Evidence for error feedback control during intrinsic neuromodulation of emotion

Keith A. Bush[1], Josh M. Cisler[2], G. Andrew James[1], Clinton D. Kilts[1]

[1]Brain Imaging Research Center, Psychiatric Research Institute, University of Arkansas for Medical Sciences, Little Rock, AR

[2]Department of Psychiatry, University of Wisconsin – Madison, Madison, WI

Introduction

Real-time functional magnetic resonance imaging (rtfMRI) is a transformative technology for exploring and regulating human cognition. By generating near instantaneous information about the activity of the brain, rtfMRI allows research participants to use a brain state read-out to guide the volitional modulation of their cognitive state[1] (i.e., intrinsic neuromodulation). This technology mimics closed-loop control systems used by engineers to impart autonomous decision-making to physical devices[2]. To test for evidence of this process in humans, we conducted post-hoc control theoretic analysis of a task involving volitional emotional responding to traumatic memories in both the presence and absence of rtfMRI-guidance.

Neuromodulation Model

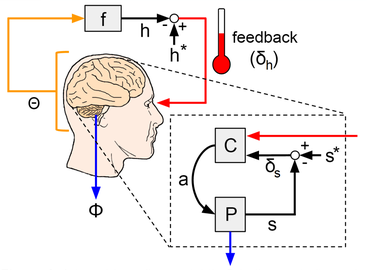


Figure 1: *Internal loop (lower right): C is the controller; a is the action; P is the plant; s is the state of the plant; s* is the internal reference (goal); and, δ_s is the error state of the plant with respect to the reference (i.e., the internal error). External loop (upper left): Θ represents real-time fMRI; f is a mapping between fMRI and the plant's measure, h, which is compared to the external reference, h*, to form the external error signal, δ_h. Φ is psychophysiological correlates of Θ.*

We hypothesize that guided neuromodulation is a dual input error feedback control system, according to the model depicted in Figure 1.

Intrinsic Neuromodulation (IN): The internal loop (Fig. 1) depicts an internal error feedback control system using the block diagram convention of engineering control theory[2].The controller, C, selects an action, a, that induces a change in the target process, P (termed the plant) to minimize internal error, δ_s.

Guided-Intrinsic Neuromodulation (GIN): The internal loop is combined with an external feedback loop (Fig 1). rtfMRI facilitates construction of an external error signal, δ_h, which is visually forwarded (as feedback) to the controller, C. The controller selects action, a, to induce a change in P that simultaneously minimizes both δ_s and δ_h.

Neuromodulation Task

Our dataset was one of convenience, drawn from an experiment in which (n=13) subjects volitionally modulated (enhanced or suppressed) their emotional response to their trauma/stress memory in either the presence or absence of rtfMRI-guidance. Subjects were female, aged 21-50, and included both healthy controls (n=6) as well as those with past exposure to assaultive violence (n=7). The experiment was conducted in three parts.

fMRI Acquisition Scans (data collection phase): two scans, each comprised of two 3 min visual presentations of the trauma/stress script (with concurrent audio narration), each preceded by instructions to enhance or suppress emotional responses elicited by the trauma/stress script.

fMRI Acquisition Scans (model building phase): the resulting fMRI data were then used as input into a support vector machine (SVM) algorithm[3] to train a classifier that discriminates brain activation states or patterns as reflecting either enhanced (+1) or suppressed (-1) emotional responding to the trauma/stress script.

rtfMRI Modulation Scans: two scans, each comprised of eight 40 s trials during which the trauma/stress script was presented. Each trial was preceded by an instruction to enhance or suppress script-related emotional responding as well as whether or not rtfMRI-guidance would be supplied. Each participant's idiosyncratic SVM classifier predicted a hyperplane distance for each TR of each trial of each scan, similar to the trace depicted in Figure 2. rtfMRI-guidance was generated by varying the transparency of the color of the font used to visually display the trauma/stress scripts (white font on a black background). Highly transparent font appeared dark. Highly opaque font appeared bright white. Font color during non-rtfMRI-guided trials used fixed transparency (appearing light gray).

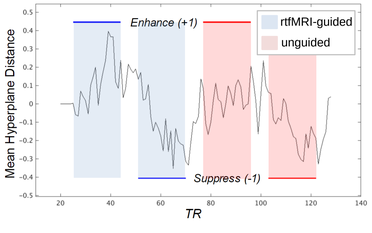


Figure 2*: Example neuromodulation task design and hyperplane distances, h, drawn from the first scan of the rtfMRI Modulation Scans of the experiment. Trials in which the subject volitionally modulates (enhances or suppresses) the emotional response to their trauma/stress script in the presence of rtfMRI-guidance are labeled in bold blue. Similar trials in which no rtfMRI-guidance is provided are labeled in bold red. Group mean hyperplane distances for each TR are plotted in black to facilitate visualization of how hyperplane distances track the enhance (+1) and suppress (-1) goals for each of the two task formats. TRs without predicted hyperplanes are comprised of rest or instruction blocks.*

Control Theoretic Analysis

Image Preprocessing: Images underwent (in order) despiking, slice timing correction, deobliquing, motion correction using rigid body alignment, alignment to normalized anatomical images, spatial smoothing (8 mm FWHM Gaussian filter), detrending, high-pass filtering (0.0078 Hz cutoff), and rescaling into percent signal change.

Independent Components Analysis: Preprocessed fMRI data were reduced to (n=17) functionally independent group-level networks via independent components analysis (ICA) implemented within the GIFT Matlab toolbox[4] as well as removal of low stability and unusable components (ventricles and CSF). Subject fMRI data was then projected into ICA-space.

Calculation of Regressors: Regressors representing theoretical components of the neuromodulation model were computed for each subject according to the equations specified in Box 1.

General Linear Modeling (GLM): Independent components were regressed onto control theoretic regressors, yielding betas, $\beta_{i,j}(k)$, where i is the subject, j is the component, and k is the regressor.

Neuromodulation Performance Prediction: Group-level beta coefficients, $\beta_j(z_{cmb})$, were fit (via iteratively reweighted least squares) to predict control performance as a function of the neuromodulation model componentry, where j is the independent component and z_{cmb} is Fischer-transformed correlation of the plant to the task design across all modulation trials.

Regression Targets

Plant:
h_t

Error Processing:
$\delta_{s,t} \approx |h_t^* - h_t|$
$\delta_{h,t} \approx f_{hrf}(|h_t^* - h_{t+lag}|)$

Cognitive Dynamics:
$a_t = \partial h_t / \partial t$.
$C_t = \partial a_t / \partial t = \partial^2 h_t / \partial t^2$

Box 1*: Regression targets. Predicted hyperplane distance at time, t (measured in TRs) is denoted, h_t. For the external error regression target, δ_h, t+lag represents the time at which feedback was presented to the subject, including lag from the real-time processing pipeline. Function, f_{hrf}, denotes convolution of the error signal with the hemodynamic response function. Derivatives are calculated numerically according to* $\partial h/\partial t \approx (h_{t+1} - h_{t-1})/2$.

Error Processing

Localization of Error Processing: We identified both internal and external error processing components that significantly predict task performance. Internal error processing was found within the anterior cingulate cortex as well as bilateral frontal pole (see Fig. 3). External error processing was found across three components comprised of lateral prefrontal cortex, striatum, superior temporal lobe, and anterior insula (see Fig. 4).

Error Processing Engagement Predicts Neuromodulation Task Performance. As is hypothesized by the guided-intrinsic neuromodulation model, combining the significant internal error processing component with the most significant external error processing component (striatum), permitted us to explain 57% of individual variation in overall neuromodulation performance (see Fig. 5).

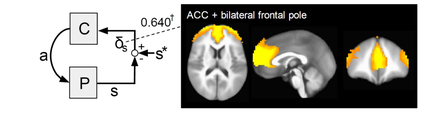


Figure 3 (above)*: Independent component linked, via GLM, with internal error processing, $\beta_{i,j}(\delta_s)$, that significantly predict task performance, $\beta_{ACCfp}(z_{cmb})$=0.640 (p<0.0001, robustfit, p<0.05 FDR_{corr}). The depicted component represents both the dorsal anterior cingulate cortex and bi-lateral frontal pole. The anatomical slices shown are X=0, Y=44, Z=16.*

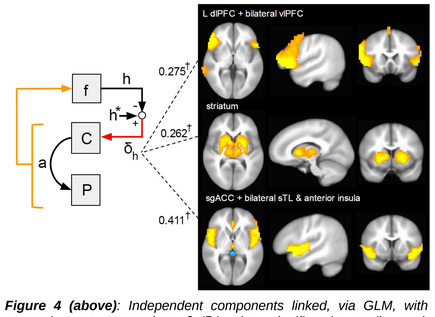


Figure 4 (above)*: Independent components linked, via GLM, with external error processing, $\beta_{i,j}(\delta_h)$, that significantly predict task performance, $\beta_j(z_{cmb})$ for subject i and ICA component j: **(top)** component representing the left vlPFC and left dlPFC, shown at X=-50, Y=19, Z=1. $\beta_{lPFC}(z_{cmb})$ = 0.275 (p=0.008, robustfit, p<0.05 FDR_{corr}); **(middle)** component representing the dorsal striatum, shown at X=-12, Y=8, Z=7. $\beta_{str}(z_{cmb})$ = 0.262 (p=0.004, robustfit, p<0.05 FDR_{corr}); **(bottom)** component representing anterior insula, superior temporal lobe and subgenual ACC, shown at X=-47, Y=10, Z=3. $\beta_{21}(z_{cmb})$ = 0.411 (p=0.004, robustfit, p<0.05 FDR_{corr}).*

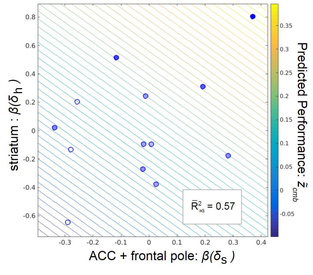
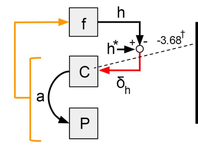


Figure 5 (left)*: Model predicting neuromodulation task performance as a function internal and external error processing. Filled circles represent the $\{\beta_{lACCfp}(\delta_s), \beta_{str}(\delta_h)\}$ coordinates for individual subjects, i. The opacity of the circles visually represents the subject's true performance (fully transparent represents z_{cmb} = -0.048 and fully opaque represents z_{cmb} = 0.419). The combined model exhibits a median leave-one-out-cross-validated (LOOCV) R^2-adjusted value of 0.57).*

Control Law

We identified a component comprised of the right frontoparietal network (FPN) that exhibits a very significant negative relationship with the neuromodulation model's theorized control law, i.e., the mapping between error processing and the controller, C (see Fig. 6). One possible explanation of this negative relationship is that the controller, C, is driven by error dynamics rather than cognitive dynamics (see Fig. 7). Further, we found that right FPN is more strongly activated by the striatum during rtfMRI-guided neuromodulation than it is by the ACC during unguided neuromodulation (see Fig. 8).



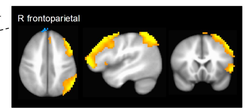


Figure 6 (right)*: Independent component linked, via GLM, with cognitive dynamics, $\beta_{i,j}(\partial^2 h/\partial t^2)$, that is significantly predicted by external error processing of the striatum, $(\beta_{str}(\beta_{FPN}(\partial^2 h/\partial t^2))$ = -3.68 (p=0.0001, robustfit, p<0.05 FDR_{corr}). The depicted component represents the right frontoparietal network (FPN). The anatomical slices shown are X=45, Y=28, Z=40.*

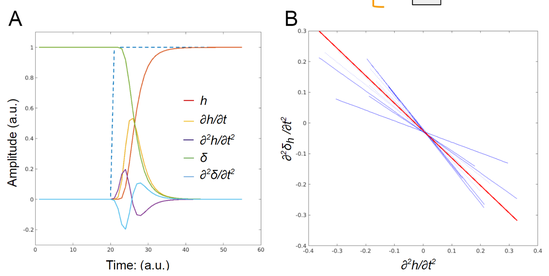


Figure 7 (left)*: Explanations of the relationship between higher-order derivatives of system state and error. **(A)** A theoretical control system executing a positive step-response, exhibiting the negative correlation (r = -1.0) between second derivatives of the plant and error. **(B)** Linear mixed effects model (LMM) of negative relationship between the second derivative of the plant, $\partial^2 h/\partial t^2$ and the external error, $\partial^2 \delta_h/\partial t^2$ observed during GIN trials.*

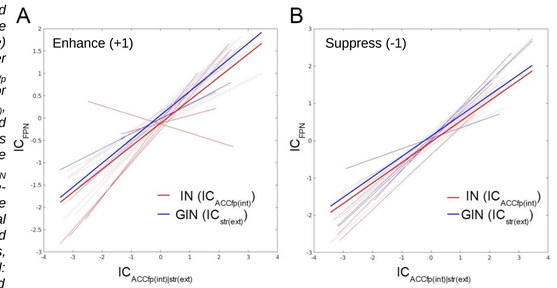


Figure 8 (right)*: The proposed neuromodulation control law. The law describes group (bold line) relationships between either $IC_{ACCfp(int)}$, representing IC_{ACCfp} activity restricted to internal error processing tasks, or $IC_{str(ext)}$, representing IC_{str} activity restricted to external error processing tasks and the second derivative of the cognitive dynamics, IC_{FPN}. IC_{FPN} activity is restricted to those time-points coinciding with the appropriate internal and external error processing tasks and enhance and suppress goals, respectively. **(A)** Enhance goal: GIN LMM (line = bold blue; fixed effect β_1 = 0.537, p<0.0001; fixed effect β_0 = [-0.013, 0.137]; random predictor and intercept effects survived model selection); IN LMM (line = bold red; fixed effect β_1 = 0.516, p<0.0001; fixed effect β_0 = [-0.187,-0.045]). **(B)** Suppress goal: GIN LMM (line = bold blue, fixed effect β_1 = 0.549, p<0.0001; fixed effect β_0 = [0.029, 0.214]; random predictor and intercept effects survived model selection); IN LMM (line = bold red, fixed effect β_1 = 0.551, p<0.0001; fixed effect β_0 = [-0.106, 0.034]). Significant random predictor effects (GIN solid blue; IN solid red) and insignificant random slope effects (GIN dashed blue; IN dashed red) are plotted along with fixed effects.*

Discussion

We conducted post-hoc control theoretic analysis of an rtfMRI-guided neuromodulation task according to a dual input error feedback control model. Analysis identified functional neurocircuitry that explained 57% of individual performance variance on the task. Anterior cingulate cortex, lateral prefrontal cortex, anterior insula, and striatum were all implicated in error processing whereas right frontoparietal network was implicated in the mapping error signals to cognitive changes, suggesting its role as part of a control law. Based on the involved neurocircuitry, neuromodulation exhibits strong functional ties to cognitive control.

References

[1] Weiskopf N. *et al.* Physiological self-regulation of regional brain activity using real-time functional magnetic resonance imaging (fMRI): methodology and exemplary data. *NeuroImage* 19, 577–586 (2003).

[2] Brogan WL. *Modern Control Theory*. (Prentice-Hall, 1991).

[3] Boser BE, Guyon IM, Vapnik VN. A Training Algorithm for Optimal Margin Classifiers. In *Proceedings of the fifth annual workshop on Computational Learning 144*–152 (1992).

[4] Calhoun VD, Adali T, Pearlson GD, Pekar JJ. A Method for Making Group Inferences from Functional MRI Data Using Independent Component Analysis. *Human Brain Mapping*, *14*, 141–151 (2001).

Acknowledgements

This work was supported in part by the Arkansas Science and Technology Authority (ASTA) via the Basic Science Grant mechanism (Grant number: 15-B-3) as well as the generous support of Helen Porter. We would like to thank Sonet Smitherman, Jennifer Payne, and Emily Hahn for their assistance in acquiring data.