resulted in no significant difference in tNSR or in measurable activation for functional images. As such, we highlight the usefulness of the RF noise test as part of a device validation protocol, and –given the relatively short duration of an RF noise measurement (approximately 6-7 minutes) and that it only requires a phantom scan– we recommend using the RF noise test to confirm proper filtering and shielding of cabling connecting electrically-active elements in the imaging room.

B. tNSR

tNSR during phantom experiments is sensitive to reductions in signal quality potentially introduced by the presence or operation of an EUT in the scanner room. Our analysis demonstrated that an arbitrary choice, such as the number of voxels considered for whole-volume analysis of tNSR data, significantly affects the false positive rate of inference tests, and consequently may result in poorly controlled levels of false negatives when establishing the MR-compatibility of a mechatronic device. Previous papers reporting on the MR-compatibility of mechatronic devices (e.g. [13]) used a low number of voxels to test the tNSR changes introduced by the presence or movement of a mechatronic system in the scanner. Our analysis shows that some of those studies might have incorrectly concluded the MR-compatibility of the presented devices based on consideration of tNSR data in too small a subset of voxels.

To avoid the unwanted dependence of the result of inference tests on the selected number of voxels, we developed an alternative analytical and experimental protocol. We based our analysis on repeated measurements of the mean tNSR within a selected ROI in interleaved experimental conditions, and estimation of the confidence interval through bootstrapping to allow subsequent pairwise comparisons between experimental conditions. The developed protocol quantifies the test-retest error of the metric obtained in different repetitions of the baseline condition, for comparison with the tNSR change deriving from the introduction of the EUT. Application of this method with the MR-SoftWrist demonstrated no significant difference in tNSR, between either the IN or MVT condition and the BL conditions.

C. In vivo validation

In the validation experiment with a human subject, we did not observe significant changes in either spatial extent or amplitude of task-related activation in the presence of the EUT, compared to the baseline condition. This result is not surprising, since the in vivo experiment is less specific with respect to noise sources relative to phantom experiments, and the EUT did not show any significant effects in the phantom experiments. In previous experiments conducted using a different scanner and coil (see Supplementary Material), we measured a significant 1.8% increase of tNSR in conditions of operation of the EUT, while the in vivo validation experiment did not show any effects. The significance of the *in vivo* experiment is to provide an indication as to what extent increases of tNSR values – measured with high specificity in phantom experiments – are capable of inducing significant degradations in the statistical parametric maps, to the point that they modify the extracted activations in human subjects. In our previous study, we observed that the 1.8% increase of tNSR is not sufficient to modify the amplitude and spatial extent of sensorimotor task-related activations.

Testing mechatronic devices in conditions that closely resemble those of operation during experiments is crucial. Validation of mechatronic devices via human experiments should always be performed as an intermediate step in the device development process to prove consistency and applicability to the intended application. Measurement of contrast during a human experiment is useful because it involves realistic conditions observed during real experiments in fMRI.

D. Study limitations

Results obtained to demonstrate the MR-compatibility of a specific mechatronic device operating during fMRI are highly specific to a given scanner, coil, and sequence combination. In particular, the 95% confidence interval and some of the observations of the frequency dependency of the signal intensities measured in the RF noise test are highly coil-specific, with different relationships potentially measurable for coils with a different number of channels or electrical configurations. Similarly, temporal noise-to-signal values can change with the use of a different sequence and phantom composition. To deal with the issues above, the developed methodology is heavily based on comparisons with matching control conditions, where the effect of potentially confounding variables (i.e. scanner field, coil type, sequence, room temperature, weather conditions) is minimized by repeating measurements in the same conditions, and interleaving temporally experimental conditions to avoid thermal drifts, when possible. As an unavoidable consequence, the banner "fMRI-compatible" that we seek to attribute to a given mechatronic device based on the presented methodology can be conferred only with explicit reference to a very specific configuration, defined by a given scanner field, coil type, and sequence. As soon as one of parameters mentioned above are changed, a subset of experiments within those presented in this paper needs to be rerun to test fMRI-compatibility in the modified conditions.

VII. Conclusions

We have presented a new quantitative set of methods to support the development and troubleshooting of an electrically active mechatronic device developed for use in sensorimotor protocols with fMRI. With the developed methodology, we have incidentally demonstrated that careful introduction of non-magnetic metals and shielded electrically active components do not result in significant degradation of signal contrast for fMRI experiments, for the MR-SoftWrist, a wrist pointing robot meant to operate just outside the scanner bore. The developed methodology is useful for the objective assessment of the fMRI-compatibility of future mechatronic devices developed for use in fMRI protocols. The methods are sufficiently general that they could also be of use in the evaluation of virtually all electrically active equipment that are intended for use during fMRI.

References

[1] S. Ogawa *et al.*, "Brain magnetic resonance imaging with contrast dependent on blood oxygenation," *Proc. Natl. Acad. Sci*, vol. 87, pp. 9868–9872, Dec. 1990.

[2] D. J. Heeger and D. Ress, "What does fMRI tell us about neuronal activity?" *Nature Rev. Neurosci.*, vol. 3, no. 2, pp. 142–151, Jan. 2002.

[3] E. Amaro, Jr. and G. J. Barker, "Study design in fMRI: Basic principles," *Brain Cognition*, vol. 60, no. 3, pp. 220–232, Apr. 2006.

[4] B. Biswal *et al.*, "Functional connectivity in the motor cortex of resting human brain using echoplanar MRI," *Magnet. Reson. Med.*, vol. 34, no. 4, pp. 537–541, Oct. 1995.

[5] M. Hampson *et al.*, "Detection of functional connectivity using temporal correlations in MR images," *Hum. Brain Mapp.*, vol. 15, no. 4, pp. 247–262, Jan. 2002.

[6] K. J. Friston, L. Harrison, and W. Penny, "Dynamic causal modelling," *NeuroImage*, vol. 19, no. 4, pp. 1273–1302, Aug. 2003.

[7] J. Sulzer *et al.*, "Delineating the whole brain BOLD response to passive movement kinematics," in *Proc. IEEE Int Conf Rehabil Robot*, 2013, pp. 1–5.

[8] N. K. Logothetis, "What we can do and what we cannot do with fMRI," *Nature*, vol. 453, no. 7197, pp. 869–878, Jun. 2008.

[9] N. Sadato *et al.*, "Frequency-dependent changes of regional cerebral blood flow during finger movements: functional MRI compared to PET," *J. Cerebr. Blood F. Met.*, vol. 17, pp. 670–679, Jun. 1997.

[10] L. Shmuelof *et al.*, "The neural correlates of learned motor acuity," *J. Neurophysiol.*, vol. 112, no. 4, pp. 971–980, Aug. 2014.

[11] J. K. Buschbeck, H. H. Ehrsson, and H. Forssberg, "Human brain activity in the control of fine static precision grip forces: an fMRI study," *Eur. J. Neurosci.*, vol. 14, pp. 382–390, Jul. 2001.

[12] J. Diedrichsen *et al.*, "Neural correlates of reach errors," *J. Neurosci.*, vol. 25, no. 43, pp. 9919–9931, Sep. 2005.

[13] C. Hollnagel *et al.*, "Brain activity during stepping: A novel MRI-compatible device," *J. Neurosci. Meth.*, vol. 201, no. 1, pp. 124–130, Sep. 2011.

[14] T. Flash and N. Hogan, "The coordination of arm movements: an experimentally confirmed mathematical model," *J. Neurosci.*, vol. 5, no. 7, pp. 1688–1703, Jul. 1985.

[15] R. Shadmehr and F. A. Mussa-Ivaldi, "Adaptive representation of dynamics during learning of a motor task," *J. Neurosci.*, vol. 14, no. 5, pp. 3208–3224, May 1994.

[16] E. Burdet *et al.*, "The central nervous system stabilizes unstable dynamics by learning optimal impedance," *Nature*, vol. 414, no. 6862, pp. 446–449, Nov. 2001.

[17] D. J. Herzfeld *et al.*, "A memory of errors in sensorimotor learning," *Science*, Sep. 2014.

[18] H. Krebs *et al.*, "Robot-aided neurorehabilitation," *IEEE T. Rehabil. Eng.*, vol. 6, no. 1, pp. 75–87, Mar. 1998.

[19] H. Kazerooni, R. Steger, and L. Huang, "Hybrid Control of the Berkeley Lower Extremity Exoskeleton (BLEEX)," *Int. J. Robot. Res.*, vol. 25, no. 5-6, pp. 561–573, May 2006.

[20] A. M. Okamura, "Methods for haptic feedback in teleoperated robot assisted surgery," *Ind. Robot.*, vol. 31, no. 6, pp. 499–508, Dec. 2004.

[21] N. Yu *et al.*, "fMRI assessment of upper extremity related brain activation with an MRI-compatible manipulandum," *Int. J. Comput. Assist. Radiol. Surg.*, Aug. 2010.

[22] S. Menon *et al.*, "Haptic fMRI: Combining Functional Neuroimaging with Haptics for Studying the Brain's Motor Control Representation," in *Proc. IEEE Eng Med Biol Soc*, 2013, pp. 1–6.

[23] S. Menon, H. Ganti, and O. Khatib, *Using Haptic fMRI to Enable Interactive Motor Neuroimaging Experiments*. Springer International Publishing, 2016, pp. 89–103.

[24] A. Khanicheh *et al.*, "fMRI-compatible rehabilitation hand device," *J. Neuroeng. Rehabil.*, vol. 3, no. 1, p. 24, Oct. 2006.

[25] B. Vigaru, J. Sulzer, and R. Gassert, "Design and Evaluation of a Cable-Driven fMRI-Compatible Haptic Interface to Investigate Precision Grip Control," *IEEE Trans. Haptics*, pp. 1–1, 2016.

[26] T. Ikeda *et al.*, "Muscle activity during gait-like motion provided by mri compatible lower-extremity motion simulator," *Adv. Robotics*, vol. 30, no. 7, pp. 459–475, Feb. 2016.

[27] N. Yu, R. Gassert, and R. Riener, "Mutual interferences and design principles for mechatronic devices in magnetic resonance imaging," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 6, no. 4, pp. 473–488, Sep. 2010.

[28] G. S. Fischer *et al.*, "MRI compatibility of robot actuation techniques– a comparative study," *Med Image Comput Comput Assist Interv.*, vol. 11(Pt. 2), pp. 509–517, 2008.

[29] B. Yang *et al.*, "Design and implementation of a pneumatically-actuated robot for breast biopsy under continuous MRI," in *Proc. IEEE Int Conf Robot Autom.*, 2011, pp. 674–679.

[30] National Electrical Manufacturers Association, "NEMA Standards Publication MS 1-2008, *Determination of Signal-to-Noise Ratio (SNR) in Diagnostic Magnetic Resonance Imaging* ," 2008.

[31] J. Weinreb *et al.*, "ACR MRI Accreditation: Yesterday, Today, and Tomorrow," *J. Am. Coll. Radiol.*, vol. 2, no. 6, pp. 494–503, Jun. 2005.

[32] L. Friedman and G. H. Glover, "Report on a multicenter fMRI quality assurance protocol," *J. Magn. Reson. Imaging*, vol. 23, no. 6, pp. 827–839, Jun. 2006.

[33] R. M. Weisskoff, "Simple measurement of scanner stability for functional NMR imaging of activation in the brain," *Magn. Reson. Med.*, vol. 36, no. 4, pp. 643–645, Oct. 1996.

[34] T. Stöcker *et al.*, "Automated quality assurance routines for fMRI data applied to a multicenter study," *Hum. Brain Mapp.*, vol. 25, no. 2, pp. 237–246, Jun. 2005.

[35] B. P. Sutton *et al.*, "Investigation and validation of intersite fMRI studies using the same imaging hardware," *J. Magn. Reson. Imaging*, vol. 28, no. 1, pp. 21–28, Jul. 2008.

[36] N. Vanello *et al.*, "Sensing Glove for Brain Studies: Design and Assessment of Its Compatibility for fMRI With a Robust Test," *IEEE/ASME Trans. Mechatronics*, vol. 13, no. 3, pp. 345–354, 2015.

[37] A. Erwin *et al.*, "Development, control and MRI compatibility of the MR-SoftWrist," in *Proc. IEEE/RAS-EMBS Int Conf Rehabilitation Robotics*, 2015, pp. 187–192.

[38] ——, "Kinesthetic feedback during 2DOF wrist movements via a novel mr-compatible robot," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. PP, no. 99, pp. 1–12, Dec. 2016.

[39] G. H. Glover, L. Foland, and F. BIRN, "*Scanner quality assurance for longitudinal or Multicenter fMRI studies*," in *Proc. Intl. Soc. Mag. Reson. Med.*, 2004, pp. 1–1.

[40] J. Reithler *et al.*, "Resistance-based high resolution recording of predefined 2-dimensional pen trajectories in an fMRI setting," *J. Neurosci. Meth.*, vol. 152, no. 1, pp. 10–17, Apr. 2006.

SUPPLEMENTARY MATERIAL

A. tNSR Validation

To confirm the validity of our proposed method in detecting changes in tNSR (as detailed in Experiment 4), we applied our analysis to a series of tNSR experiments in which noise was intentionally introduced to the scanner environment by the VCO, and through the operating conditions of the EUT.

1) Experiment S1: tNSR in presence of RFI: The proposed tNSR method was used to estimate the change in mean tNSR for phantom images collected in baseline and RFI conditions— as detailed in Experiment 1. The image sequence and definition of the ROI volume were consistent with the methodology described in Experiment 3.

The 95% confidence intervals for all three conditions (BL, RFI -70 dBm, RFI -40 dBm) are shown in Fig.S1. In this experiment, the tNSR measure identified a significant increase in mean tNSR for the -40 dBm condition, in which slight artifacts could be detected by visual inspection of the phantom images. The test showed no significant difference between baseline and the -70 dBm condition, which were also indistinguishable via visual inspection. This experiment demonstrates both the validity and short-comings of the tNSR metric— it is capable of detecting changes in tNSR where it is known to exist (-70 dBm), but unable to detect instances of RF noise below a certain threshold (-40 dBm). These results highlight the usefulness of the RF noise metric in conducting a strict check for instances of RFI that may otherwise be undetectable.
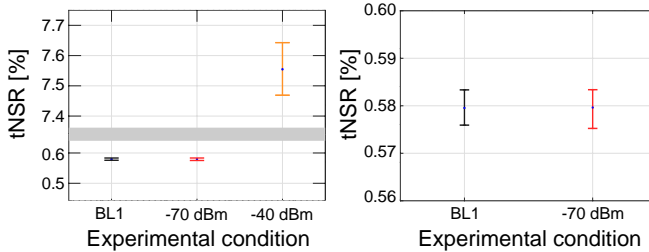

Fig. S1. Left: Bootstrapped confidence intervals for mean tNSR in the ROI volume for all 3 conditions at the $p < 0.05$ level. Right: Pairwise comparison between the 95% confidence intervals for BL and -70 dBm experimental conditions shows no significant difference.

2) Experiment S2: tNSR in presence of EUT: We applied our methodology to estimate the change in tNSR between baseline and EUT conditions.

Dsata for this analysis were collected in 2015 using different scanner hardware, but with the same EUT, the MR-Softwrist, as detailed previously (Sec II: Methods). The 32-channel head coil and Siemens Trio 3T scanner at the neuroimaging center, Core for Advanced MRI, at the Baylor College of Medicine were used to image a sherical phantom filled with agarose gel doped with a 0.2mM $Ni(NiO_3)_2$ and 0.5wt% NaCl. Imaging parameters included: single-shot Echo Planar Imaging (EPI); 2.5 mm isotropic voxel size; 38 slices with no gaps; image size: 80x80 px; 200x200x95 mm^3 total imaging volume; 78 degree flip angle, TE=35 ms, TR=2000 ms, bandwidth per pixel=1453 Hz/pixel.
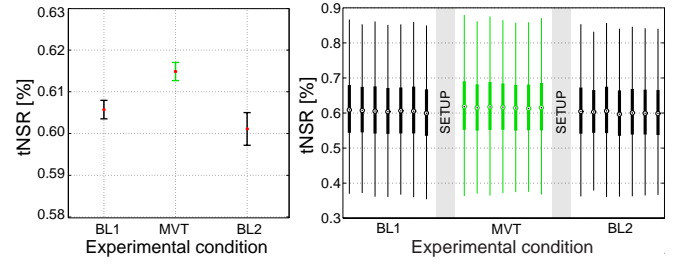

Fig. S2. tNSR results of a previous fMRI compatibility experiment conducted with the MR-Softwrist on a different MRI scanner. Left: Bootstrapped confidence intervals for mean tNSR in the ROI volume for each condition at the $p < 0.05$ level, and Right: The distribution of tNSR for all repeated measures (seven for each experimental condition) ordered in terms of acquisition time

In this experiment, we acquired seven repeated scans for each of the following experimental conditions related to the status of the EUT:

1) BL1, first baseline condition without the EUT;
2) MVT, with the EUT position-controlled to move its end effector through a sinusoidal trajectory (peak-to-peak amplitude: 1 cm, frequency: 0.5 Hz);
3) BL2, a second baseline condition, after removing the EUT from the scanner, to evaluate the test-retest reliability of the procedure.

Pairwise comparisons between the MVT and both baseline conditions are statistically significant at the p<0.05 level, demonstrating that the presence of the MR-SoftWrist in the experimental condition does indeed increase noise levels during fMRI (Fig. S2). The increase in tNSR measured in the active MVT condition is about 1.8% of the mean baseline value, which is greater than the estimated test-retest reliability of the measurement (estimated to be equal to 0.8%). This result demonstrates that the developed methodology is capable of quantifying an increase in tNSR introduced by the operation of an EUT in the scanner.

In the more recent experimental results presented in Sec. IV, we did not observe a significant increase in tNSR for experimental modes in which the EUT was included in the scanner environment. The discrepancy may be accounted for by the differences in the scanner hardware or imaging sequences used between the two experiments. These results highlight the specificity of the presented fMRI compatibility tests to the scanner and sequence under test, and the need to re-validate the fMRI compatibility of the EUT in the experimental conditions in which it is planned to operate.