A Bayesian statistical analysis of behavioral facilitation associated with deep brain stimulation

Anne C. Smith[a,*], Sudhin A. Shah[b], Andrew E. Hudson[c], Keith P. Purpura[d], Jonathan D. Victor[d], Emery N. Brown[e,f,g], Nicholas D. Schiff[d]

[a] Department of Anesthesiology and Pain Medicine, University of California, Davis, CA 95616, United States
[b] Department of Physiology and Biophysics, Weill Cornell Med College, 1300 York Avenue, New York, NY 10065, United States
[c] Department of Anesthesiology, Weill Cornell Med College, 1300 York Avenue, New York, NY 10065, United States
[d] Department of Neurology and Neuroscience, Weill Cornell Med College, 1300 York Avenue, New York, NY 10065, United States
[e] Neuroscience Statistics Research Laboratory, Department of Anesthesia and Critical Care, Massachusetts General Hospital, Boston, MA 02114, United States
[f] Department of Brain and Cognitive Sciences, MIT, Cambridge, MA 02139, United States
[g] MIT-Harvard Division of Health Sciences and Technology, MIT, Cambridge, MA 02139, United States

ARTICLE INFO

ABSTRACT

Deep brain stimulation (DBS) is an established therapy for Parkinson's Disease and is being investigated as a treatment for chronic depression, obsessive compulsive disorder and for facilitating functional recovery of patients in minimally conscious states following brain injury. For all of these applications, quantitative assessments of the behavioral effects of DBS are crucial to determine whether the therapy is effective and, if so, how stimulation parameters can be optimized. Behavioral analyses for DBS are challenging because subject performance is typically assessed from only a small set of discrete measurements made on a discrete rating scale, the time course of DBS effects is unknown, and between-subject differences are often large. We demonstrate how Bayesian state-space methods can be used to characterize the relationship between DBS and behavior comparing our approach with logistic regression in two experiments: the effects of DBS on attention of a macaque monkey performing a reaction-time task, and the effects of DBS on motor behavior of a human patient in a minimally conscious state. The state-space analysis can assess the magnitude of DBS behavioral facilitation (positive or negative) at specific time points and has important implications for developing principled strategies to optimize DBS paradigms.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Deep brain stimulation (DBS) with macroelectrodes is now an accepted treatment for Parkinson's Disease and is currently being evaluated as a treatment for several neuropsychiatric disorders (Perlmutter and Mink, 2006). Objective measures of effectiveness are important in both the routine and investigational uses of DBS—to determine whether DBS is effective at all, and, if so, to optimize stimulation parameters (Volkmann et al., 2006). Developing such measures is challenging, because behavior is usually quantified by a relatively small number of irregularly spaced observations along a discrete rating scale, the time course of DBS effects is typically unknown, and individual differences may preclude averaging across subjects. Moreover, different behavioral effects may evolve with different time courses and in some cases, these time courses may even be bi-phasic or multi-phasic (Greenberg et al., 2006; Mayberg et al., 2005; Schiff et al., 2007).

Behavioral analyses of DBS responses generally rely upon classical neuropsychological tests or assessments of percentage of correct performances and reaction times using standard statistical methods (e.g., Frank et al., 2007; Page and Jahanshahi, 2007). Similarly, behavioral analysis of the effects of microstimulation experiments in subcortical structures or cortical regions typically do not assess dynamic effects (Nakamura and Hikosaka, 2006) or collapse dynamic effects over multiple trials (Williams and Eskandar, 2006). More recently, Schiff et al. (2007) have applied a logistic regression approach to determine the effect of DBS on performance. While their approach, which is applicable to animal studies and clinical trials, suffices to identify the presence or absence of DBS effects, it is limited in that it postulates a specific form for the dynamics.

Here, we extend the method of Schiff et al. (2007) by introducing models that can assess the temporal dynamics of the behavioral response to DBS. Our strategy is to consider four models of increasing complexity and increasingly weaker assumptions concerning the dynamics. Three of these models are logistic regression mod-

els (Fahrmeir and Tutz, 2001; Nelder and Wedderburn, 1972); the fourth is a state-space autoregressive model (Kitagawa and Gersch, 1996). The first two were considered by Schiff et al. (2007) and identify a stimulation effect by including stimulation history as a covariate. The third model captures some detail about the response dynamics by allowing each stimulation bout to have a different effect size. The fourth model avoids assuming a parametric form for the response dynamics, and instead uses a state-space approach. The state-space approach is an extension of earlier models (Smith et al., 2004, 2005, 2007; Wirth et al., 2003) designed to estimate learning in behavioral experiments.

For internal consistency and ease of programming, we implement the four models under a Bayesian framework and estimate the data likelihood using a Monte Carlo Markov chain (MCMC) approach. While this method can be computationally expensive, being able to use existing software (WinBugs; Lunn et al., 2000) reduces the programming burden. In addition, the final results, in the form of samples from the estimated marginal distributions, can be easily used to make statistical comparisons between different experimental time points. For model selection purposes, the software automatically estimates the deviance information criterion (DIC; Spiegelhalter et al., 2002).

We illustrate our methods by analyzing responses from data sets arising from two different studies aimed at facilitating behavioral responsiveness in the context of the injured brain. The first study involves the responses of normal monkeys undergoing DBS performing a visuomotor reaction-time task. The second study involves analysis of behavioral metrics recorded from the crossover phase of a DBS pilot clinical trial involving a severely brain-injured human subject (Schiff et al., 2007). In both cases, the state-space method and the logistic regression method reveal the presence of a behavioral effect of DBS, but the state-space approach identifies a more detailed picture of its dynamics.

2. Materials and methods

2.1. Experimental protocols

We describe in this section the two experimental protocols we study.

2.2. Central thalamic deep brain stimulation in a macaque monkey to induce sustained attention

This protocol represents an animal model for the effects of DBS on attention. As part of an IACUC approved (Weill Cornell Medical College), NIH sponsored study (NS20712; PI, N. Schiff) and subsequent industry sponsored study (IntElect Medical, Inc), two rhesus monkeys (*Macaca mulatta)* were trained to perform a behavioral paradigm requiring sustained visual attention (forewarned visuomotor reaction-time task). The task consisted of making a saccade to a visual target (a red square at one of nine preselected locations chosen at random) followed by a variable period of fixation on the target and detection of a change in target color followed by a bar release. This standard task, which takes approximately 4 s, requires sustained attention because in order to receive a reward, the animal must release the bar within a brief time window cued by the change in target color. The change in target color occurs after a variable delay time chosen at random from a normal distribution (Parasuraman and Davies, 1977). The variable time interval begins after the monkey maintains a stable fixation of the initial red square. Typically, performance is maintained at a high level at the start of the recording session and diminishes over the 2 h session. Breaks of fixation and, more rarely, inappropriately timed bar releases or failures to release the bar on time, all lead to termination of a trial and failure to receive a reward. Thus our behavioral data sets for this

experiment are composed of time series of binary observations with a 1 corresponding to reward being delivered and a 0 corresponding to reward not being delivered at each trial, respectively. The goal of the experiment is to determine whether, once performance has diminished as a result of spontaneous fatigue, DBS allows the animal to recover its pre-fatigue level of performance.

Recordings of extracellular potentials and local field potentials from sharp electrodes placed in the central thalamus were used to localize sites for electrical stimulation (Schiff et al., 2002, 2001). Central thalamic sites were first identified utilizing a 3-dimensional MRI reconstruction and construction of a customized computer-designed chamber attachment to guide the electrodes. Recording sites were referenced to the animal's MRI images and in comparison with standard rhesus monkey atlas coordinates (Paxinos et al., 2000). The central thalamus, including the anterior intralaminar nuclei (central lateral, paracentralis) and related paralaminar regions of the median dorsalis nucleus were targeted for investigation (Schiff and Purpura, 2002). At some recording sites in the central thalamus, the local field potential had an increase in power in the 30–70 Hz band during the delay period of successful trials. These sites were used for electrical stimulation (Schiff et al., 2002). Bouts of electrical stimulation were repeatedly administered to the central thalamus across blocks of trials during the ongoing task performance (see results) delivered through a bipolar recording/stimulating microsyringe (Crist Instruments Co.) using constant current stimulation (FHC Co. Model Pulsar 6 bp) in the intensity range 200–500 μA (M1) or a customized model bipolar stimulation electrode (M2). Stimulation was applied as a brief train of 5 biphasic stimulation pulses (50 μs/phase) at a stimulation frequency of 50 Hz (chosen to reflect the average range of elevated firing rates seen in neuronal populations with increasing firing rates during the delay period, and corresponding to the middle range of high frequencies identified in local field potential recordings in the characterization experiments). The choice of the train of 5 pulses was further inspired by recordings in intralaminar central lateral nucleus (CL) of alert cats that demonstrated bursts of ∼5 spikes at ∼800–1000 Hz at a rate of approximately 40 Hz (Steriade et al., 1993).

2.3. Central thalamic deep brain stimulation in a human clinical trial

A 38-year-old right-handed man who sustained a closed head injury following an assault 6 years prior to the study was enrolled in a clinical trial designed to determine if central thalamic deep brain stimulation can improve arousal regulation and facilitate behavioral responsiveness (Schiff et al., 2007). It was hypothesized that stimulation in this region may allow recruitment of a wider array of networks in the brain and thereby improve the patient's responsiveness. We summarize the key features of this patient and the clinical trial here; further details can be found in Schiff et al. (2007). The patient remained in a minimally conscious state (MCS) defined by Giacino et al. (2002) as a "condition of severely altered consciousness in which minimal but definite behavioral evidence of self or environmental awareness is demonstrated." Quantitative behavioral assessments in MCS patients capture a range of behaviors from small movements in response to environmental stimuli to inconsistent communication through verbalization or gesture. Pre-surgical assessments were conducted to enable pre-selection of the three primary outcome variables. Subsequent to completion of the pre-surgical baseline, the patient underwent implantation of bilateral DBS electrodes. The electrodes remained off until initiation of a DBS titration period which began 50 days after implantation, except for a 2-day period of stimulation testing in the immediate post-operative period. Different combinations of frequency, intensity, electrode contact activation and periods of ON and OFF times were tested

in order to optimize behavioral outcome responses. The primary outcome measures were subscales of the JFK Coma Recovery Scale (Revised). In addition, three secondary outcome measures were used (object naming, limb control and oral feeding). Further details of these measures can be found in the Supplementary information to Schiff et al. (2007).

At the end of the titration phase, a 6-month double-blinded alternating ON/OFF cross-over trial began using optimal parameters selected during the titration phase. We analyze in detail one of the secondary outcome measures during this period, oral feeding index, assessed by a speech pathologist during scheduled feeding sessions. These indices lie on a scale from 1 to 5, ranging from normal manipulation of food (chewing and swallowing) to oral feeding not attempted due to underarousal or failure to open mouth. Previously, Schiff et al. (2007) analyzed this data as binary data having converted indices 2 or higher to zero and index 1 to one. Here we consider the full data set as binomial observations. The index values are transformed to a scale from 0 to 4 for consistency so that a 0 indicates poor performance and a 4 indicates normal (good) performance.

We also analyze limb movement and arousal data during the same 6-month period. The limb movement data was collected weekly and consisted of between 21 and 24 binary observations of the patient's ability to demonstrate common tasks (as assessed by team of therapists). In total, there were 25 count observations across the 6 month cross-over period. The arousal data consists of between 3 and 4 binary measurements per day.

3. Models

3.1. A state-space model to characterize arousal and attention state dynamics

We construct a state-space model to characterize arousal and attention state dynamics in a subject receiving thalamic stimulation. To define the state-space model we require observation and state equations. The state equation is defined by the following discrete first order autoregressive (AR (1)) process

$$x_k = \rho x_{k-1} + \varepsilon_k \tag{1}$$

where x_k is the arousal state of the patient at time $k = 1, \ldots, K$, ε_k is zero-mean, Gaussian noise with variance σ_ε^2 and ρ, is the correlation between successive observations. The initial state, x_0, is assumed to be unknown.

We consider two possible formats for the observations: data in the form of binary responses (i.e., the monkey behavioral datasets described above) and data in the form of proportions (i.e., the human datasets described above). For the binary response data the observation model is assumed to be Bernoulli where

$$Pr(n_k|x_k, \theta) = p_k^{n_k}(1 - p_k)^{1-n_k} \tag{2}$$

where $\theta = (\rho, \sigma_\varepsilon^2)$ and where $n_k = 0$ if the response is incorrect at time k and $n_k = 1$ if the response is correct for $k = 1, \ldots, K$. Here, p_k is the probability of a correct response and is related to the state using the logistic transform

$$\log[p_k(1 - p_k)^{-1}] = x_k. \tag{3}$$

For the proportion response data the observation model is assumed to be binomial where

$$Pr(n_k|N_k, x_k, \theta) = \binom{N_k}{n_k} p_k^{n_k}(1 - p_k)^{N_k-n_k} \tag{4}$$

where n_k and N_k are slightly different for each of the human data sets and are described as follows. The human oral feeding data indices range from 0 (unarousable) to 4 (chewing, swallowing, and completing a meal). In this case, n_k is the observed index on day k and $N_k = 4$ on all days $k = 1, \ldots, K$. For the human limb control data, n_k is the sum of the binary observations in week k and N_k varies between 21 and 24 total observations in that week. For the arousal data, n_k is the sum of observed binary indices per day and N_k is the total number of observations (3 or 4) on that day. We compute a smoothed estimate of this normalized performance data p_k as a function therefore of day k for the oral feeding and arousal data and as a function of week k for the limb control data.

For the latter two data sets (limb control and arousal) the use of a binomial model is the principled choice for independent binary observations. For the oral feeding data, the value of n_k is based on an observed index between 0 and 4, and there is no a priori expectation for the distribution. We also chose to implement a binomial link for this data for simplicity and consistency with the other types of data and models. (We also considered a Poisson link, and found that, according to the DIC, it provided a worse fit.)

3.2. A logistic regression approach to characterize arousal and attention state dynamics

For comparison we also fitted a logistic regression model (LR) to the data using a similar structure to the likelihood-based model used in Schiff et al. (2007). The proposed model for the cognitive state is

$$x_k = A + Bk + C\sum_{k_\tau=0}^{k} S_{k_\tau} R \exp[-R(k - k_\tau)] \tag{5}$$

where A, B and C are parameters to be estimated. The final term in Eq. (5) determines how much stimulation history affects the cognitive state. Indicator S_k takes values +1 and 0 depending on whether stimulation is on or off, respectively. The parameter, R, in Eq. (5) determines how fast the electrical stimulation affects the cognitive state, and has units of 1/(sampling interval). Schiff et al. (2007) found that model fits were equally adequate over a broad range of values of R, and chose $R = 1$; we do the same here. We assume Bernoulli/binomial observation models as in the state-space model above (Eqs. (2) and (4)). The estimated performance is linked to the data with the logistic transformation (Eq. (3)).

As in Schiff et al. (2007) we fitted the models without stimulation effect (i.e., with parameters A and B, but not C) and with stimulation effect (i.e., with parameters A, B and C). We refer to these two models as LR AB and LR ABC1, respectively. In addition, we also considered a third version of the model that allowed for different values of C for each stimulation period. To implement this, we define multiple stimulation vectors S_τ^i for $i = 1, \ldots, S$, where S is the total number of stimulation ON periods. Each vector S_τ^i takes values zero except during the ith stimulation ON period where it is +1. The cognitive state Eq. (5) becomes

$$x_k = A + Bk + \sum_{i=1}^{S} C_i \sum_{k_\tau=0}^{k} S_{k_\tau}^i \exp[-(k - k_\tau)]. \tag{6}$$

We refer to this model as LR ABC2.

3.3. Bayesian model fitting: state-space model

Previously we have computed estimates of the unknown state and parameters using maximum-likelihood approaches and the EM algorithm (Dempster et al., 1977; Smith and Brown, 2003; Smith et al., 2004). We make use here of a Bayesian approach and Monte Carlo Markov chain estimation methods (e.g., Congdon, 2003; Robert and Casella, 2004; Smith et al., 2007). We let $n = \{n_1, \ldots, n_K\}$ for the binary data or $n = \{n_1, \ldots, n_K, N_1, \ldots, N_K\}$ for

the binomial data and $x = \{x_1, \ldots, x_K\}$. Defining the prior distribution for the parameters as $p(\theta)$, the posterior distribution of the parameters and the state, x, given the data is

$$p(\theta, x|n) = \frac{p(n|x, \theta)p(x|\theta)p(x_0)p(\theta)}{p(n)} \tag{7}$$

where

$$p(n|x, \theta) = \prod_{k=1}^{K} p_k^{n_k}(1 - p_k)^{1-n_k} \tag{8}$$

for the Bernoulli model

$$p(n|x, \theta) = \prod_{k=1}^{K} \binom{N_k}{n_k} p_k^{n_k}(1 - p_k)^{N_k - n_k} \tag{9}$$

for the binomial model

$$p(x|\theta) = \prod_{k=1}^{K} (2\pi\sigma_\varepsilon^2)^{-1/2} \exp\left\{ -\frac{1}{2} \frac{(x_k - \rho x_{k-1})^2}{\sigma_\varepsilon^2} \right\} \tag{10}$$

and

$$p(n) = \int p(\theta, x, n)\, d\theta\, dx. \tag{11}$$

Under the Bayesian framework, it is required that we specify priors on the components of $\theta = (\rho, \sigma_\varepsilon^2)$ in addition to a prior distribution on the initial state, x_0. For the autoregressive decay parameter ρ we assumed a uniform prior on the interval $[-1, 1]$. For the inverse variance (= precision), τ, we used a broad gamma prior with location and shape parameters 0.1 and 0.01 respectively corresponding to a mean of 1 and variance of 100 for both models. The distribution of the initial state is assumed to be Gaussian with zero mean and variance $\sigma_\varepsilon^2/(1 - \rho^2)$. The Bayesian analysis provides an estimate of the posterior probability density of all the unknown parameters of the model. From these sample estimates it is possible to directly estimate the mean and median as well as 100% $(1 - \alpha)$ Bayesian credible intervals.

We can evaluate Eq. (7) using standard Monte Carlo Markov Chain methods in WinBUGS (Lunn et al., 2000). Implementation in this software merely requires specifying the model equations (Eqs. (1)–(3) or (1), (3), (4)) as well as priors for the unknown parameters $p(\theta)$. The Bayesian approach treats the missing observations as another parameter to be estimated. For the problems considered in Section 4 we found that satisfactory convergence of each parameter's posterior probability density was achieved using 10,000 iterations of 3 chains after 5000 burn-in iterations. Convergence was confirmed by visual analysis of the mixing of the 3 chains for a subset of the parameters as well as a requirement that the Brooks–Gelman–Rubin (BGR) statistic be <1.2 for all parameters (Brooks and Gelman, 1998; Gelman and Rubin, 1992). Our choice of number of chains and iterations was made based recommendations made in Kass et al. (1998). For a binary data set of length 1350 trials (as in our first example) run time ranges from 1 min (LR AB) to 40 min (state-space) on a 2.39 GHz desktop computer with 2 GB RAM. We implemented the models using the software package WinBUGS because of its user-friendly model declaration language. WinBUGS is not designed to handle extremely large models and data sets (e.g., >2000 trials). Other software (or a filtering approach) may be preferable in these situations. For the models here we provide the WinBUGs code and a Matlab script (The MathWorks, Natick, MA) to drive it using Matbugs (Murphy and Mahdaviani, 2005) at the following web address: http://www.neurostat.mit.edu/behaviorallearning.

3.4. Bayesian model fitting: logistic regression models

For comparison with the state-space results we chose to implement the logistic regression models under the same Bayesian framework instead of the maximum-likelihood approach used previously (Schiff et al., 2007). Priors on the parameters A, B and C were zero-mean Gaussian with very large variance (100,000), ensuring that the prior did not materially influence the behavioral state in any particular direction.

3.5. Model selection

Previously Schiff et al. (2007) used the likelihood ratio test to compare their non-Bayesian (nested) regression models and used this method to demonstrate the significance of the stimulation effect in the data. Since we now have 4 (non-nested) Bayesian models to compare, we make use of the estimated mean of the posterior deviance and the deviance information criterion (Spiegelhalter et al., 2002). The estimated mean of the posterior deviance ($\overline{D}$) is the Monte Carlo estimate of the deviance ($-2\log(\text{likelihood})$). The DIC is an approximate Bayesian analogue of the Akaike Information Criterion (AIC) used in model selection and is a technique for penalizing the deviance when a large number of model parameters is employed. Since in the Bayesian framework, the number of parameters is not necessarily clear, the DIC estimates the number of effective parameters from the Monte Carlo simulation. The values of $\overline{D}$ and DIC are computed as part of the MC simulation in WinBUGS. Typically, one model can be regarded as superior to a second model if its DIC value is lower by more than 5–7 (Spiegelhalter et al., 2002).

3.6. Comparing performance across trials

For both the state-space model and the logistic regression models, comparison between estimated probabilities of the cognitive state at any given trial is straightforward using the Monte Carlo samples computed by the algorithm. Since these samples are drawn from the full posterior joint distribution of cognitive state across all trials, they automatically account for potential covariance between trials and therefore a correction for multiple comparisons is not required. That is, correction for multiple comparisons is typically performed when many independent observations are made. In this situation the covariance between the cognitive states at all trials is estimated, the states are not assumed to be independent, and for each pairwise comparison we are able to ask the simple question: given two covarying states from the full distribution of states, how likely are their distributions to be different? The probability that the distribution of the state at trial x_i is larger than the distribution of the state at trial x_j is then computed by pairing the Monte Carlo samples of x_i and x_j drawn from their joint distribution and making a count, S_{ij}, of how many times $x_i > x_j$. Assuming we have S samples from the distribution of each state, the required probability, $\Pr(x_i > x_j)$, is computed from S_{ij}/S. Performing this calculation for all pairs (i, j) enables us to plot a probability surface that allows comparison of performance across the entire length of the clinical trial/experiment.

3.7. Global comparison of stimulation ON performance with stimulation OFF performance

In addition to comparing estimated cognitive state on a trial-by-trial basis, it is also possible to compare stimulation ON performance with stimulation OFF performance under the Bayesian framework, again, avoiding the issue of multiple comparisons. To do this, we identify two states, x_{ON} and x_{OFF}, defined as the mean of all the cognitive states when stimulation is ON and OFF, respectively.

The probability that the two state distributions are different is computed using the sampling procedure described in the paragraph above.
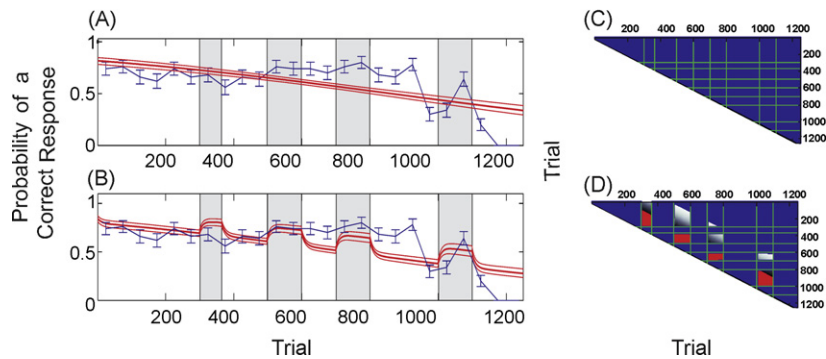
4. Results

4.1. Analysis of central thalamic deep brain stimulation in a macaque monkey to induce sustained attention
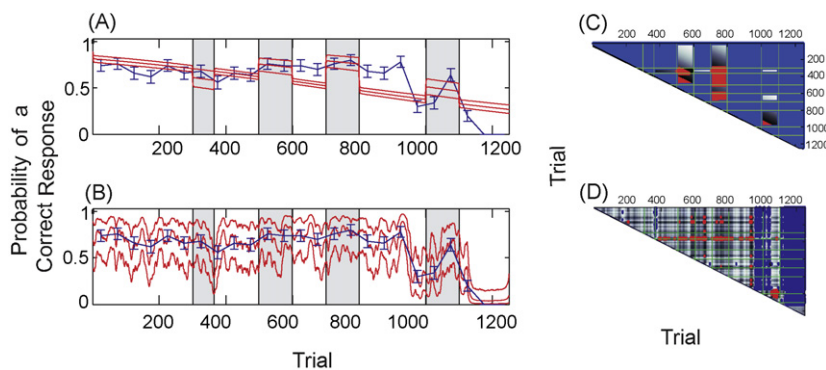
To illustrate the technique applied to binary performance measures, we consider responses from a monkey performing the attention paradigm described in Section 2 (Figs. 1 and 2). In this experiment, the monkey performed 1250 trials. Stimulation was applied during 4 periods across trials 300–364, 498–598, 700–799 and 1000–1099, indicated by shaded gray regions in Figs. 1A and B and 2A and B. The raw data is summarized by the blue lines with error bars, representing the 50-trial block average and its standard error. Dividing the results into periods when stimulation is applied ("ON") and not applied ("OFF"), there are 240 correct responses out of 367 trials during the ON periods and 501 correct responses from 883 trials during OFF periods. According to the standard binomial proportion test the responses during ON periods are more often correct than during OFF periods with high statistical significance ($p < 0.005$).

Fig. 1A and B shows response curves and 95% credible intervals computed using the models LR AB and LR ABC1 described in Section

2. In Fig. 1A (LR AB), no DBS effect is assumed; in Fig. 1B (LR ABC1) stimulation is assumed to have an effect (which can be either positive or negative,) that abates with an exponential time course. For the LR AB model the parameter estimates (and standard errors) for A and B were 1.502 (0.132) and −0.001742 (1.77e−4), respectively. To test the accuracy of the MCMC-based results we also implemented more standard maximum-likelihood methods (glmfit.m: Matlab, The Mathworks) which yielded values for A and B of 1.495 (0.1312) and −0.001733 (1.78e−4), respectively, indicating that the two approaches give similar results. For LR ABC1 the estimated stimulation effect parameter, C, has mean value of approximately 0.37, with 95% credible interval [0.2048, 0.5405]. Since the estimated credible interval of C does not include zero, we may conclude that the stimulation has a significant effect on the fitted response curve. Fig. 2A shows the response curve estimates computed using model LR ABC2. In this model, we computed four values C_1, C_2, C_3 and C_4 corresponding to the stimulation effect on the state during each of the four stimulation-ON periods (Fig. 2A: gray-shaded areas). Under this formulation, the response curve decreases during the first stimulation period (as does the raw data) and increases more in the third stimulation-ON period than the surrounding stimulation periods. The estimated mean values for C_1, C_2, C_3 and C_4 were approximately −0.27, 0.41, 0.80 and 0.36 with corresponding 95% credible intervals [−0.60, 0.07], [0.12, 0.72], [0.50, 1.12] and [0.07, 0.66]. From these results we conclude that at the 95% confidence level the first stimulation-ON period has a performance that

Fig. 1. Models applied to monkey visuomotor reaction-time task. Raw data is shown with blue error bars in Panels A and B representing the average response over 50 trials and its standard error. Stimulus ON periods are indicated by gray shading. (A and B) Performance curves (median and 95% credible intervals) for the LR AB and LR ABC1, respectively. (C and D) Trial-by-trial within experiment comparisons for the performance curves shown in A and B, respectively. Each panel shows a probability surface computed, by comparing by Monte Carlo, the performance curve at a value on the *x*-axis with a value on the *y*-axis. Dark (light) areas indicate that the performance at the trial on the *x*-axis is lower (higher) than the performance at the trial on the *y*-axis. Dark (light) regions were the probability extends beyond the 95% credible interval levels are highlighted in red (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Fig. 2. Models applied to monkey visuomotor reaction-time task. Raw data is shown with blue error bars in Panels A and B representing the average response over 50 trials and its standard error. Stimulus ON periods are indicated by gray shading. (A and B) Performance curves (median and 95% credible intervals) for the LR ABC2 and state-space models, respectively. (C and D) Trial-by-trial within experiment comparisons for the performance curves shown in A and B, respectively. Each panel shows a probability surface computed, by comparing by Monte Carlo, the performance curve at a value on the *x*-axis with a value on the *y*-axis. Dark (light) areas indicate that the performance at the trial on the *x*-axis is lower (higher) than the performance at the trial on the *y*-axis. Dark (light) regions were the probability extends beyond the 95% credible interval levels are highlighted in red (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

is not significantly different from baseline, while the three later stimulation-ON periods are associated with a significant increase in performance.

In Fig. 2B, we plot the estimated probability of a correct response computed using the state-space formulation. We show both the median (wide red curves) and the 95% credible intervals (lighter red curves). In this case the estimated probability is less constrained and tracks the data independent of the stimulation-ON/OFF information. On average the response curve lies around the 0.75 level but decreases are observed at the end of the first stimulation-ON period around trial 375, at the end of the 4th OFF period around trial 950 and for the remainder of the experiment from trial 1100 onwards.

The several models allow for different levels of detail in assessing the dynamics of behavior. Model LR AB cannot even address whether there is a stimulation effect, since it attempts to fit the data without a stimulation-related covariate. Model LR ABC1 can address whether there is a stimulation effect, but not whether it differs across bouts, since it fits all stimulation effects with a single parameter. Model LR ABC2 can address whether there is a difference in stimulation effect across bouts (since each bout is associated with a different model parameter), but cannot assess the dynamics at the level of individual trial. Of the models considered, only the state-space approach can do the latter.

To see this in more detail, we construct surface plots that compare performance between pairs of trials (Figs. 1C and D and 2C and D) using the algorithm described in Section 2. Figs. 1C and D and 2C and D show a surface representing these probability values where dark (light) indicates that the performance on the *x*-axis is lower (higher) than the performance at the trial on the *y*-axis. Dark (light) regions where the probability extends beyond the 95% credible interval levels are highlighted in red (blue).

Comparison across bouts can be performed for the two models that allow for different dynamic effects for each bout, namely, model LR ABC2 (Fig. 2A) and the state-space model (Fig. 2B). For example, Fig. 2C indicates that the performance in the third ON period is significantly greater than ($p > 0.95$) the performance in the first ON period. The results of state-space modeling are somewhat consistent with this: they indicate that the third ON period is greater than the end of the first ON period but that also the second OFF, second ON and third OFF are greater than the end of the first ON period.

Thus, the state-space model provides for an even more detailed analysis – trial-by-trial comparisons – and reveals that the dynamics are quite complex. In Fig. 2D, there are two strong vertical blue blocks from trials 950 to 1050 and from around trial 1100 to the end of the experiment. These correspond to significant drops in performance in the last 2 stimulation OFF periods. Another important effect in the data is highlighted by the red block around the last half of the final ON period (around [*x*, *y*] coordinates [1050, 1000]). This shows that the performance during the final stimulation ON period is significantly higher than the performance at the end of the OFF period before and at the beginning of this final ON period. The most surprising feature of the analysis is that the performance around trial 350 at the end of the ON stimulation period is lower than the performance of the surrounding trials, as indicated by the red horizontal stripe along $y = 350$. Thus, the binomial test indicates that while stimulation ON produces an overall improvement in performance, this does not preclude the presence of periods when performance during stimulation ON is significantly lower than during stimulation OFF. This is an important point and may turn out to yield insight into the dynamics of the behavioral changes produced by DBS and the underlying physiological mechanisms which may include effects on synaptic efficacy and gene expression (Shirvalkar et al., 2006).

Table 1
Computed values of posterior mean of deviance ($\bar{D}$), point estimate of the deviance ($\hat{D}$), estimated effective parameters (p_D) and deviance information criterion for the 4 models applied to the monkey performance data.

	$\bar{D}$	$\hat{D}$	p_D	DIC
LR AB	1583	1581	2	1585
LR ABC1	1565	1562	3	1569
LR ABC2	1546	1540	6	1552
State-space	1340	1237	102	1443

Numbers are rounded to the nearest integer. p_D is computed from $\bar{D} - \hat{D}$.
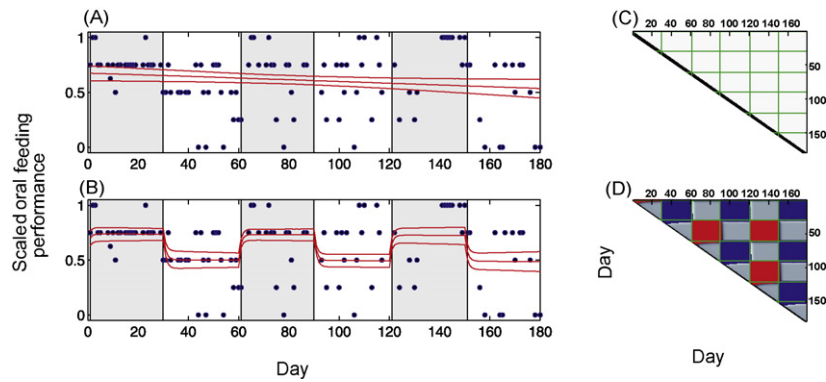
We next consider model selection by examining measures of model fit. Table 1 shows the Monte Carlo derived estimates of posterior mean of the deviance ($\bar{D}$), point estimate of the deviance ($\hat{D}$), estimated effective parameters (p_D) and deviance information criterion for the four models. Based on the likelihood ($\bar{D}$) and DIC, we may conclude that the state-space model provides a better model fit, i.e., a better description of the effects of stimulation, than the LR approaches. We also note that the model fit improves as the stimulation is added (LR AB to LR ABC1) and as the model allows for differential effects of the stimulation on the performance (LR ABC1 to LR ABC2).

As a final test of the state-space model, we compared by Monte Carlo the (multivariate) distribution of the states when the stimulation was ON against the distribution when the stimulation was OFF. Using the pairwise Monte Carlo comparison method described above for Figs. 1C and D and 2C and D, we computed the state during stimulation ON to be significantly greater than the state during stimulation OFF with $p < 0.002$ (based on 6000 Monte Carlo samples).
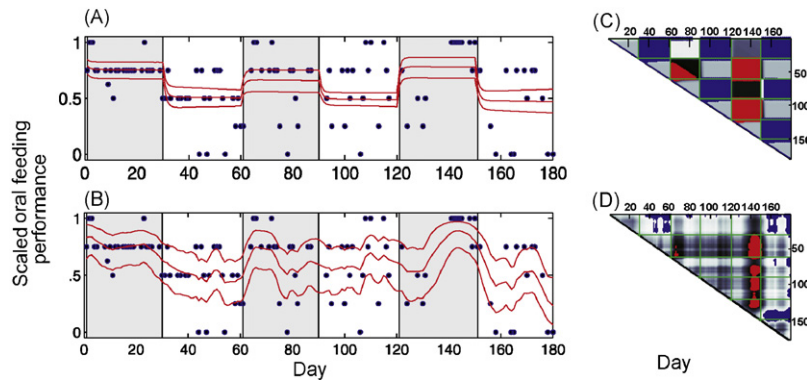
In conclusion, the state-space model provides a better model fit and is able to address more detailed questions than the regression-based models. It is worth noting that, while the state-space model allows more flexibility in tracking the data, the results it yields are largely consistent with the simpler regression models. For example, all the models able to account for stimulation effect (LR ABC1, LR ABC2 and the state-space model) indicate that the stimulation has a positive influence on the performance, consistent with the standard binomial test result. Likewise, both LR ABC2 and the state-space model show that the performance does not improve during the first stimulation ON period. Overall, however, the state-space model results highlight an abrupt step-like decline in performance towards the end of the experiment, around trial 950, which undergoes a significant increase during the final stimulation period before a final significant drop to zero.

4.2. Analysis of central thalamic deep brain stimulation in a human clinical trial

In the second example, the state-space model is applied to a dataset in which behavior is described by a five-level rating scale. The raw data obtained on a daily basis during the crossover period of the clinical trial and described in Section 2 is shown as blue dots in Figs. 3A and B and 4A and B. Note that on some days observations were not made and we treat these data as missing at random. Stimulation was applied during three periods across trials 1–29, 60–87 and 117–151 (shaded regions in Fig. 2A–C). As before, we first divide the data into observations when stimulation was ON ($n = 55$) and observations when stimulation was OFF ($n = 52$). Comparison of the two data sets using the distribution-independent Kolmogorov–Smirnov (KS) test indicates that they are highly significantly different ($p < 0.0007$). We also compare individual ON/OFF periods using the KS test. Of the 15 possible pairwise comparisons, only data from periods 1 and 2 yielded a significant effect ($p < 0.00003$, judged at a level of 0.003, the Bonferroni-corrected criterion corresponding to 0.05).

Fig. 3. Models applied to human oral feeding data. Raw data is shown as blue filled circles in Panels A and B. Stimulus ON periods are indicated by gray shading. (A and B) Performance curves (median and 95% credible intervals) for the LR AB and LR ABC1 models respectively. (C and D) Day by day within experiment comparisons for the feeding performance curves shown in A and B, respectively. Each panel shows a probability surface computed, by comparing by Monte Carlo, the performance curve at a value on the x-axis with a value on the y-axis. Dark (light) areas indicate that the performance at the trial on the x-axis is lower (higher) than the performance at the trial on the y-axis. Dark (light) regions were the probability extends beyond the 95% credible interval levels are highlighted in red (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Fig. 4. Models applied to human oral feeding data. Raw data is shown as blue filled circles in Panels A and B. Stimulus ON periods are indicated by gray shading. (A and B) Performance curves (median and 95% credible intervals) for the LR ABC2 and the state-space models, respectively. (C and D) Day by day within experiment comparisons for the feeding performance curves shown in A and B, respectively. Each panel shows a probability surface computed, by comparing by Monte Carlo, the performance curve at a value on the x-axis with a value on the y-axis. Dark (light) areas indicate that the performance at the trial on the x-axis is lower (higher) than the performance at the trial on the y-axis. Dark (light) regions were the probability extends beyond the 95% credible interval levels are highlighted in red (blue). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

As for the previous example, the data was initially fitted with generalized linear models: without stimulation (LR AB; Fig. 3A), assuming uniform stimulation effect (LR ABC1; Fig. 3B) and assuming possibly different stimulation effects (LR ABC2; Fig. 4A). For these data there is no evident overall trend. However, the estimate of variable, C, which reflects the effect of the stimulation on the regression curve in LR ABC1, is significantly larger than zero, with mean 0.6352 and 95% credible interval [0.3576, 0.9045]. For LR ABC2, the three parameters C_1, C_2 and C_3 have mean values 0.6804, 0.4170 and 0.8547 with 95% credible intervals greater than zero, indicating a significant effect of the stimulation on the cognitive state during all the stimulation ON periods.

Day to day comparison for LR AB (Fig. 3C) indicates no significant trend in the data. For the model with a uniform stimulation effect (LR ABC1), the computed probability surface (Fig. 3D) has a checkerboard-type pattern indicating ON and OFF periods are significantly different from one another. Analysis with the LR ABC2 model suggests that the stimulation effects in the three stimulation bouts are not equally strong. For example, the performance at the beginning of the second ON period is greater than the performance during most of the first OFF period. However, at the end of the second ON period the performance is not significantly greater than the performance in the first OFF period.

The state-space model performance plot (Fig. 4B and D) shows that the dynamics of the process are more complex. The perfor-

mance during the ON periods tends to be higher than during OFF periods but with some exceptions. For example, performance in the 2nd ON period is variable: initially slightly higher than the previous OFF period and, towards the end, lower than the 1st ON period (Fig. 4D, red block at (65, 45) and blue block at (80, 5), respectively). As the third ON period progresses performance dramatically improves (Fig. 4D, long vertical red block) and drops off again significantly during the final OFF period (Fig. 4D, blue vertical blocks on right hand side).

As for the monkey example, we examine model fit using the estimated deviance and DIC (Table 2). We conclude from both the mean deviance $\bar{D}$ and the DIC that the logistic regression model fit is improved by the addition of the stimulation parameter, C, and that

Table 2
Computed values of posterior mean of deviance ($\bar{D}$), point estimate of the deviance ($\hat{D}$), estimated effective parameters (p_D) and deviance information criterion for the 4 models applied to the human oral feeding data.

	$\bar{D}$	$\hat{D}$	p_D	DIC
LR AB	327	325	2	329
LR ABC1	307	304	3	310
LR ABC2	306	301	5	311
Binomial state-space	263	232	31	294

Numbers are rounded to the nearest integer.

Table 3
Summary of analysis of all nine data sets with DIC computed by all four models.

Data set	DIC				Stimulation effect in LR ABC1?	Stimulation affects response from SS?
	LR AB	LR ABC1	LR ABC2	SS		
1. Monkey N0123	1585	1569	1552	**1443**	Positive ($p < 0.001$)	Positive ($p < 0.002$)
2. Monkey 080709	2369	2258	2088	**2037**	Negative ($p < 0.001$)	Negative ($p < 0.01$)
3. Monkey 0806111	1813	1812	1810	**1765**	None	Positive ($p < 0.01$)
4. Monkey 080724	1643	1641	**1637**	1648	Positive ($p < 0.05$)	None
5. Monkey 080703	941	937	933	**891**	Negative ($p < 0.01$)	Negative ($p < 0.01$)
6. Monkey 080825	1861	1861	1861	**1730**	N/A—Stimulation intensity varies between bouts	N/A—Stimulation intensity varies between bouts
7. Human-oral feeding	329	310	311	**294**	Positive ($p < 0.001$)	Positive ($p < 0.0002$)
8. Human limb-moving	217	209	194	**168**	Positive ($p < 0.001$)	Positive ($p < 0.01$)
9. Human-arousal	227	222	225	**205**	Positive ($p < 0.004$)	Positive ($p < 0.003$)

The number of stimulation bouts in each data set varies between two and seven. The stimulation intensity was the same across all bouts with the exception of data set #6 where the there were seven stimulation bouts with different intensities. The lowest value of DIC for each data set is optimal and indicated in bold. The sixth column indicates whether the estimate of parameter C is significantly different than zero as estimated by model LR ABC1. The rightmost column indicates whether overall the state-space estimates when stimulation is on are different than the state-space estimates when stimulation is off.

the state-space model is a better fit than both. In this case there is little difference between the fits for LR ABC1 and LR ABC2.

This example demonstrates that, while the logistic regression approach and the state-space approach both identify a positive effect of stimulation, the state-space approach is able to detail its dynamics. This detail demonstrates that the actual dynamics are more complex than postulated in the logistic regression models, and that there are differences in the effect of stimulation in the several ON/OFF periods. Additionally, the state-space analysis reveals that functional oral feeding declines during the second and third OFF periods, beginning approximately two weeks after offset of stimulation. The logistic regression-based analyses could not identify these changes, since they did not conform to the functional form of the DBS effects that were incorporated into the model.

As with the monkey data set, we performed a final test of the state-space model by comparing by Monte Carlo the (multivariate) distribution of the states when the stimulation was ON against the distribution when the stimulation was OFF. The state during the stimulation ON was significantly greater than the state when the stimulation was OFF with $p < 0.0002$ based on 30,000 Monte Carlo sample comparisons.

Application of LR ABC1, LR ABC2, and the state-space models all indicate that overall the stimulation has a positive effect on oral feeding performance, consistent with the KS test applied to the raw data. The state-space model, which provides the best model fit, indicates significant dynamic effects occur within both ON and OFF bouts of stimulation.

4.3. Additional analyses

To test the generality of our methods we fit the four models to an additional seven data sets (two additional data sets from the same human subject and five from a new monkey) summarized in Table 3 along with the two data sets presented above. The human data sets span the same cross-over period as the oral feeding data presented and therefore contain three bouts of stimulation. The monkey data sets range from 800 to 1700 trials with between two and eight bouts of stimulation. As assessed by the DIC the state-space model fit the data best in eight out of the nine total cases. For the data set (4), where the state-space model was not the best fit, the model LR ABC2 was superior, indicating that the almost linear-step-like structure imposed by that model was better able to describe the data. In addition to DIC in Table 3, we also show whether the two models LR ABC1 and SS indicate that stimulation has a positive or negative effect on the observed response in eight of the nine examples. (In the example 6, we did not consider this question because stimulation intensity was varied between stimulation bouts resulting in a dose dependent performance effect.) There is general agreement

between LR ABC1 and SS model conclusions in six cases. In two of these cases stimulation had a significantly negative effect on the responses. In the two remaining cases where one model indicated an effect and the other did not (3 and 4), the model selected based on lowest DIC indicates a significant positive effect of stimulation.

5. Discussion

We have presented four Bayesian models for the analysis of behavioral data from two DBS experiments. Our first model was a simple Bayesian generalized linear (logistic regression) model. The second two models included effects of the stimulation, assuming, respectively, that the stimulation effects were static and varied between stimulation periods. The fourth model was a state-space model. Overall, our results indicated that the final state-space model was a best fit to 8 out of 9 data sets considered even when the larger number of parameters used in the state-space formulation is taken into consideration.

We implemented all four models using the same Markov chain Monte Carlo software package (WinBugs; Lunn et al., 2000). This allowed for ease of implementation and model comparison, since the software automatically computes the DIC. These models could also be estimated using maximum-likelihood approaches with varying levels of computational complexity. Models LR AB, ABC1 and ABC2 can easily be computed using available generalized linear model (logistic regression) estimation software which computes the maximum-likelihood solution. Estimation of the state-space model can be made using the Expectation-Maximization algorithm (Dempster et al., 1977; Smith and Brown, 2003). This approach requires estimation of the expectation of the posterior distribution which can be done using the Kalman filter and smoothing algorithms (Dempster et al., 1977; Nalatore et al., 2009; Smith and Brown, 2003; Smith et al., 2004, 2005). Even though these methods may rely on different model assumptions (e.g., Gaussian posterior) we have found they yield comparable results (Smith et al., 2007).

For each model we compared the estimated performance between respective experimental data points on a trial-by-trial (day-by-day) basis for the monkey (human) data sets. This is straightforward to do using Monte Carlo Markov chain methods and is theoretically justified since the comparisons take into account covariance in the model between the trials (or days) being compared. In addition to allowing comparison between trials (or days), it is also possible to compare blocks of time, such as all the OFF periods against all the ON periods. In contrast these computations would require lengthy calculation of the covariance matrix in the typical non-Bayesian logistic regression framework. This approach allowed us to ask both global questions (e.g., did the stimulation have an overall effect on performance?) often answered by off-the-

shelf statistical techniques and more detailed questions about local time effects.

It is instructive at this point to consider the implications of the model selection (DIC) results. For the majority of data sets, the state-space model (which had very few assumptions about the structure of the data) was assessed to be the best fit. Because of this generality, the estimated 95% credible intervals are broader than those for the LR models, resulting in more conservative estimates of the overall effect of stimulation. Nevertheless, within their common domain of applicability (i.e., whether there was an effect of stimulation), the two approaches lead to similar conclusions.

An important advantage of the LR ABC2 and state-space models is that these models do not make the assumption that all the stimulation effects are the same. This makes it possible to assess effects on behavior produced by varying stimulation duration or changing electrode placement and current intensity over the course of the experiment.

As reviewed below, current knowledge of the effects of DBS suggests that the behavioral response will have complex dynamics. We emphasize that the goal of the state-space approach is to measure the behavioral response with unknown dynamics; it is not intended to be interpreted as a mechanistic account of them. Specifically, the state-space approach rigorously distinguishes between chance fluctuations in performance that occur even when the probabilities of the behavioral outcomes are constant, and fluctuations in performance that are due to changes in the underlying probabilities of the behaviors.

There are many reasons that these underlying probabilities in our application may have complex dynamics. At the level of physiologic mechanism, the central lateral nucleus of the thalamus is positioned to play a key role in arousal regulation (reviewed in Schiff, 2008). Arousal regulation during behavioral performance depends importantly on descending inputs to both the central thalamus and the brainstem/basal forebrain arousal systems originating from medial frontal cortical systems (Nagai et al., 2004; Paus et al., 1997). These medial frontal cortical regions and the striatum are the among the primary targets for activation with central thalamic DBS (Schiff and Purpura, 2002) and effective stimulation contacts show activation of medial frontal regions with central thalamic DBS in human studies (Schiff et al., 2007). We anticipate that continuous DBS in the central lateral nucleus may maintain a level of depolarization of neurons in cortical and striatal targets that resists a variety of sources of reduction of excitation within these structures as performance wears on (for example increasing satiety reducing motivational signals, or boredom reducing attentional effort). We interpret the observed effects of DBS in both the monkey and human datasets analyzed above under the same physiological model. The additional depolarization of cortical and striatal neurons may assist weak signaling as performance degrades (or remain poor due to underlying brain injuries), shifting the internal arousal state to a level which has a higher probability for the completion of the behaviors. As this signal is not informative for the tasks and its effect is expected to be limited to adjustment of the background state, a variety of complex internal dynamics will continue to impact the likelihood of correct performance.

Electrical stimulation of the central thalamus is among the most recent applications of DBS for the treatment of human brain disorders (reviewed in Schiff and Fins, 2007). DBS has a long antecedent history beginning with experimental studies of the 'ascending reticular activating system' (Moruzzi and Magoun, 1949) and earlier experimental demonstrations of behavioral facilitation in monkeys (Fuster, 1958; Fuster and Uyeda, 1962; Stamm and Rosen, 1972). Quantization of effects in earlier studies relied on measurement of reaction times and percentages of correct trial performance using parametric explorations of current intensity at different sites of stimulation. In these studies, the loss of facilitatory effects on performance correlated with visible behavioral changes (twitches, head turning, etc.) that likely reflected spreading of electrical current into adjacent neuronal populations. The methods presented here would allow on-line, continuous and detailed assessment of the impact of current stimulation parameters on performance. For the 6-month human trial data, real-time (i.e., daily) assessment of performance is possible since the model estimation takes less than 10 min on 2.39 GHz desktop computer. For the longer binary data sets, CPU time is close to 40 min, precluding continuous analysis during a typical 2 h session, but still allowing analysis on a day-to-day basis. For continuous real-time analysis, one might in future consider using faster methods such as the Expectation-Maximization algorithm (Smith and Brown, 2003), particle filtering (Ergun et al., 2007) or stochastic point process filters (Eden et al., 2004). These methods offer a powerful strategy to select periods of behavioral facilitation for evaluation of physiological data obtained during effective stimulation for comparison against periods of either ineffective stimulation or when the stimulator is OFF.

The effects produced by DBS will change as stimulation parameters are varied. However, as demonstrated in both clinical and experimental studies the same set of stimulation parameters may produce different behavioral effects over time adding another important dynamical component that must tracked over the course of DBS application. The initial design of the human clinical trial of central thalamic DBS anticipated that the effects of DBS would be rapidly reversible, making a cross-over design ideal (Schiff et al., 2007). But, cross-over studies made over the course of days and months may be influenced by long and short-term carry-over effects. In the human trial, if the effects of DBS had been completely irreversible, achieving maximum effect after just a short period of stimulation without decay, the crossover design would have been completely misleading and falsely negative due to long and short-term carry-over effects. The intermediate result obtained with evidence of decay recommends further use of cross-over designs but with the added caution that statistical tools be available to identify and quantify carry-over effects that may endure. Rodent studies of central thalamic DBS also show both acute dynamic effects and slower carry-over effects, with a persistence of DBS effects after discontinuation of stimulation (Shirvalkar et al., 2006). For central thalamic stimulation which demonstrates these effects, and acts on a main route of neocortical activation associated with attention, learning and memory, it is anticipated that a mix of time courses of response to DBS will be typical and will require new methodologies as presented here to quantitatively assess the causal relationships of DBS and behavioral changes over multiple timescales. Similarly, human trials of DBS in neuropsychiatric disorders such as depression and obsessive compulsive disorder have also identified multiple acute and subacute effects of DBS (Mayberg et al., 2005). Having the ability to precisely measure the evolving dynamics of treatment may provide important further insight into mechanisms underlying treatment with DBS.

In animal experiments, state changes may also effect brief periods of DBS testing as an anticipated decline in performance may mix effects of DBS with waning vigilance, motivation, increasing boredom and fatigue (as seen above). The methods shown here can be used to model these conditions. Moreover, if behavioral facilitation is observed acutely in animal experiments it may presage a shifting of baseline performance over days, weeks, and, in the case of non-human primate experiments, months, of continued exposure to stimulation in effective sites. The methods described here would allow for carefully assessing these possibilities and incorporating this important information into the interpretation of DBS results. Precise tracking of the dynamics of quantitative measurements of behavior of will be of particular importance in linking these observations to quantitative physiological measurements in experimental studies of the underlying mechanisms of DBS effects.

More generally, it may be possible to develop improved generalized linear (or other) models by incorporating covariates that represent additional features revealed by state-space analyses. Such covariates (e.g., a ramp up of performance during simulation, or an interaction of DBS effects and waning vigilance) may give a better model fit and more precise credible intervals.

Acknowledgements

References

Brooks SP, Gelman A. General methods for monitoring convergence of iterative simulations. Journal of Computational and Graphical Statistics 1998;7:434–55.

Congdon P. Applied Bayesian Modelling. Chichester, England: Wiley; 2003.

Dempster AP, Laird NM, Rubin DB. Maximum Likelihood from Incomplete Data Via Em Algorithm. Journal of the Royal Statistical Society Series B—Methodological 1977;39:1–38.

Eden UT, Frank LM, Solo V, Brown EN. Dynamic analyses of neural encoding by point process adaptive filtering. Neural Computation 2004;16:971–98.

Ergun A, Barbieri R, Eden UT, Wilson MA, Brown EN. Construction of point process adaptive filter algorithms for neural systems using sequential Monte Carlo methods. IEEE Transactions on Biomedical Engineering 2007;54:419–28.

Fahrmeir L, Tutz G. Multivariate statistical modelling based on generalized linear models. 2nd ed. New York: Springer; 2001.

Frank MJ, Samanta J, Moustafa AA, Sherman SJ. Hold your horses: impulsivity, deep brain stimulation, and medication in parkinsonism. Science 2007;318:1309–12.

Fuster JM. Effects of stimulation of the brain stem on taschistoscopic perception. Science 1958;127.

Fuster JM, Uyeda AA. Facilitation of tachistoscopic performance by stimulation of midbrain tegmental points in the monkey. Experimental Neurology 1962;6:384–406.

Gelman A, Rubin DB. Inference from iterative simulation using multiple sequences (with discussion). Statistical Science 1992;7:457–511.

Giacino JT, Ashwal S, Childs N, Cranford R, Jennett B, Katz DI, et al. The minimally conscious state—definition and diagnostic criteria. Neurology 2002;58:349–53.

Greenberg BD, Malone DA, Friehs GM, Rezai AR, Kubu CS, Malloy PF, et al. Three-year outcomes in deep brain stimulation for highly resistant obsessive-compulsive disorder. Neuropsychopharmacology 2006;31:2384–93.

Kass RE, Carlin BP, Gelman A, Neal RM. Markov chain Monte Carlo in practice: a roundtable discussion. American Statistician 1998;52:93–100.

Kitagawa G, Gersch W. Smoothness priors analysis of time series. New York: Springer; 1996.

Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. Statistics and Computing 2000;10:325–37.

Mayberg HS, Lozano AM, Voon V, McNeely HE, Seminowicz D, Hamani C, et al. Deep brain stimulation for treatment-resistant depression. Neuron 2005;45:651–60.

Moruzzi G, Magoun HW. Brain stem reticular formation and activation of the EEG. Electroencephalography and Clinical Neurophysiology 1949;1:455–73.

Murphy K, Mahdaviani M. Matbugs.m.; 2005. http://www.cs.ubc.ca/~murphyk/Software/MATBUGS/matbugs.html.

Nagai Y, Critchley HD, Featherstone E, Fenwick PBC, Trimble MR, Dolan RJ. Brain activity relating to the contingent negative variation: an fMRI investigation. Neuroimage 2004;21:1232–41.

Nakamura K, Hikosaka O. Role of dopamine in the primate caudate nucleus in reward modulation of saccades. Journal of Neuroscience 2006;26:5360–9.

Nalatore H, Ding M, Rangarajan G. Denoising neural data with state-space smoothing: method and application. Journal of Neuroscience Methods 2009;179:131–41.

Nelder JA, Wedderburn RW. Generalized linear models. Journal of the Royal Statistical Society Series A 1972;135:370–84.

Page D, Jahanshahi M. Deep brain stimulation of the subthalamic nucleus improves set shifting but does not affect dual task performance in Parkinson's disease. IEEE Transactions on Neural Systems and Rehabilitation Engineering 2007;15:198–206.

Parasuraman R, Davies DR. A taxonomic analysis of vigilance. In: Mackie RR, editor. Vigilance: Theory, operational performance, and physiological correlates. New York: Plenum; 1977. p. 559–74.

Paus T, Zatorre RJ, Hofle N, Caramanos Z, Gotman J, Petrides M, et al. Time-related changes in neural systems underlying attention and arousal during the performance of an auditory vigilance task. Journal of Cognitive Neuroscience 1997;9:392–408.

Paxinos G, Huang XF, Toga AW. The rhesus monkey brain in stereotaxic coordinates. San Diego: Academic Press; 2000.

Perlmutter JS, Mink JW. Deep brain stimulation. Annual Review of Neuroscience 2006;29:229–57.

Robert CP, Casella G. Monte Carlo statistical methods. 2nd ed. New York: Springer; 2004.

Schiff ND. Central thalamic contributions to arousal regulation and neurological disorders of consciousness. Molecular and Biophysical Mechanisms of Arousal, Alertness, and Attention 2008;1129:105–18.

Schiff ND, Fins JJ. Disorders of consciousness. Mayo Clinic Proceedings 2007;82:250–1.

Schiff ND, Giacino JT, Kalmar K, Victor JD, Baker K, Gerber M, et al. Behavioural improvements with thalamic stimulation after severe traumatic brain injury. Nature 2007;448:600-U10.

Schiff ND, Hudson AE, Kalik SF, Purpura KP. Modeling wakeful unresponsiveness: characterization and microstimulation of the central thalamus. In: Society for neuroscience 32th annual meeting 62.12; 2002.

Schiff ND, Kalik SF, Purpura KP. Sustained activity in the central thalamus and extrastriate areas during attentive visuomotor behavior: correlation of single unit activity and local field potentials. In: Society for neuroscience 31th annual meeting; 2001.

Schiff ND, Purpura KP. Towards a neurophysiological basis for cognitive neuromodulation through deep brain stimulation. Thalamus and Related Systems 2002;2:55–69.

Shirvalkar P, Seth M, Schiff ND, Herrera DG. Cognitive enhancement with central thalamic electrical stimulation. Proceedings of the National Academy of Sciences of the United States of America 2006;103:17007–12.

Smith AC, Brown EN. Estimating a state-space model from point process observations. Neural Computation 2003;15:965–91.

Smith AC, Frank LM, Wirth S, Yanike M, Hu D, Kubota Y, et al. Dynamic analysis of learning in behavioral experiments. Journal of Neuroscience 2004;24:447–61.

Smith AC, Stefani MR, Moghaddam B, Brown EN. Analysis and design of behavioral experiments to characterize population learning. Journal of Neurophysiology 2005;93:1776–92.

Smith AC, Wirth S, Suzuki WA, Brown EN. Bayesian analysis of interleaved learning and response bias in behavioral experiments. Journal of Neurophysiology 2007;97:2516–24.

Spiegelhalter DJ, Best NG, Carlin BR, van der Linde A. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society B 2002;64:583–616.

Stamm JS, Rosen SC. Cortical steady state potential shifts and anodal polarization. Acta Neurobiologiae Experimentalis (Warsz) 1972;32:193–209.

Steriade M, Dossi RC, Contreras D. Electrophysiological properties of intralaminar thalamocortical cells discharging rhythmic (approximate-to 40 Hz) Spike-Bursts at approximate-to 1000 Hz during waking and rapid eye-movement sleep. Neuroscience 1993;56:1–9.

Volkmann J, Moro E, Pahwa R. Basic algorithms for the programming of deep brain stimulation in Parkinson's disease. Movement Disorders 2006;21:S284–9.

Williams ZM, Eskandar EN. Selective enhancement of associative learning by microstimulation of the anterior caudate. Nature Neuroscience 2006;9:562–8.

Wirth S, Yanike M, Frank LM, Smith AC, Brown EN, Suzuki WA. Single neurons in the monkey hippocampus and learning of new associations. Science 2003;300:1578–81.