◆ Human Brain Mapping 28:1033–1044 (2007) ◆

Real-Time fMRI Using Brain-State Classification

Stephen M. LaConte,* Scott J. Peltier, and Xiaoping P. Hu

Department of Biomedical Engineering, Georgia Institute of Technology, Emory University, Atlanta, Georgia

Abstract: We have implemented a real-time functional magnetic resonance imaging system based on multivariate classification. This approach is distinctly different from spatially localized real-time implementations, since it does not require prior assumptions about functional localization and individual performance strategies, and has the ability to provide feedback based on intuitive translations of brain state rather than localized fluctuations. Thus this approach provides the capability for a new class of experimental designs in which real-time feedback control of the stimulus is possible—rather than using a fixed paradigm, experiments can adaptively evolve as subjects receive brain-state feedback. In this report, we describe our implementation and characterize its performance capabilities. We observed ~80% classification accuracy using whole brain, block-design, motor data. Within both left and right motor task conditions, important differences exist between the initial transient period produced by task switching (changing between rapid left or right index finger button presses) and the subsequent stable period during sustained activity. Further analysis revealed that very high accuracy is achievable during stable task periods, and that the responsiveness of the classifier to changes in task condition can be much faster than signal time-to-peak rates. Finally, we demonstrate the versatility of this implementation with respect to behavioral task, suggesting that our results are applicable across a spectrum of cognitive domains. Beyond basic research, this technology can complement electroencephalography-based brain computer interface research, and has potential applications in the areas of biofeedback rehabilitation, lie detection, learning studies, virtual reality-based training, and enhanced conscious awareness. *Hum Brain Mapp 28:1033–1044, 2007.*

Key words: real-time fMRI; support vector machine; brain state; biofeedback

INTRODUCTION

The ability to predict brain states from short-time intervals of functional magnetic resonance imaging (fMRI) data reflects a fundamental correspondence between these data and the spatiotemporal activity of neuronal populations. It has been noted [Cox and Savoy, 2003] that this correspondence makes fMRI well suited for "brain reading" experiments, based on the modality's noninvasiveness and spatiotemporal qualities. While many fundamental methodological studies of fMRI classification exist [Kjems et al., 2002; Kustra and Strother, 2001; LaConte et al., 2003, 2005b; Martinez-Ramon et al., 2006; Mitchell et al., 2004; Mourao-Miranda et al., 2005; Shaw et al., 2003; Strother et al., 2002, 2004], much interest in predicting brain states and studying mental representations was catalyzed by the desire to evaluate the evidence for a localized versus distributed coding scheme for the (high-order) extrastriate visual cortex [Cox and Savoy, 2003; Downing et al., 2001; Hanson et al., 2004; Haxby et al., 2001; Ishai et al., 1999; Kanwisher et al., 1997; O'Toole et al., 2005]. Consequently,

Contract grant sponsor: Georgia Research Alliance; Contract grant sponsor: NIH; Contract grant numbers: RO1EB002009, R21NS050183-01.

*Correspondence to: Stephen M. LaConte, Emory University Hospital Education Annex, 531 Asbury Circle, Suite N305, Atlanta, GA 30322. E-mail: slaconte@bme.emory.edu

Received for publication 23 March 2006; Revised 8 July 2006; Accepted 24 July 2006

DOI: 10.1002/hbm.20326

Published online 28 November 2006 in Wiley InterScience (www.interscience.wiley.com).

there has been a remarkable surge in cognitive neuroscientific interest and inventive experimental designs focused on classification of brain states from fMRI data. The applications have been broad and include lie detection [Davatzikos et al., 2005], unconsciously perceived sensory stimuli [Haynes and Rees, 2005], behavioral choices in the context of emotional perception [Pessoa and Padmala, 2005], early visual areas [Kamitani and Tong, 2005], information-based mapping [Kriegeskorte et al., 2006], and memory recall [Polyn et al., 2005].

Simultaneously, continued advances in MR imaging systems and experimental sophistication with blood oxygenation level dependent (BOLD) [Ogawa et al., 1990a,b] imaging have led to the emergence of real-time fMRI as a viable tool for real-time biofeedback [deCharms et al., 2004, 2005; Posse et al., 2003; Weiskopf et al., 2003; Yoo et al., 2004; Yoo and Jolesz, 2002]. The most recent work of deCharms et al. [2005] represents a particularly compelling example of the utility of real-time fMRI for therapeutic applications. That study examined the impact of providing BOLD signal level changes in the rostral anterior cingulate cortex as feedback to affect conscious perception of pain. The deCharms study showed that when subjects increased (decreased) activity in this region, there was a corresponding increase (decrease) in pain perception, for a given pain stimulus. Such training was effective enough to lead chronic pain patients to report decreases in ongoing pain, even after completion of the experiment.

This paper describes a technological advance that merges brain-state prediction with real-time biofeedback. The prediction of brain states for feedback is fundamentally different from existing real-time fMRI implementations [deCharms et al., 2004, 2005; Posse et al., 2003; Weiskopf et al., 2003; Yoo et al., 2004; Yoo and Jolesz, 2002], which use time series fluctuations in localized brain regions to derive biofeedback signals. By explicitly using distributed brain-state patterns, our scientific perspective is that of brain reading rather than that of localized activation. The first advantage of this for real-time applications is that prior assumptions about functional localization and individual performance strategies are not required—the system learns these directly from the volunteer. This provides for a high degree of experimental flexibility across the spectrum of cognitive domains. The second advantage is that feedback can rely on a direct, intuitive translation of brain state, rather than a representation based on increasing or decreasing local activity. The potential benefit to fMRI research is quite high, as this approach provides the capability for a new class of experimental designs in which real-time feedback control of the stimulus is possible—rather than using a fixed paradigm, experiments can adaptively evolve as subjects receive brain-state feedback.

Within this manuscript, our aim is to describe our implementation, demonstrating the feasibility of using brain-state classification for real-time experiments, and to characterize the capabilities of this implementation. Our primary experimental task for this study consisted of rapid button press blocks that alternately used the left or right index finger. The left/right conditions were cued by a target that appeared on either the left or right portion of the visual display. During "training" runs, a support vector machine (SVM) classification model was trained, and an arrow in the center of the display pointed toward the left or right target (its orientation agreed with the cue). During a subsequent "testing" run, each acquired image volume was classified by the SVM model, and the arrow was no longer static during the condition—at each acquisition time, its position and orientation was updated based on the classifier's "left" or "right" decision. With additional subjects, we used the same display parameters, but changed the task instructions associated with the left and right conditions to examine mood, language, and imagined motor tasks. Our main conclusions are that (i) human-machine training can be accomplished in minutes, (ii) near-perfect prediction accuracy is attainable during sustained periods of activation, (iii) stimulus feedback can respond to changes in brain state much earlier than the time-to-peak limitations of the BOLD response, and (iv) this approach is flexible enough to accommodate a broad range of psychological tasks, while requiring no change in experimental procedures.

METHODS

Conceptually, our aims were relatively straightforward. We wanted the ability (1) to train a classification model based on early fMRI data (e.g., a training run; Fig. 1A), and thereafter (2) to use the classifier to predict brain state with each acquired image and (if desired) alter the stimulus based on this brain state (Fig. 1C). Our approach was to use the scanner's dedicated image reconstruction hardware for classification (both training and testing), and to transmit classification results to a stimulus display computer. The nature of the classification approach made these design choices reasonable; the major consideration was to avoid interfering with normal scanner function, but the computationally intensive aspect for classification is model training (which we perform after collecting all images in the run) and the real-time aspect (applying the classifier and sending and I/O signal) is computationally relatively inexpensive.

Real-Time Implementation

We modified the MR scanner's image reconstruction software to allow for brain-state training or testing during data collection (Fig. 1). We used the C-based SVMlight software [Joachims, 1999) for classification, modifying it to handle fMRI image data, and to avoid disk I/O during active data collection. The modified SVMlight was compiled into the Siemens' Image Calculation Environment to handle reconstructed data.

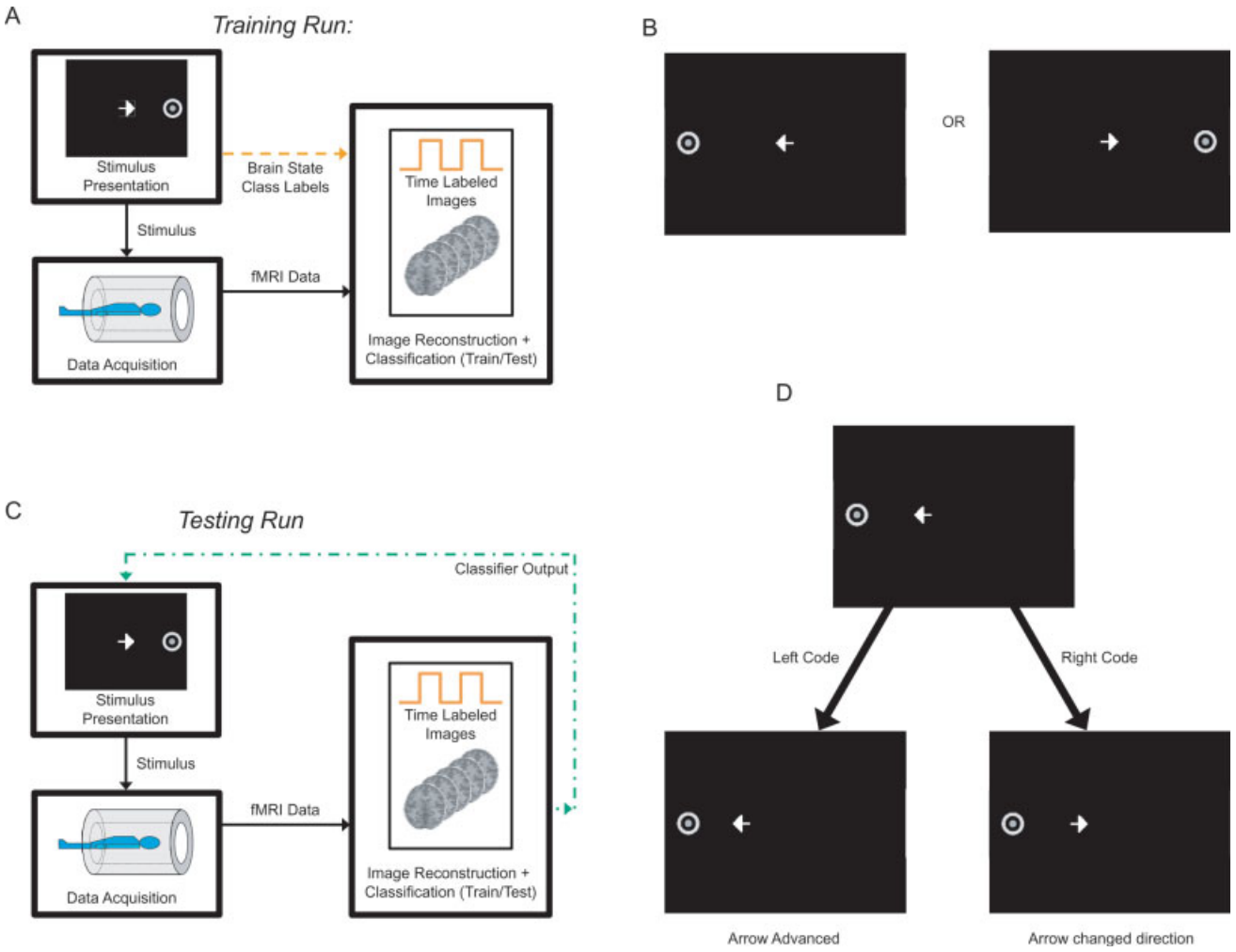

Figure 1.

Real-time brain-state imaging and voxel selection. Black arrows represent conventional fMRI: a stimulus is presented to a volunteer during image acquisition. (**A**) During training experiments (orange dashed line), the experimental condition (brain state) is used to label corresponding image times, and train a classifier on the scanner's image reconstruction hardware. (**B**) For training Runs 1 and 3 and for Run 2 (where the scanner was operated under testing mode, but no feedback was presented), volunteers were presented with the left button press condition or the right button press condition. For these runs, the arrow was always in the center of the visual field and oriented toward the target. (**C**) During testing with feedback (green dash-dot line), the brain state is not known, but is estimated from each image during image reconstruction. The volunteer's brain state, then, is used as feedback to control the stimulus. (**D**) For the 4th feedback-testing run, the stimulus target still alternated between left and right conditions. Shown is a specific example of a possible display update from one TR to the next. In this case, the goal is to move the arrow toward the left target. Given that the current display has the arrow directed to the left, a subsequent left code will advance the arrow toward the target. On the other hand, if a right code is sent from the MR scanner to the display computer, the arrow will change its direction. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Brain masking

As part of these train/test capabilities, we implemented an expedient means for brain/nonbrain segmentation (Fig. 2). Initially, our approach was simply to calculate an intensity threshold mask during training runs and use this same mask for subsequent testing runs. Using this approach with Subjects 1 and 2, however, we found that the eye regions, which are usually preserved in the threshold masks, were contributing to the training models. For all subsequent subjects, we removed eye regions by adding a short additional fMRI run. The basic principle of this additional run is to intentionally introduce variance in eye regions of the image, allowing the combination of an intensity-based mask with an additional variance-based mask. While we envision several future refinements, this current

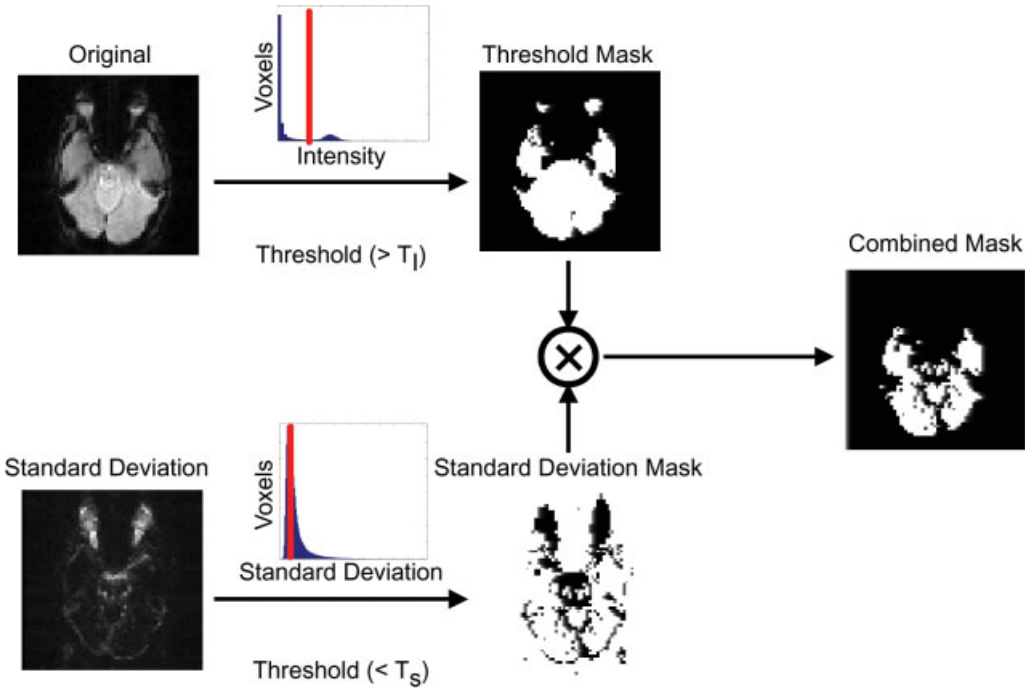

Figure 2.

Brain masking procedure to discard background, eyes, and tissue sources of high variance from fMRI data. A single image time volume is used to obtain an intensity threshold mask calculated from the first acquired image. In addition, the standard deviation of each voxel (using the entire run) is used to calculate the standard deviation mask. The final mask result is obtained by applying AND operator pixelwise to the two intermediate masks. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

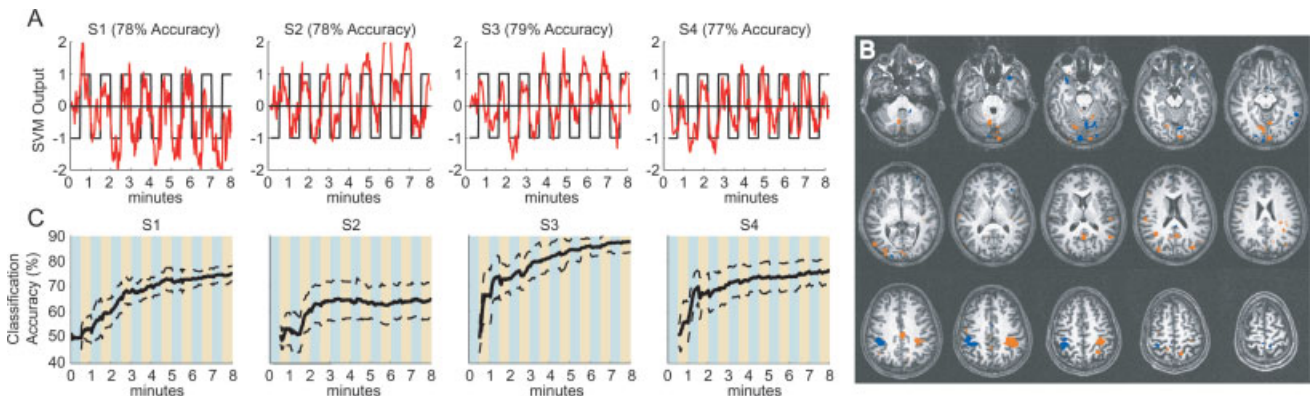

Figure 3.

Visually guided motor experiment. (**A**) Classifier output for the feedback testing (Run 4) for four subjects (labeled as S1–S4), demonstrating consistently accurate results across subjects. (**B**) SVM map indicating relevant spatial locations for discriminating between left and right conditions for Subject 3, training Run 3. Using radiographic convention, the right hemisphere is shown on the left. Positive model values are displayed in red and negative in blue (This corresponds with the training convention that the left/right tasks are assigned −1/+1 class labels, respectively). (**C**) Learning curves for the four subjects generated with successively longer increments of training data. Mean (solid) +/− standard deviation (dashed) results were generated using the 12 possible train-test permutations. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

approach is extremely simple and does not require a user interface. Volunteers were asked to focus their gaze on a fixation symbol that moved to locations on the display in a random order, updating at each image time sample. We used Presentation® (www.neurobs.com) to present a white fixation symbol on a black background at 31 locations covering the field of view (the back-projected display has an approximate visual field of 20° horizontally and 15° vertically). The order in which the symbol was displayed at a given location was randomized, and each location was used once. An intensity threshold, T_I, was used to create a binary mask from the first image volume. This mask consisted of setting all voxels below T_I to 0, and all voxels above to 1. The standard deviation threshold, T_s, was applied to a standard deviation image estimated from the entire 2-min run. In this case, all voxels below T_s were set to 1 (all above were set to 0). We used empirically derived threshold values that provided good results across several subjects. The combined mask, then, consisted of a voxel-wise multiplication (logical AND) of the threshold mask and the standard deviation mask.

In general, brain stem, eyes, ventricles, sagittal sinus, and peripheral areas are prevalent sources of unwanted variance that are efficiently removed with this approach. Immediately after the run, the mask image is available for visual inspection and limited modification. In practice, the number of slices and slice positioning should match the subsequent fMRI experiments in the imaging session.

Training and testing runs

For training runs, the brain mask and pre-assigned brain-state labels corresponding to the upcoming paradigm are read during scan preparation. During data acquisition, the mask of brain voxels is applied to the data, and these voxels and their corresponding class label are incorporated into appropriate software data structures on an ongoing basis as images are acquired. After scanning is complete, the data are used to train a linear SVM classifier. For test runs, a specified SVM model is read during scan preparation. During actual image acquisition (immediately after completion of reconstruction), images are applied to the SVM model and a serial I/O code is transmitted. This serial communication (RS-232) between the image reconstruction computer and the stimulus presentation computer (running Presentation) was used to control the display.

Brain-State Classification

We used the approach described by LaConte et al. [2005c] for SVM classification. Each imaged time volume was represented as a vector, $\mathbf{x_t}$, whose components were the intensity values for each brain voxel at that time, t. The experimental condition (behavioral state) associated with each $\mathbf{x_t}$ was represented as a scalar class label, y_t. For example, the left–right button press task (described later) was cast as a binary classification problem (left: $y_t = -1$ or right: $y_t = +1$).

The SVM algorithm attempts to find a linear decision boundary (separating hyperplane) using the decision function $D(\mathbf{x_t}) = (\mathbf{w} \cdot \mathbf{x_t}) + w_o$, where $\mathbf{w}$ defines the linear decision boundary. For all results reported, we used the linear, soft-margin SVM with parameter $C = 100$. As described in LaConte et al. [2005c], the soft margin SVM determines $\mathbf{w}$ by minimizing the sum $\frac{C}{T}\sum_{t=1}^{T} s_t + \frac{1}{2}\|\mathbf{w}\|^2$.

Each s is a "slack variable" representing the training error for observation t. The free parameter, C, controls the degree to which the training errors, averaged over all time points T ($1 \leq t \leq T$) penalize the minimization.

Once the SVM model is determined from the training images, independent testing images can be classified into left or right categories using the estimated decision function (classifier output) with $D(\mathbf{x_t}) < 0$ assigned to left and $D(\mathbf{x_t}) > 0$ assigned to right. Percent classification accuracy was reported by calculating

$$\frac{[\text{number of correctly classified scans}]}{[\text{total number of scans}]} \times 100.$$

Data Collection

Seven healthy males participated in this study after giving informed consent in accordance with Emory University's Institutional Review Board. Their ages ranged from 24 to 45 years, with a mean age of 32.6 years. These subjects were imaged on a 3 T Siemens Trio (Siemens, Germany). The fMRI runs used an echo planar imaging sequence (28 axial slices, TR/TE = 2,000/31 ms, voxel = $3.4 \times 3.4 \times 5$ mm^3). T_1 relaxation effects were negligible as the sequence automatically discards the first several scans.

To investigate the essential capabilities of this approach and allow for extensive offline analysis, we performed four runs of a block design experiment on four subjects. Table I describes the full imaging session for these four subjects. The visual display used a simple visual cue, consisting of an arrow and a target (Figs. 1B,D). The task required alternating 30-s periods of sustained rapid button presses on a fiber optic button box (Current Designs, www.curdes.com), using the left or right index finger, corresponding to the left or right position of the target (eight periods of each condition - 16 total). For Runs 1–3 (Fig. 1B), an arrow in the central visual field was oriented toward a target (located about 10° to the left or right of center). To the volunteer, Runs 1–3 were identical, although Run 2 was acquired with the reconstruction software running in testing mode (Runs 1 and 3 were training sessions). In Run 4, feedback in the form of updated arrow orientation and position was presented to the subject immediately after collecting each 2-s volume based on the classification of that time point as either left or right. This classification used the model generated from Run 3. Note that the visual feedback was updated using computer I/O from the image reconstruction computer to send either a

TABLE I. Experimental summary (Subjects 1–4)

Run no.	Image recon mode	Stimulus mode	Stimulus description
0	Brain mask	Fixation	Moving fixation/eye movement
1	Train SVM	No feedback	Arrow always points to target
2	Test SVM	No feedback	Arrow always points to target
3	Train SVM	No feedback	Arrow always points to target
4	Test SVM	Feedback	Arrow orientation and position updated each TR

fMRI sessions consisted of an intial run to generate a binary mask distinguishing brain from non-brain voxels and four experimental runs. For the volunteer, Runs 1–3 appeared identical, and provide additional data to compare with feedback in Run 4.

left or right code to the stimulus display computer. Thus the subjects observed the arrow advance toward the target for correctly classified images. Based on the brain-state code transmitted by the scanner to the paradigm display computer, the arrow either continued in the current orientation direction (position updates used a fixed step size of about $2/3^\circ$ out of the $\sim 20^\circ$ total horizontal visual field) or flipped its left–right orientation (Fig. 1D). After 30 s, the target position alternated and the arrow was recentered, pointing to the new target.

Three additional volunteers were used to examine the flexibility of the online implementation with respect to behavioral paradigm. These three also performed a training and testing feedback run using the button press task described (data not shown). After this, they performed another training and testing feedback run with another task (happy vs. sad, English vs. Mandarin, and effected vs. imagined motor).

Image Visualization and Processing

Visualization of data and SVM models was accomplished with Matlab (MathWorks, Natick, MA) and AFNI [Cox, 1996]. For some of the results, motion correction was performed using AFNI, to estimate and correct for rigid-body misalignment in specified runs. When this was done, the first image of the first run was used as the target volume for all images in all runs aligned. As previously reported [LaConte et al., 2005c], for the case of the linear SVM, the vector $\mathbf{w}$ has the same dimensionality and spatial correspondence as the image data $\mathbf{x}$. Thus, the model itself can be topographically mapped onto the brain. Since the SVM decision function is the dot product of the vector $\mathbf{w}$ and an image x, voxels that are highly correlated with the experimental conditions of the training data (e.g., the left/right task conditions) will tend to be reflected by higher absolute values in $\mathbf{w}$. In this sense, the maps obtained by $\mathbf{w}$ can bear similarities to conventional, mass univariate t-maps. There are important differences, though; the components of $\mathbf{w}$ (the values for every voxel) are obtained simultaneously and do not make the same distributional assumptions as the t-test.

RESULTS

Basic Characterization Using a Visually Guided Motor Experiment

The online classification results from the fourth run are shown in Figure 3A. Negative values of the classifier moved the arrow to the left, and positive values moved the arrow to the right. Across the four subjects, results were remarkably consistent with prediction accuracies of 78, 78, 79, and 77%, respectively. Figure 3B shows the training model for Subject 3 (Run 3), which reflects a visual-motor task. Concerning the feedback run, all subjects reported high confidence in their ability to control the arrow cursor movement and indicated that the feedback component made the button press task considerably more engaging.

An important design parameter for future studies is what constitutes a sufficient amount of training data. In offline analysis, resampling is possible across the four runs. That is, each of the four runs can be designated as either train or test, regardless of time order, allowing for 12 train-test permutations.[1] To examine the training issue, we generated learning curves (Fig. 3C); for each permutation, we incrementally trained with an increasing number of initial training run images and tested with the entire length of the testing run. Despite individual variations in terms of prediction accuracy and performance, these curves indicate that training with significantly shorter experimental runs is possible. We observe dramatic improvement occurring with 2 min of data and note that accuracy tends to asymptote by $\sim$4 min of training data (representing four repetitions of both left and right experimental conditions) for all subjects.

Classifier Drift

We found it necessary to detrend the classifier output in our data (Fig. 4). Primarily, we see a large offset of the

[1]We have initial data suggesting that these permutations may not be completely identical. Specifically we see a slight tendency for higher prediction accuracy in Run 4 (regardless of training run) compared to Runs 1–3 (across all training run permutations). If this is the case, though, our current four subject data set lacks the power to demonstrate it statistically. We have examined the issue of train-test asymmetries across runs in [LaConte, 2005c] and intend to perform further studies for the case here (of feedback runs vs. nonfeedback runs).

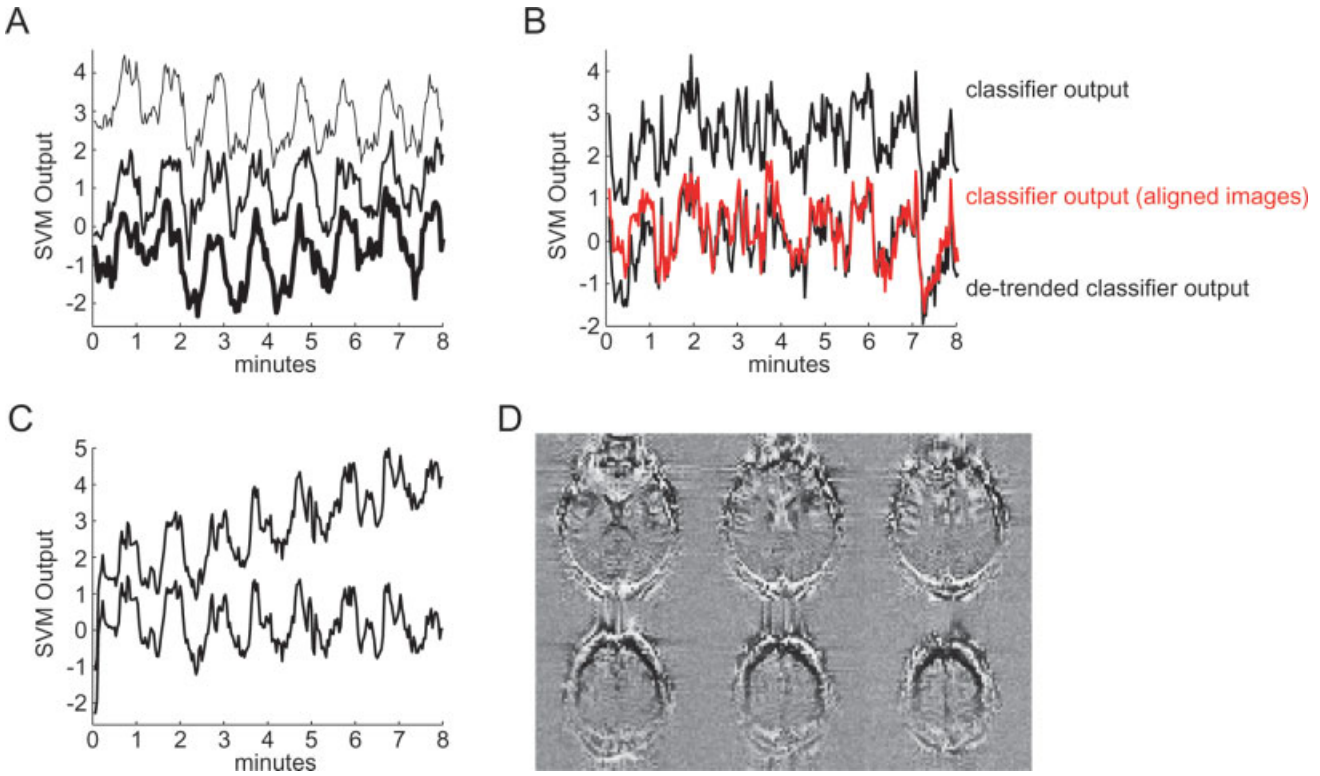

Figure 4.

Classifier drift. (**A**) Classifier output without detrending for Subject 3, Run 4, training with Run 1 (thin line), Run 2 (medium line), and Run 3 (thick line), exemplifying significant offset and slight drift. (**B**) Classifier output for Subject 1, Run 3, using Run 1 as training data. The red line represents results after aligning images in Runs 1 and 3 to the first image in Run 1. The two black lines are uncorrected data, with and without detrended classifier output. The good correspondence between the aligned result and the detrended classifier output suggests alignment as an important factor contributing to this effect. (**C**) Classifier output for Subject 4, Run 4 (trained on Run 3), with and without detrending. (**D**) Subtraction of the two correlation maps from Run 4 using the time courses in (C). This result demonstrates discrepancies at the edges of the slices, suggesting that the drift in the raw classifier output arises from motion. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

mean value of the classifier (normally, the classifier output should be centered at 0, again so that $D(\mathbf{x}_t) < 0$ is assigned to left and $D(\mathbf{x}_t) > 0$ is assigned right) and a slow linear drift within the run tested. This classifier drift was removed by computing the least-squares fit of a straight line to the data and subtracting the resulting function from the data. In the online setting, the offset and slope are continually re-estimated, classifying each new image acquisition and adjusting its value based on all accumulated, uncorrected classifier decision function results. Reported online results, in which the volunteers were given feedback, reflect this real-time detrending. For offline analyses, global linear trend removal (using the classifier output for all images in the run) was used to estimate and correct for classifier drift. The online detrending gives similar results to the global approach, but is less stable at the beginning of the run, particularly during the first experimental condition.

The results in Figure 4 examine the classifier drift. An example is shown in Figure 4A in which the same run is tested, using each of the other three runs as training. In Figure 4B, classifier output is generated both with and without motion correction preprocessing (Methods). Notice that the classifier output from the aligned data corresponds well to the nonaligned, but detrended result. The contribution of motion to classifier drift is further examined in Figures 4C,D. Using classifier output, both before and after detrending as reference time series (Fig. 4C), we generated correlation maps for the testing run. The subtraction of these two maps is shown in Figure 4D, showing differences primarily at the perimeter of the brain, further implicating motion as a key contributor to classifier drift.

Investigation of Hemodynamic Effects

Because the BOLD response relies on changes in the cerebral vasculature to indirectly measure neuronal activity, the fMRI signal exhibits a temporal delay. To examine the consequences of this limitation, we calculated the frequency of errors for the 16 total left and right test condi-

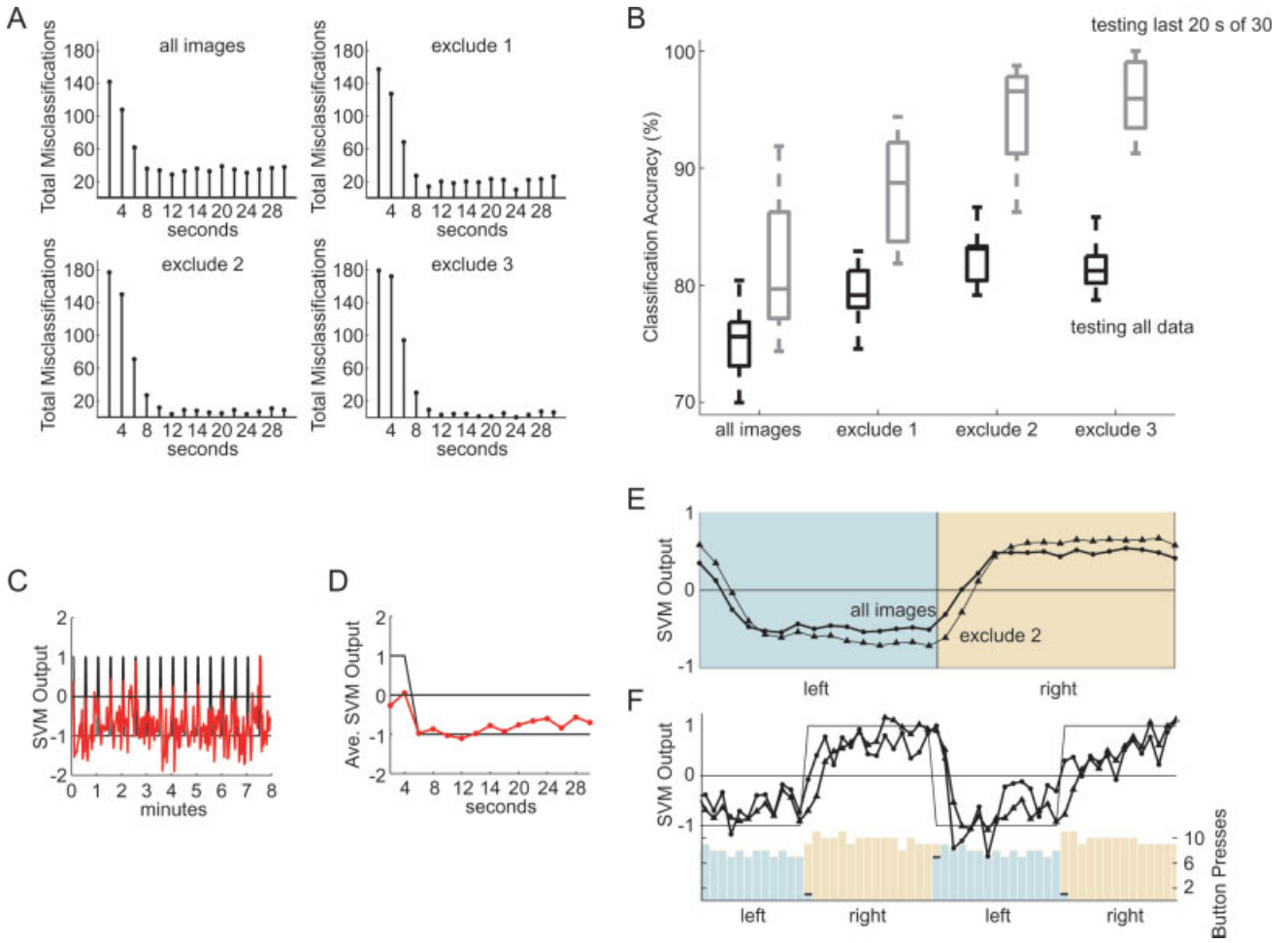

Figure 5.

Hemodynamic effects. (**A**) Number of classification errors across all 16 left and right conditions for all 12 train-test permutations for Subject 1 when training with all images and without the first one, two, or three transition images in each condition. (**B**) Classification accuracies for the 12 permutations of each case in (A), testing with all data in each run (black) and testing over the sustained activation of the last 20 s (10 images) in each experimental condition (gray). (**C**) Classification of transition (first 2 images for each 15 image condition) vs. "no transition" images (rather than left vs. right), obtained by aligning all four runs to the first image of the first run, using the first 3 runs as training data, and testing with Run 4. No correction was made for the unbalanced ratio (2:13) of training exemplars. (**D**) Time-locked average of (C). (**E**) Average classifier output for two of the training conditions in (A)—using all images (closed circle) and excluding two transition images per task condition (closed triangle). (**F**) Classifier output with behavioral data showing an unaveraged example of the effect in (E). Output from the model that is not trained with the first two transition images in each condition is less responsive, lagging behind output from the full model. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

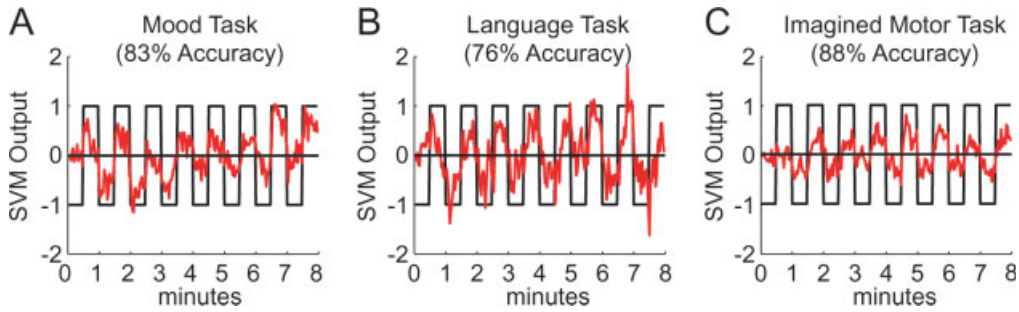

Figure 6.

Brain-state classification across a variety of cognitive domains. With the exact same experimental setup (different instructions), subjects can learn to move the arrow with very high accuracy with an 8-min training run, using (**A**) Mood (thinking happy vs. sad thoughts), (**B**) Language (bilingual subject progressing through a narrative, alternating between Mandarin and English), and (**C**) Imagined Motor (training with both the button press task and the imagined motor task, this subject transitioned to only the imagined motor task during the test run). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

tions across the 12 train-test permutations (leading to a maximum of 192 possible errors for each of the 15 time points for a given condition). The plots in Figure 5A were generated by training with all scans as well as omitting the first, second, and third transition images, respectively, during training (while testing on complete runs). These plots demonstrate a clear difference in the number of misclassifications for initial transition images compared to later "steady-state" images. The exclusion of transition images during the training stage leads to a notable improvement for nontransition periods, while elevating misclassification during the transition times. Thus one potential strategy for improving overall accuracy is to remove transition images. Though we believe the ramifications of this have not been previously studied, omitting transitions for both training and testing is common practice for offline studies [LaConte et al., 2003, 2005c; Mitchell et al., 2004]. Figure 5B shows overall prediction accuracy and accuracy during the last 20 s (10 images) of sustained activity for each condition for the four training strategies. The results for this subject indicate that removing 4 s (two transition images) is optimal, based on the overall prediction accuracy estimates and leads to a median prediction accuracy that is above 95% for the sustained portions of the task. It is important to note that the increased error rate for the transition images when they are excluded from the training set indicates that the original model was accounting for transition effects, even though these images were under-represented in terms of the number of exemplars and also likely would be better represented by a separate class structure.

Figures 5C,D demonstrates that it is actually possible to train a classifier to discriminate transition vs. nontransition images from these data. In this case, Runs 1–3 were aligned and combined as training data to test Run 4. The first two "transition" images in each condition were assigned to one class, while the remaining images were assigned to the second class. Although noisy, Figure 5C shows predictable spikes at the transition images. Figure 5D is the time-locked average of Figure 5C, demonstrating good average classification of steady-state images and clear separation from the transition images. The fact that we have not corrected for the gross mismatch in the number of samples between the two classes likely contributes to the bias of the transition class toward the steady-state class. The possibility of classifying between transition and nontransition scans is exciting, as this will allow for future online strategies that combine different class structures to achieve more accurate results.

Figure 5E shows the block-averaged classifier output derived from the training conditions of Figure 5A (training with all data and training without the first two transition images in each block, respectively). This average demonstrates a more responsive classifier when transition data are included in the model, evinced by the fact that the nontransition models tend to lag the full model outputs. We note that often considerable changes occur in the classifier output using the "transition aware" model during the condition switch—even though subjects' reaction time results in a delayed response to the abrupt transition, leading to a mixture of left–right button presses during the 2-s image acquisition (Fig. 5F).

Inherent Flexibility With Respect to Task

As a multivariate technique, our implementation matches the spatiotemporal nature of fMRI data, is computationally suited for real-time feedback, and its scalar-valued output directly relates to task condition (brain state), making it interpretable to both volunteers and experimenters (as opposed to activation patterns—especially subtle variations thereof—whose meaning is inaccessible to a lay person and often debated among specialists). One important aspect of our approach is that classification is performed without restriction to anatomical or functional regions of interest (ROIs). We examined the implications of this with three additional volunteers. The first is experienced in both fMRI and clinical practice, and, in this case, mood control was substituted for the button press task—for both a training run and a test run, the left target was used as a cue for the subject to think sad thoughts, while the right target cued for happy thoughts. This resulted in excellent control of the arrow movement during the test run (83% accuracy) and required no procedural change for the experimenters (Fig. 6A). Figure 6B represents a language task. This second volunteer progressed through a (silent) narrative recount of a recent travel experience, alternating between Mandarin and English periods. In Figure 6C, a third volunteer performed the same button press task as the first 4 subjects and additionally imagined left/right motor activities during the training run, while for the testing run, only the imagined motor component was performed.

As a final observation, note that in Figure 6C the amplitude of the classifier output is less than the plots of Figures 6A,B and 3A. We believe that this is related to the extent and distribution of activation relative to the training data. In this case, the training data is comprised of two simultaneous tasks (effected and imagined motor), while the test data only consisted of the imagined task. The training model in this case is still correlated with the test data, but it is likely that the test data fluctuations between the left and right conditions are less spatially distributed (resulting in a reduced amplitude for $D(x_t)$).

DISCUSSION

Real-Time Implementation

Our implementation uses SVM classification [Cherkassky and Mulier, 1998; Hastie et al., 2001; Joachims, 1999; Vapnik, 1995], without feature selection (all brain voxels are included). Further, all brain-state classification is performed on the scanner's image reconstruction hardware,

with direct I/O communication to a dedicated paradigm display computer. We consider the specific choice of classification algorithm to be a modular aspect of our software design, but chose SVM for several reasons. First, SVM has been prevalent in the recent brain-state fMRI literature [Cox and Savoy, 2003; Davatzikos et al., 2005; LaConte et al., 2005c; Mitchell et al., 2004]. In addition, our comparison of SVM with linear discriminant analysis [LaConte et al., 2005c] indicated that SVMs tend to be less sensitive to preprocessing issues, which is highly desirable for real-time applications. Finally, SVM implementations utilize convex optimization, which is computationally tractable and unique in its single optimal solution [Collobert et al., 2006] (as opposed to nonconvex algorithms such as back-propogation for neural networks). Indeed in practice, while training SVM models can be computationally intensive, this step is performed at the end of the training run, and thus does not interfere with image acquisition and reconstruction. For our system, there is usually no perceptible training delay associated with these experiments, and we have never observed training that would interrupt the natural progression of fMRI runs.

Another implementation issue is that of feature selection—a preprocessing step used to reduce data dimensionality, which often improves classification and computational performance. We have studied this issue in LaConte et al. [2005a], and have a basic mechanism to focus on ROIs via the mask file in our real-time setup. This enables explicit localization that is similar to other real-time fMRI systems, but still has the inherent advantage of the feedback signal being estimated from individual image times. While it is very likely that feature selection (in terms of anatomical ROIs) could improve prediction accuracy in some cases, we have already noted low error rates during steady-state portions of the BOLD response (Figs. 5A,B), and expect enhanced performance from future improvements and innovations. Further, in this report, we desire to stress the adaptive potential for future experiments, not only of the machine learning system, but also of the human volunteer. Such human adaptation could include factors such as context-sensitive processing, plasticity, or fatigue. Aggressive feature selection constitutes prior knowledge, which can be useful and even essential for specific hypotheses. In other cases, though, feature selection can introduce experimenter bias [Lange et al., 1999], imposing an obstacle to unexpected experimental findings. An important advantage of a multivariate approach is that both localized and distributed models are possible to the extent that the algorithm can weight the contribution of each variable (image voxel).

Finally, our approach is tightly integrated with our scanner's image reconstruction system. Because the major computational burden is in the training phase, with testing essentially requiring the calculation of a vector dot product, our implementation does not interfere with normal image reconstruction. By integrating our setup with the MR hardware, we can avoid the time delay caused by disk and network I/O associated with processing on a remote machine. Serial communication with the paradigm display computer consists of a simple transmission of scalar values (the classified brain state). This also provides for flexibility, both in terms of display set-up as well as providing for future experimental flexibility in terms of alternative sensory feedback modalities, and even affected output to devices (e.g. controlling a robot [Taylor et al., 2003]).

Temporal Limitations of Feedback

Ultimately, the hemodynamic delay associated with fMRI restricts capabilities for instantly responding to changes in brain state. Exactly how far this limitation can be pushed, however, is still an open issue. Our results suggest that discarding transition data does provide a cleaner training set for steady-state scans, but leads to a loss in information. The results of Figure 5A arise as a direct consequence of this well-known hemodynamic inertia. Generally, BOLD signal changes take 6–12 s to reach maximum intensity, and can remain relatively constant for sustained periods of activity. Cessation of tasks requires ~8–20 s to return to baseline signal levels [Chen et al., 1998; Kollias et al., 2000]. However, early but weak signal changes have been reported to occur roughly 0.5–2 s after the onset of neuronal activity [Kollias et al., 2000; Yacoub and Hu, 1999]. In addition, it has been reported that certain, limited experimental conditions can allow for the fMRI signal to track neuronal interactions at the millisecond time scale [Ogawa et al., 2000].

In the field of brain computer interface (BCI), bit rate is an important characteristic of the system [Wolpaw et al., 2000]. Increasing bit rate requires improved accuracy as well as faster brain-state switching capabilities, multiple classification levels (representing multiple bits), or both. We have only demonstrated two-class classification, but this does not reflect an inherent limitation in fMRI [Cox and Savoy, 2003]. In terms of task switching, a closer look at the transition issue indicates a picture much more encouraging than the necessity of a 6–12 s delay time. Although here we only present results from a single data set, we have consistently observed similar behavior in other data. Considering our results, the increased errors in classifying transition images when they are excluded from the training models in Figure 5A, the ability to train on transition vs. no-transition class labels (Figs. 5C,D), and the noted responsiveness of the full training model (Figs. 5E,F) all indicate that BOLD switching possesses reliable data structure suitable for classification techniques that occur much faster than does the time interval required to reach steady state. The exact nature of this structure needs further exploration, but one possibility is that distributed patterns may arise based on the mismatched rise and fall properties of the BOLD response. Further, such abrupt transitions in our task are somewhat artificial and represent a worst-case scenario. For example, if we were to modify the arrow task to be more analogous to natural,

goal-directed behavior (e.g. swimming laps), the individual would be able to plan for the upcoming target based on his/her current pace and prepare to change directions. Based on the subtle brain states that we and others have been able to detect, it is likely that such motor planning would provide adequate information content and lead time for correct classification of the volunteer's intent. Thus, the ultimate limit on data transfer is very much an unexplored matter and an exciting area for future research.

Another consideration is that brain-state classification can be done on a TR-by-TR basis. A running time series is not required, since similar data were observed during training. Note that localized fMRI approaches pay an analogous price to this training run, since an initial run is usually designated for functional localization.

Compared to electroencephalography (EEG) systems that allow for direct user control, our training times are extremely short. EEG systems primarily use endogenous electrical signals in specific frequency bands and usually require extensive training [Wolpaw et al., 2000]. Our data indicate that we can have a trained BCI after just several minutes of data collection, and it is likely that this can be done with a variety (and possibly multiple combinations) of cognitive tasks.

Flexibility of Cognitive Paradigms

We have demonstrated that real-time brain-state classification does not require a specific hypothesis about functional localization and individual performance strategies. This leads to experimental flexibility across the spectrum of cognitive domains. The primary results reported in Figures 3–5 arise from a simple experimental task that was primarily designed to provide robust activations and easy confirmation that the volunteers were following instructions (via recorded button presses). Based on our findings in Figure 6, though, we expect that experience in one domain will be relevant across a broad range of cognitive and behavioral tasks. Beyond the procedural convenience of task flexibility, another major advantage of brain reading is that feedback can rely on a direct, intuitive translation of brain state, rather than a display based on increasing or decreasing local activity. A subtle, but important methodological consideration from deCharms et al. [2005] was that subjects required cognitive strategy guidelines to be successful. No similar coaching is necessary with our brain-state approach.

CONCLUSION

We have demonstrated a flexible technological development for online classification, which adds the capability of providing adaptive feedback based on a subject's classified brain state. Conventional fMRI experiments provide a stimulus paradigm as "input" and measure the fMRI response as the brain's "output." Subsequent analyses then frequently rely on the assumption of a linear system. Feedback allows a new level of sophisticated exploration of brain function that goes beyond the input–output relationships of linear systems. For nonlinear systems identification and control, feedback is commonly utilized. Indeed, feedback provides the underpinning of all living systems, and, in the future, we are optimistic that such techniques will provide insights unattainable through traditional stimulus–response experiments. The design trade-offs we have considered, such as choosing the amount of training data, treatment of transition images, and selection of ROIs will vary based on experimental situation, but all of these factors will undoubtedly continue to improve with ongoing research in this area. Beyond adding flexibility to basic research experiments, this development can be extended to complement active research in EEG-based BCI. Other applications include biofeedback rehabilitation, lie detection, learning studies, virtual reality-based training, and enhanced conscious awareness.

ACKNOWLEDGMENTS

We thank G. Berns, V. Cherkassky, S. Hamann, S.-C. Ngan, G. Pagnoni, K. Sathian, W.K. Simmons, and S.C. Strother for their helpful advice.

REFERENCES

Chen W, Zhu XH, Kato T, Andersen P, Ugurbil K (1998): Spatial and temporal differentiation of fMRI BOLD response in primary visual cortex of human brain during sustained visual simulation. Magn Reson Med 39:520–527.

Cherkassky V, Mulier F (1998): Learning From Data: Concepts, Theory, and Methods. New York: Wiley.

Collobert R, Weston J, Bottou L (2006): Trading convexity for scalability. In: Proceedings of the 23rd International Conference on Machine Learning, Pittsburgh, PA.

Cox DD, Savoy RL (2003): Functional magnetic resonance imaging (fMRI) "brain reading:" detecting and classifying distributed patterns of fMRI activity in human visual cortex. Neuroimage 19:261–270.

Cox RW (1996): AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res 29:162–173.

Davatzikos C, Ruparel K, Fan Y, Shen DG, Acharyya M, Loughead JW, Gur RC, Langleben DD (2005): Classifying spatial patterns of brain activity with machine learning methods: Application to lie detection. Neuroimage 28:663–668.

deCharms RC, Christoff K, Glover GH, Pauly JM, Whitfield S, Gabrieli JDE (2004): Learned regulation of spatially localized brain activation using real-time fMRI. Neuroimage 21:436–443.

deCharms RC, Maeda F, Glover GH, Ludlow D, Pauly JM, Soneji D, Gabrieli JDE, Mackey SC (2005): Control over brain activation and pain learned by using real-time functional MRI. Proc Natl Acad Sci USA 102:18626–18631.

Downing P, Liu J, Kanwisher N (2001): Testing cognitive models of visual attention with fMRI and MEG. Neuropsychologia 39: 1329–1342.

Hanson SJ, Matsuka T, Haxby JV (2004): Combinatorial codes in ventral temporal lobe for object recognition: Haxby (2001) revisited: Is there a "face" area? Neuroimage 23:156–166.

Hastie T, Tibshirani R, Friedman J (2001): The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer-Verlag.

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001): Distributed and overlapping representations of faces and objects in ventral temporal cortex. Science 293:2425–2430.

Haynes JD, Rees G (2005): Predicting the stream of consciousness from activity in human visual cortex. Curr Biol 15:1301–1307.

Ishai A, Ungerleider LG, Martin A, Schouten HL, Haxby JV (1999): Distributed representation of objects in the human ventral visual pathway. Proc Natl Acad Sci USA 96:9379–9384.

Joachims T (1999): Making large-scale SVM learning practical. In: Scholkopf B, Burges C, Smola A, editors. Advances in Kernel Methods—Support Vector Learning. Cambridge, MA: MIT Press.

Kamitani Y, Tong F (2005): Decoding the visual and subjective contents of the human brain. Nat Neurosci 8:679–685.

Kanwisher N, McDermott J, Chun MM (1997): The fusiform face area: A module in human extrastriate cortex specialized for face perception. J Neurosci 17:4302–4311.

Kjems U, Hansen LK, Anderson J, Frutiger S, Muley S, Sidtis J, Rottenberg D, Strother SC (2002): The quantitative evaluation of functional neuroimaging experiments: Mutual information learning curves. Neuroimage 15:772–786.

Kollias SS, Golay X, Boesiger P, Valavanis A (2000): Dynamic characteristics of oxygenation-sensitive MRI signal in different temporal protocols for imaging human brain activity. Neuroradiology 42:591–601.

Kriegeskorte N, Goebel R, Bandettini P (2006): Information-based functional brain mapping. Proc Natl Acad Sci USA 103:3863–3868.

Kustra R, Strother S (2001): Penalized discriminant analysis of [O-15]-water PET brain images with prediction error selection of smoothness and regularization hyperparameters. IEEE Trans Med Imaging 20:376–387.

LaConte S, Anderson J, Muley S, Ashe J, Frutiger S, Rehm K, Hansen LK, Yacoub E, Hu X, Rottenberg D, Strother S (2003): The evaluation of preprocessing choices in single-subject BOLD fMRI using NPAIRS performance metrics. Neuroimage 18:10–27.

LaConte S, Chen J, Peltier S, Hu X (2005a): Discriminating one finger from another: Support vector classification of event related fMRI. In: Proceedings of the 13th Scientific Meeting of the International Society for Magnetic Resonance in Medicine, May 7–13, 2005, Miami. p 320.

LaConte S, Peltier S, Hu X (2005b): Real-time classification of brain states for fMRI experiments. Neuroimage 26, S42.

LaConte S, Strother S, Cherkassky V, Anderson J, Hu XP (2005c): Support vector machines for temporal classification of block design fMRI data. Neuroimage 26:317–329.

Lange N, Strother SC, Anderson JR, Nielsen FA, Holmes AP, Kolenda T, Savoy R, Hansen LK (1999): Plurality and resemblance in fMRI data analysis. Neruoimage 10:282–303.

Martinez-Ramon M, Koltchinskii V, Heileman G, Posse S (2006): fMRI pattern classification using neuroanatomically constrained boosting. Neuroimage 31:1129–1141.

Mitchell TM, Hutchinson R, Niculescu RS, Pereira F, Wang XR, Just M, Newman S (2004): Learning to decode cognitive states from brain images. Mach Learn 57:145–175.

Mourao-Miranda J, Bokde ALW, Born C, Hampel H, Stetter M (2005): Classifying brain states and determining the discriminating activation patterns: Support vector machine on functional MRI data. Neuroimage 28:980–995.

O'Toole AJ, Jiang F, Abdi H, Haxby JV (2005): Partially distributed representations of objects and faces in ventral temporal cortex. J Cogn Neurosci 17 580–590.

Ogawa S, Lee T-M, Kay AR, Tank DW (1990a): Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proc Natl Acad Sci USA 87:9868–9872.

Ogawa S, Lee T-M, Nayak AS, Glynn P (1990b): Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields. Magn Reson Med 14:68–78.

Ogawa S, Lee T-M, Stepnoski R, Chen W, Zhu X-H, Ugurbil K (2000): An approach to probe some neural systems interaction by functional MRI at neural time scale down to milliseconds. Proc Natl Acad Sci USA 97:11026–11031.

O'Toole AJ, Jiang F, Abdi H, Haxby JV (2005): Partially distributed representations of objects and faces in ventral temporal cortex. J Cogn Neurosci 17:580–590.

Pessoa L, Padmala S (2005): Quantitative prediction of perceptual decisions during near-threshold fear detection. Proc Natl Acad Sci USA 102:5612–5617.

Polyn SM, Natu VS, Cohen JD, Norman KA (2005): Category-specific cortical activity precedes retrieval during memory search. Science 310:1963–1966.

Posse S, Fitzgerald D, Gao KX, Habel U, Rosenberg D, Moore GJ, Schneider F (2003): Real-time fMRI of temporolimbic regions detects amygdala activation during single-trial self-induced sadness. Neuroimage 18:760–768.

Shaw ME, Strother SC, Gavrilescu M, Podzebenko K, Waites A, Watson J, Anderson J, Jackson G, Egan G (2003): Evaluating subject specific preprocessing choices in multisubject fMRI data sets using data-driven performance metrics. Neuroimage 19:988–1001.

Strother SC, Anderson J, Hansen LK, Kjems U, Kustra R, Sidtis J, Frutiger S, Muley S, LaConte S, Rottenberg D (2002): The quantitative evaluation of functional neuroimaging experiments: The NPAIRS data analysis framework. Neuroimage 15:747–771.

Strother S, LaConte S, Hansen LK, Anderson J, Zhang J, Pulapura S, Rottenberg D (2004): Optimizing the fMRI data-processing pipeline using prediction and reproducibility performance metrics. I. A preliminary group analysis. Neuroimage 23:S196–S207.

Taylor DM, Tillery SIH, Schwartz AB (2003): Information conveyed through brain-control: cursor versus robot. IEEE Trans Neural Syst Rehabil Eng 11:195–199.

Vapnik V (1995): The Nature of Statistical Learning Theory. New York: Springer-Verlag.

Weiskopf N, Veit R, Erb M, Mathiak K, Grodd W, Goebel R, Birbaumer N (2003): Physiological self-regulation of regional brain activity using real-time functional magnetic resonance imaging (fMRI): methodology and exemplary data. Neuroimage 19:577–586.

Wolpaw JR, Birbaumer N, Heetderks WJ, McFarland DJ, Peckham PH, Schalk G, Donchin E, Quatrano LA, Robinson CJ, Vaughan TM (2000): Brain-computer interface technology: A review of the first international meeting. IEEE Trans Rehabil Eng 8:164–173.

Yacoub E, Hu X (1999): Detection of the early negative response in fMRI at 1.5 tesla. Magn Reson Med 41:1088–1092.

Yoo SS, Jolesz FA (2002): Functional MRI for neurofeedback: Feasibility study on a hand motor task. Neuroreport 13:1377–1381.

Yoo SS, Fairneny T, Chen NK, Choo SE, Panych LP, Park HW, Lee SY, Jolesz FA (2004): Brain-computer interface using fMRI: Spatial navigation by thoughts. Neuroreport 15:1591–1595.