Haptic fMRI : Using Classification to Quantify Task-Correlated Noise during Goal-Directed Reaching Motions

Samir Menon[1], Paul Quigley[1], Michelle Yu[1], and Oussama Khatib[1]

***Abstract*— Neuroimaging artifacts in haptic functional magnetic resonance imaging (Haptic fMRI) experiments have the potential to induce spurious fMRI activation where there is none, or to make neural activation measurements appear correlated across brain regions when they are actually not. Here, we demonstrate that performing three-dimensional goal-directed reaching motions while operating Haptic fMRI Interface (HFI) does not create confounding motion artifacts. To test for artifacts, we simultaneously scanned a subject's brain with a customized soft phantom placed a few centimeters away from the subject's left motor cortex. The phantom captured task-related motion and haptic noise, but did not contain associated neural activation measurements. We quantified the task-related information present in fMRI measurements taken from the brain and the phantom by using a linear max-margin classifier to predict whether raw time series data could differentiate between motion planning or reaching. fMRI measurements in the phantom were uninformative (2σ, 45–73%; chance=50%), while those in primary motor, visual, and somatosensory cortex accurately classified task-conditions (2σ, 90–96%). We also localized artifacts due to the haptic interface alone by scanning a stand-alone fBIRN phantom, while an operator performed haptic tasks outside the scanner's bore with the interface at the same location. The stand-alone phantom had lower temporal noise and had similar mean classification but a tighter distribution (bootstrap Gaussian fit) than the brain phantom. Our results suggest that any fMRI measurement artifacts for Haptic fMRI reaching experiments are dominated by actual neural responses.**

I. Introduction

Haptic fMRI experiments that involve large three-dimensional motions present a novel opportunity to study human motor planning and visuomotor integration, making it imperative to reject the possibility of obtaining spurious scientific results due to potential limb motion artifacts. Such artifacts may arise due to haptic interface electronics, magnetic field changes induced by limb displacement, or by systematic perturbations to scanner calibration during motor tasks. Given that most artifacts have immediate effect, it is feasible to segregate them from neural activation using an impulse response model because fMRI measures slow changes in blood deoxygenation [1], [2]—neural activation outlasts the transient artifacts. It is, however, a challenge to determine whether limb motions induce slow task-correlated magnetic field fluctuations that mirror actual neural activation [3], or test if artifacts can substantially disrupt non-parametric analyses.

Fig. 1. Motor Neuroimaging with Haptic fMRI. The experiment involved planning and executing goal-directed reaches to three distinct spatial locations (left, L; middle, M; right, R) that spanned the MRI workspace. A three degree-of-freedom fMRI-compatible haptic interface, HFI, monitored hand motions. A real-time virtual simulation provided visual feedback on a screen behind the scanner.

Past research has attempted to eliminate motor artifacts in fMRI measurements to begin with. Approaches included avoiding electromagnetic actuation for haptic interfaces [4], [5]; placing actuators outside the scanner room [6]; using simple devices and planar motions [7]; or increasing signal-to-noise by using low fMRI resolutions (voxels>27mm, sample time>2s) and smoothing data in post-processing [4], [8], [9]. Such approaches, however, usually lead to haptic interfaces that perform well only for a limited set of motor tasks. The alternative, general purpose three degree-of-freedom haptic interfaces with high-fidelity electromagnetic actuation [10], [11], are novel devices, and it is an active area of research to ensure that they reject artifacts in experiments involving natural [12] and unconstrained three degree-of-freedom motions [13].

Here, we demonstrate that it is feasible to conduct three degree-of-freedom reaching tasks (Fig. 1) with Haptic fMRI while avoiding spurious interpretations induced by artifacts. To do so, we used a max-margin classifier [14] to test how informative fMRI measurements were in the brain and in a phantom that was placed next to it during fMRI acquisition. Since task-correlated fMRI artifacts must appear in both brain and phantom, obtaining good classification using measurements from the phantom would imply that artifacts dominate neural activity. Moreover, we did not assume canonical

*This work was supported by a Stanford University BioX fellowship (S. Menon) and a Stanford University BioX Neuroventures Research Grant (O. Khatib)

[1]S. Menon, P. Quigley, M. Yu, and O. Khatib are with the Artificial Intelligence Laboratory, Department of Computer Science, Stanford University, Stanford, CA 94305, USA smenon@stanford.edu, piq93@stanford.edu, mxyu@stanford.edu, ok@cs.stanford.edu

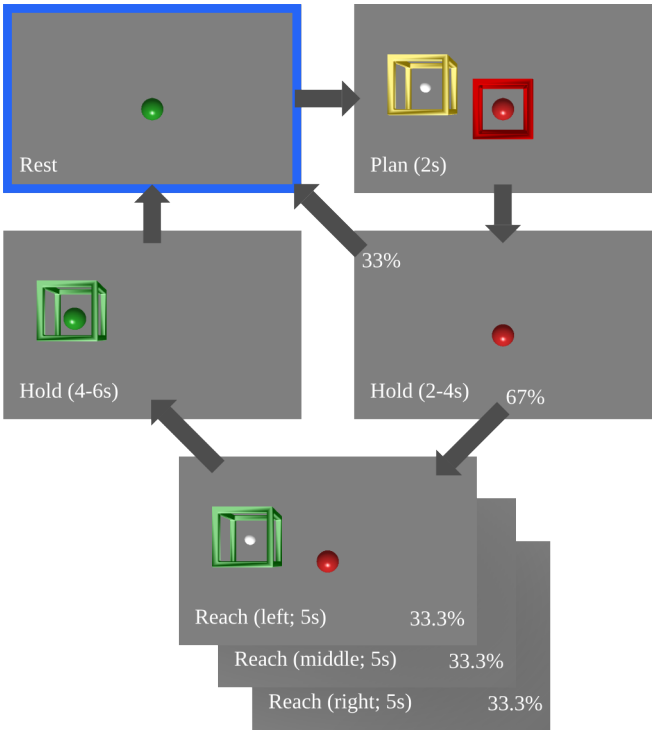

Fig. 2. A Motor Task Section. Task instructions were provided on a monitor visible from inside the scanner bore through a dual-mirror setup. Panels show different stages of the motor task. Subjects started at rest (blue highlight), planned a motion to one of three spatial positions (left, middle, or right) with equal probability, and then either executed a reach to the goal position or returned to rest. The reach was followed by a hold at the spatial location, after which subjects returned to the initial rest position. Subjects executed four such sections for each spatial position (twelve total) in an fMRI scan run, and executed 8–10 runs in a session (see Appendix for details).

haemodynamic responses, which are likely to filter artifact related noise. This increased the likelihood that our analysis would capture instantaneous motion artifacts. The opposite held true in our results. fMRI measurements in phantoms could not predict whether subject were planning motions or reaching, which validated our Haptic fMRI interface's (HFI) performance [11]. We attribute our results to a highly reliable experiment protocol [13] and low temporal noise during scanning.

II. Experiment Design

Our primary objective was to establish Haptic fMRI as an experimental technique that can enable motor neuroscientists to non-invasively study the human brain during a variety of motor tasks. Doing so can complement classical (invasive) electrophysiology research in animals by validating their results in humans. In addition, it is our goal to enable complex tasks that animals find challenging to perform, but which humans can quickly learn and perform—overtraining can alter the subject's neural motor representation. Such motor neuroscience experiments are feasible using a transparent fMRI-compatible haptic interface that supports arbitrary three degree-of-freedom motor tasks involving motion or force control. Yet, given fMRI's susceptibility to a variety of artifacts [2], any such experiments must demonstrate that they do not confound subsequent analyses.

For this paper, we developed a motor experiment protocol for human subjects that spans the space of unconstrained, goal-directed reaching motions possible in an MRI scanner (Fig. 2). The protocol is a subset of a more general protocol that we developed [13]. Here, we used fMRI measurements obtained during planning, unconstrained visually-guided reaching, maintaining a hold at the goal positions, and returning to a resting state. The protocol's intuitive design elicits reliable motions across multiple subjects with less than ten minutes of training, despite differences in subject physiology.

For the reaching task, we selected spatially disparate goal locations (left, y=-0.14m; middle, y=-0.01m; right, y=0.12m; all, z=0.035m) that span the MRI workspace, which should induce—if it is possible to do so—artifacts related to large arm motions and hand-tracking with the haptic interface. The reaching duration (5s) is sufficient to sample any magnetic susceptibility artifacts that might arise in the MRI scanner, given our temporal resolution (1.57s). As such, it is a suitable testbed to determine how much information task-correlated noise conveys in fMRI measurements made while operating HFI.

III. Using Phantoms to Segregate Haptic fMRI Artifacts from Neural Noise

It is challenging to estimate noise levels during Haptic fMRI experiments by only looking at the brain. This is because subjects can not solely fixate on tasks at hand, which implies that fMRI measurements contain much neural activation that is unrelated to the task. Moreover, there are often systematic artifacts due to heart- and breathing-rate fluctuations, which are not a pressing concern because they can be effectively factorized using general linear model based denoising techniques [15].

To obtain Haptic fMRI noise estimates that do not include neural noise, we designed flexible phantoms that we placed next to a subject's brain and imaged simultaneously with the brain's motor regions (see Appendix–*Phantom Construction* for details). Since the phantom and the brain were spatially only a few centimeters apart, they shared receiver noise characteristics and were similarly affected by magnetic field inhomogeneities due to improper shimming, or due to task-correlated drift induced by limb motion within the MRI bore. As such, testing whether fMRI measurements in a phantom could classify whether a subject was doing nothing, planning, or moving, provided a metric to estimate task-correlated noise. The phantoms should provide no information and should thus be unable to classify any experiment condition from the others.

We also segregated limb motion artifacts from artifacts induced by our haptic interface's motors, encoders, and associated electronics, by scanning a conventional fBIRN agar phantom while an operator performed the same same experiment while standing outside the scanner bore. The operator's motions were one and a half meters away from

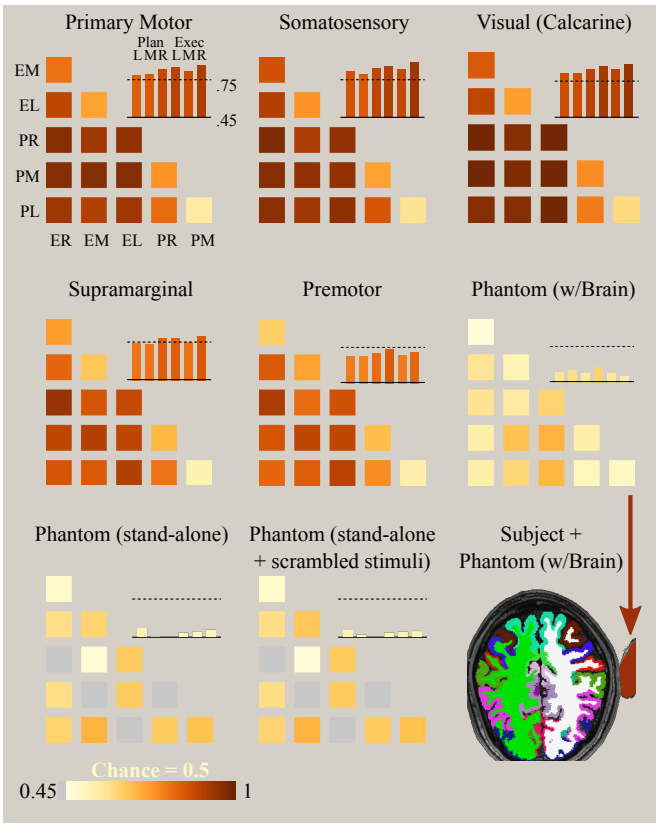

Fig. 3. Classification Accuracy. We classified subject responses to tasks in various brain regions, using a pairwise max-margin classifier. Confusion matrices show consistently high pairwise classification accuracies for the planning and execution tasks in the motor, somatosensory, and visual brain regions (bootstrap medians shown). A phantom that was simultaneously imaged with the brain had substantially lower classification accuracies, indicating that its raw fMRI timeseries was noisy. A stand-alone phantom exhibited similarly low classification accuracies. Moreover, the stand-alone phantom's accuracy did not degrade even after scrambling the stimulus timing specifications, which indicates that its time series was purely noise. A subject's segmented brain slice is shown (bottom-right). The phantom that was simultaneously imaged with it is marked with an arrow.

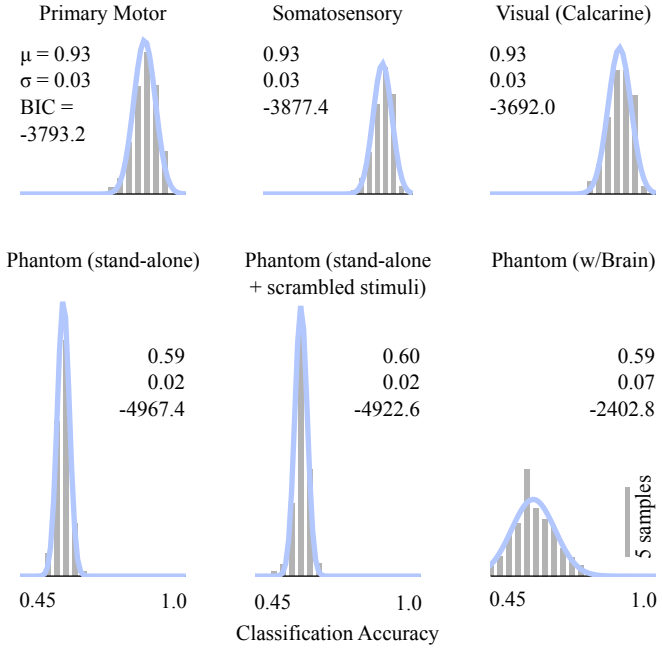

Fig. 4. Classification Accuracy Distribution. Bootstrap distribution of classification accuracies for different brain regions. Classification accuracies for the phantoms were close to random chance. The phantom that was simultaneously imaged with the subject's brain was within 1.5 standard deviations. The classification accuracy distributions for individual brain regions were fit well by a Gaussian, which provided a model to estimate the mean accuracy and standard deviation (insets). The fitting procedure, Gaussian mixture regression, achieved highly negative Bayesian Information Criterion scores (also inset), which indicate goodness of fit. The stand-alone phantom trials had a significantly tighter distribution, potentially due to their homogeneous composition.

the phantom, and thus any magnetic field changes induced by them promised to have a limited effect on fMRI measurements within the phantom. The haptic interface, in contrast, was placed at the same location for both the brain phantom and the stand-alone phantom. As such the noise that it generated would affect fMRI measurements in a similar manner. Moreover, this test also provided a comparison with earlier phantom-based measurements [6], [10], and helped test whether the homogeneous internal composition of conventional phantoms biases noise estimates.

IV. Classification with Brain and Phantom Time Series

A. Classification Accuracy Across Tasks

We classified blood-oxygen-level dependent (BOLD) fMRI measurements across brain regions using a maximum-margin hyperplane (support vector machines; SVM) with leave-one-out cross-validation. For a given task t of task type either p or q we classified pairwise using the following approach. We selected specific brain and phantom regions from our segmentation on which to train the SVM. We bootstrapped [16] 100 voxels from each region. For each instance of either task p or q, from our bootstrapped set of voxels $\{v_1, ..., v_{100}\}$, we constructed a vector of time-series responses for each voxel. Each time series was 10s in length, and began at the start of task t. We simply concatenated these time series into a vector, used these vectors to train a Support Vector Machine using Sequential Minimal Optimization [14] in order to find our maximally seperating hyperplane. Each SVM was trained on approximately 72 such vectors.

Looking at median classification performance across planning and execution tasks, we found that brain regions consistently outperformed both phantoms (Fig. 3). This indicates that Haptic fMRI artifacts, if any, do not substantially interfere with neural activation measurements. We also scrambled the stimulus timing patterns and re-ran the classification on the stand-alone phantom, with similar results as before. This indicates that what seems like artifact-related information in the fMRI time series is actually simply an overfit SVM.

B. Classification Accuracy Across Brain Regions

We further explored the distribution of classification accuracies by looking at bootstrapped histograms for motor, somatosensory, and visual cortex (Fig. 4). To do so, we grouped classifications by whether the two tasks being com-

pared were both plan tasks, both execute tasks, or one plan tasks and one execute task. We used every classification accuracy for each comparison within a group across our 1000 bootstraps to optimize a Gaussian Mixture Model using the Expectation Maximization algorithm. The Gaussian function it returned was then an estimated probability density function of classification accuracies for a random sample from within that region.

Classifiers trained on motor and sensory activity achieved classification accuracy, over 90% in the median case. When separating plan and execute tasks, of the 45,000 classifiers trained on brain regions, every single classifier achieved over 68% accuracy, which was greater than the mean accuracy of these classifiers trained on the phantom region.

C. Classifying a Phantom's Responses

The brain phantom had a wide classification distribution, which mimicked associated brain regions, but had a substantially lower classification performance (within the range of chance). The stand-alone phantoms were also close to random, but, in contrast, had tighter distributions. This was potentially due to their lower temporal noise. These results make us confident that our Haptic fMRI experiment avoids any major confounding artifacts.

Classifiers trained on responses from voxels from the phantom region performed worse on those trained on brain regions; some phantom classifiers performing worse than the 50% random guessing baseline, and no single pair classified better in the median case than on any brain region. In addition, the Bayesian Information Criterion for a gaussian fit was substantially worse in the brain phantom than in actual brain tissue (indicating greater ambient noise levels). This could be because the MRI scanner shimming process is tuned to grey matter, and the phantom is not a perfect match in terms of magnetic susceptibility.

V. Temporal Noise

A common strategy for estimating how a device's RF emmissions interfere with fMRI is to scan a passive dummy object while operating the device. However, because these phantoms tend to be homogeneous, error-correcting field homogenization (*shimming*) becomes simplified, resulting in underestimated noise levels. In contrast, scanning human heads leads to more realistic shim correction, but noise levels are often overestimated because baseline brain activity and head motion become noise covariates. By scanning a phantom and a subject's head simultaneously, we ensured that our estimate of RF noise in the phantom was a conservative overestimate (Fig. 5).

VI. Conclusion

Efforts to combine haptics with fMRI have made steady progress in the recent past, resulting in numerous reports of Haptic fMRI being used for motor neuroimaging studies [4], [8], [9], [17]. Our results support such efforts by demonstrating that carefully designed Haptic fMRI experiments elicit neural activation patterns where informative neural signals dominate any spurious noise. The results also demonstrate that using motor-task classification is a promising technique to quantify motor-task related information in neural activation patterns across the brain.

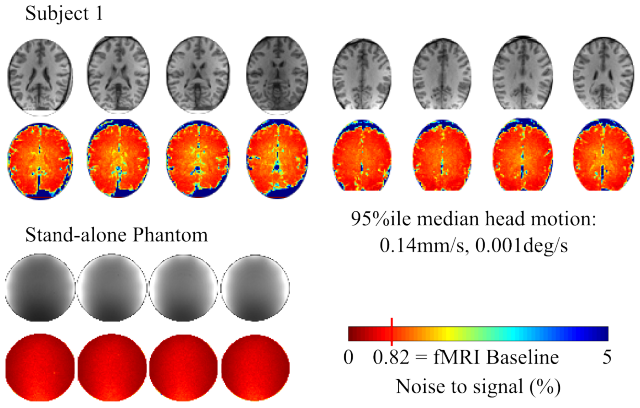

Fig. 5. Temporal Noise. The subject's low head motion helped minimize temporal noise during Haptic fMRI motor experiments. The noise distribution across cortex is similar to the fMRI scanner baseline, demonstrating that HFI does not contribute to temporal noise. Noise was measured across nine 10min runs. The stand-alone phantom had significantly lower noise levels. The phantom's homogeneous interior simplifies magnetic field homogenization, improving the quality of fMRI measurements.

APPENDIX

MRI Protocol

All fMRI scans were conducted at Stanford University's Center for Cognitive and Neurobiological Imaging on a GE Discovery MR750 3 Tesla MRI scanner, with a 32 channel Nova Medical head coil. The scan protocol was gradient echo EPI with a 16cm field of view sampled at a 64×64 resolution (2.5×2.5×2.5 mm^3 voxels), a 1.57s repetition time, a 28ms echo time, and a 72^o flip angle. All scan runs were preceeded by 2^{nd}-order polynomial shimming and were sandwiched by spiral fieldmap scans (2.5×2.5×5 mm^3 voxels). After scanning, the fMRI images were slice time corrected, motion corrected (SPM), spatially undistorted using fieldmaps, and analyzed to compute temporal noise-to-signal.

Haptics motions

Subject used HFI [11] to execute right handed motions across the MRI scanner's workspace. HFI is MRI-compatible [18] and operates without RF interference in the scanner room [11]. HFI's haptic control (and motion monitoring) rate was 350Hz. HFI's low position error (0.025mm [11]) enabled precise hand speed measurements. Hand velocities were resampled to the fMRI TR using cubic spline interpolation in order to compare hand velocity with neural responses.

Human Subject

Subject was a healthy right-handed male with no history of motor disorders: 20y, 165lb, 6'0". Informed consent was obtained in advance on a protocol approved by the Institutional Review Board (IRB) at Stanford University.

Data Collection

Each data run included twelve task sections, four each for the three different reach locations. Each section included two feedback control tasks, so each task type (y- or z-axis) was repeated four times across each spatial reach location. As such, for each spatial location, a run produced eight plans, four reaches, and four feedback control tasks of each type. Subject executed one practice run inside the MRI scanner, and then executed nine scan runs. Each run was six hundred and thirty seconds long.

Head Motion

Subject used a bite-bar with a custom dental dam to minimize head motion. We analyzed head motions by associating SPM's motion correction estimates with hand speeds and velocities measured with HFI. We resampled the hand speeds and velocities to match motion correction estimates using cubic spline interpolation.

Segmentation

We performed cortical parcellation on the subject's brains according to Freesurfer's Desikan-Killiany Atlas. The segmentation was based on a high resolution T1 anatomical scan. We performed a watershed skull trip, and then removed any remaining skull voxels by hand, before applying the segmentation software to the T1. We then coregistered this segmentation to the fMRI inplane using SPM. Using itk-SNAP, we augmented the co-registered segmentation by inserting a unique label for the phantom.

Phantom Construction

Phantoms were made using 14.17g (0.5oz) xanthan gum powder and 710mL (3c) distilled water, resulting in a 2% xanthan gum gel[19]. The xanthan gum powder was added to the water at room temperature and mixed thoroughly for thirty minutes using a 250W hand mixer. The mixture was left to settle for thirty minutes to allow large bubbles to escape. Two phantoms were made from two nested quart size, polyethylene plastic bags, each roughly containing 250 mL of the xanthan gel mixture. Each phantom was folded in half, with the xanthan gum mixture completely contained in one half, resulting in long, rectangular phantoms.

Phantom Setup

Phantoms were placed on either side of the subject's head and held in place using small head support cushions. The head coil cradled the cushions such that the phantoms were snug against the subject's head and the xanthan gum gel was relatively evenly distributed inside each phantom. Our shim and functional scans included the left phantom only.

ACKNOWLEDGMENT

We acknowledge Hari Ganti's assistance with conducting experiments, and Jack Zhu's contribution to our analysis pipeline. We thank Kendrick Kay, Laima Baltusis and Robert Dougherty for helping develop fMRI scanning and data processing protocols.

REFERENCES

[1] N. K. Logothetis and B. A. Wandell, "Interpreting the bold signal," *Annu. Rev. Physiol.*, vol. 66, pp. 735–769, 2004.

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

[3] J. Diedrichsen and R. Shadmehr, "Detecting and adjusting for artifacts in fMRI time series data." *NeuroImage*, vol. 27, no. 3, pp. 624–34, Sept. 2005.

[4] J. Diedrichsen, Y. Hashambhoy, T. Rane, and R. Shadmehr, "Neural correlates of reach errors," *The Journal of Neuroscience*, vol. 25, no. 43, pp. 9919–9931, 2005.

[5] S. Menon, A. Stanley, J. Zhu, A. Okamura, and O. Khatib, "Mapping stiffness perception in the brain with an fMRI-compatible particle-jamming haptic interface," in *Proceedings of the 14th Annual Conference of the IEEE Engineering in Medicine and Biology Society*, August 2014.

[6] R. Gassert, L. Dovat, O. Lambercy, Y. Ruffieux, D. Chapuis, G. Ganesh, E. Burdet, and H. Bleuler, "A 2-dof fmri compatible haptic interface to investigate the neural control of arm movements," in *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, 2006, pp. 3825–3831.

[7] H. Imamizu, S. Miyauchi, T. Tamada, Y. Sasaki, R. Takino, B. Pütz, T. Yoshioka, and M. Kawato, "Human cerebellar activity reflecting an acquired internal model of a novel tool," *Nature*, vol. 403, pp. 192–195, 2000.

[8] R. Moser, R. Gassert, E. Burdet, L. Sache, H. R. Woodtli, J. Erni, W. Maeder, and H. Bleuler, "An MR compatible robot technology," in *Robotics and Automation, 2003. Proceedings. ICRA '03. IEEE International Conference on*, vol. 1, Sept 2003, pp. 670–675 vol.1.

[9] E. Burdet, R. Gassert, G. Gowrishankar, and H. Bleuler, "fMRI compatible haptic interfaces to investigate human motor control," *Experimental Robotics IX*, vol. 21, pp. 25–34, 2006.

[10] A. Hribar, B. Koritnik, and M. Munih, "Phantom haptic device upgrade for use in fmri," *Medical and Biological Engineering and Computing*, vol. 47, pp. 677–684, 2009.

[11] S. Menon, G. Brantner, C. Aholt, K. Kay, and O. Khatib, "Haptic fMRI : Combining functional neuroimaging with haptics for studying the brain's motor control representation," in *Proceedings of the 13th Annual Conference of the IEEE Engineering in Medicine and Biology Society*, July 2013, pp. 4137–42.

[12] M. Kostic, D. Popovic, and M. Popovic, "Influence of planar manipulandum to the hand trajectory during point to point movement," in *IEEE International Conference on Rehabilitation Robotics*, July 2011, pp. 1–4.

[13] S. Menon, M. Yu, K. Kay, and O. Khatib, "Haptic fMRI : Accurately estimating neural responses in motor, pre-motor, and somatosensory cortex during complex motor tasks," in *Proceedings of the 14th Annual Conference of the IEEE Engineering in Medicine and Biology Society*, August 2014.

[14] J. Platt, "Sequential minimal optimization: A fast algorithm for training support vector machines," Microsoft Research, TechReport MSR-TR-98-14, April 1998.

[15] K. Kay, A. Rokem, J. Winawer, R. Dougherty, and B. Wandell, "Glmdenoise: a fast, automated technique for denoising task-based fmri data," *Frontiers in Neuroscience*, vol. 7, no. 247, 2013.

[16] B. Efron and R. Tibshirani, *An introduction to the bootstrap*. Chapman & Hall/CRC, 1994, vol. 57.

[17] N. Yu, N. Estévez, M. Hepp-Reymond, S. Kollias, and R. Riener, "fmri assessment of upper extremity related brain activation with an mri-compatible manipulandum," *International journal of computer assisted radiology and surgery*, vol. 6, no. 3, pp. 447–455, 2011.

[18] R. Gassert, E. Burdet, and K. Chinzei, "Mri-compatible robotics," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 27, no. 3, pp. 12–14, 2008.

[19] A. Hellerbach, V. Schuster, A. Jansen, and J. Sommer, "Mri phantoms are there alternatives to agar?" *PLoS ONE*, vol. 8, no. 8, p. e70343, 08 2013.