

Application of multivariate methods to an fMRI Brain-Computer Interface

Dissertation

zur Erlangung des Grades eines Doktors
der Naturwissenschaften

der Fakultät für Biologie
und
der Medizinischen Fakultät
der Eberhard-Karls-Universität Tübingen

vorgelegt

von

Sangkyun Lee
aus Gumi, South Korea

August – 2010

„Gedruckt mit Unterstützung des
Deutschen Akademischen Austauschdienstes“

Tag der mündlichen Prüfung:	10 – November –2010
Dekan der Fakultät für Biologie:	Prof. Dr. F. Schöffl
Dekan der Medizinischen Fakultät:	Prof. Dr. I. B. Autenrieth
1. Berichterstatter:	Prof. Dr. Niels Birbaumer
2. Berichterstatter:	Prof. Dr. Christoph Braun
Prüfungskommission:	Prof. Dr. Niels Birbaumer Prof. Dr. Christoph Braun Prof. Dr. Matthias Bethge Prof. Dr. Boris Kotchoubey PD Dr. Ute Strehl

I hereby declare that I have produced the work entitled: "Application of multivariate methods to an fMRI Brain-Computer Interface", submitted for the award of a doctorate, on my own (without external help), have used only the sources and aids indicated and have marked passages included from other works, whether verbatim or in content, as such. I swear upon oath that these statements are true and that I have not concealed anything. I am aware that making a false declaration under oath is punishable by a term of imprisonment of up to three years or by a fine.

Tübingen, 25.08.2010

Date

Signature

Table of contents

	page
1. Abstract	6
2. Synopsis	7
3. Personal contributions to papers and manuscripts	23
4. Papers and manuscripts	24
4. 1. Real-Time Regulation and Detection of Brain States from fMRI Signals	25
4. 2. Effective functional mapping of fMRI data with support-vector machines	54
4. 3. Detection of cerebral reorganization induced by real-time fMRI feedback training of insula activation: A multivariate investigation	65
4. 4. Real-time support vector classification and feedback of multiple emotional brain states	75
4. 5. Multivariate prediction of movement intention in the human fronto- parietal cortex	89
5. Acknowledgements	113

1. Abstract

The development of real-time functional Magnetic Resonance Imaging (rtfMRI) and the advance in computer technology allow us to acquire functional brain images and analyze them during an ongoing task. Studies with rtfMRI have shown that a healthy human participant can learn to self-regulate the activity of a single brain area. The regulation training is guided by the feedback signal (e.g., visual feedback), which reflects the blood-oxygen-level dependent (BOLD) signal of the target area. In these approaches univariate methods were used to generate feedback signals and further analysis. As univariate methods perform statistical tests on a single voxel independently, it is not considered how the target area interacts with other brain areas and how the interaction changes over the learning. In contrast, multivariate methods can determine the brain states from a combination of activity of multiple brain voxels/areas.

Based on these points, this dissertation is dedicated to develop a new multivariate pattern method based on the support vector machine to better understand spatial interactions of multiple brain areas. This method is used to analyze the changes of activation patterns in the whole brain induced by the self-regulation training in the right anterior insular cortex. In the second phase, the multivariate pattern analysis is used to build an fMRI Brain-Computer Interface (BCI) system by classifying the fMRI signals and providing visual feedback in real time. This system successfully classifies multiple discrete emotional states from the fMRI signal. In the last part, the multivariate pattern classifier is used to look over a potential BCI application by trying to find the brain area which is associated with movement intention. Through these approaches, it is demonstrated that the multivariate pattern analysis can be successfully used to improve the current fMRI-BCI and understand the brain changes induced by neurofeedback training.

2. Synopsis

2. 1. Introduction

With the development of the real-time functional magnetic resonance imaging (fMRI) and the improvement of other preprocessing techniques (Sitaram, Lee et al. 2011), several studies (Yoo and Jolesz 2002; Posse, Fitzgerald et al. 2003; Weiskopf, Veit et al. 2003; Yoo, Fairney et al. 2004; deCharms, Maeda et al. 2005; Caria, Veit et al. 2007; Rota, Sitaram et al. 2009) have demonstrated that human subjects using real-time fMRI feedback can learn voluntary self-regulation of localized brain regions: the amygdala (Posse, Fitzgerald et al. 2003), the anterior cingulate cortex (ACC) (Weiskopf, Veit et al. 2003), the anterior insular cortex (Caria, Veit et al. 2007; Caria, Sitaram et al. 2010; Ruiz, Lee et al. submitted), sensorimotor regions (Yoo and Jolesz 2002), cortical activations related to auditory attention (Yoo, O'Leary et al. 2006), right inferior frontal gyrus (IFG) associated with language processing (Rota, Sitaram et al. 2009). Recent studies (deCharms, Maeda et al. 2005; Rota, Sitaram et al. 2009; Caria, Sitaram et al. 2010; Ruiz, Lee et al. submitted) reported that the learning to regulate a circumscribed brain region can lead to specific behavioral consequences. DeCharms and colleagues (deCharms, Maeda et al. 2005) demonstrated that the self-regulation of rostral ACC was significantly associated with changes in the perception of pain. Rota et al. (2009) reported that experimental subjects, while they were trained to self-regulate IFG, significantly improved their accuracy for the identification of affective prosodic stimuli (but not for syntactic stimuli). In Caria et al. (2010), it was shown that learning of self-regulation of the anterior insula induced change in valence ratings of the aversive pictures (either an emotionally negative or a neutral picture). Ruiz et al. (submitted) showed that learned self-regulation led to changes in the perception of emotional faces in schizophrenic patients.

Standard neuroimaging experiments with fMRI use univariate methods where all the statistical tests are separately performed at each voxel. In contrast, multivariate methods can recognize spatial and temporal patterns of activity from multiple distributed voxels in the brain. Multivariate methods accumulate weak information available at multiple locations to jointly decode cognitive states although information at any single location cannot differentiate between the states (Haynes and Rees 2006). Recent studies (Mitchell, Hutchinson et al. 2003; Kamitani and Tong 2005; Haynes and Rees 2006; Haynes, Sakai et al. 2007;

Soon, Brass et al. 2008; Lee, Halder et al. 2010) have applied multivariate methods to increase sensitivity of fMRI analysis. Laconte and his colleagues (LaConte, Peltier et al. 2007) developed an fMRI BCI system by employing a multivariate pattern classification method called support vector machines (SVM). While most fMRI-BCI studies to-date have investigated self-regulation of brain activity at one or two region-of-interests (ROIs) using univariate analysis, multivariate methods allow for real-time feedback of a whole network of brain activity pertaining to a task.

2. 2. Research questions and direction of my PhD

Even though it has been demonstrated that healthy human participants and patients can volitionally regulate BOLD signals from a single target area, it still remains unknown how other regions of the brain respond during the regulation training of the target area and how they functionally interact. For instance, with the respect to the insular cortex regulation, emotional episodes were used as a cognitive strategy (Caria, Veit et al. 2007; Caria, Sitaram et al. 2010). However, these studies did not fully consider the interaction between the brain regions associated with emotion during regulation of the target area.

To better understand brain changes induced by the fMRI-BCI training, my PhD research focused on the cerebral reorganization during learning of self-regulation of the insular cortex by analyzing brain changes in the system level with multivariate pattern analysis (Lee, Ruiz et al. in press) and multivariate Granger causality modeling (Lee, Ruiz et al. in press; Ruiz, Lee et al. submitted) (**Chapter 2. 4**). For the multivariate pattern analysis of fMRI signal, I developed a novel algorithm called 'Effect Mapping' (EM) (**Chapter 2. 3**) (Lee, Halder et al. 2010) based on the support vector machine (SVM) (Vapnik 1998; Schölkopf, Burges et al. 1999; Schölkopf and Smola 2002).

In addition, I built an fMRI-BCI based on the multivariate pattern analysis by using the Effect-map and the SVM classification (**Chapter 2. 5**) (Sitaram, Lee et al. in press). This system was used to recognize multiple discrete emotional states, such as happiness, disgust, and sadness from fMRI signals, in healthy individuals instructed to recall emotionally salient episodes from their lives.

Lastly, I investigated a potential BCI application for rehabilitation of patients suffering from stroke and other movement disabilities. As the detection of brain activity related to movement intention could be used to activate neuroprostheses in patients with difficulties performing motor imagery, I examined whether a multivariate pattern classifier can decode movement intention and the type of movement (i.e., left or right) (**Chapter 2. 6**) (Ruiz, Lee et al. in preparation).

2. 3. Effect-Mapping based on support vector machines

As a multivariate pattern analysis method, support vector machines (SVMs) have been widely used for brain signal analysis and decoding (LaConte, Strother et al. 2005; Mourao-Miranda, Bokde et al. 2005; Mourao-Miranda, Reynaud et al. 2006; Haynes, Sakai et al. 2007; Mourao-Miranda, Friston et al. 2007; Soon, Brass et al. 2008). In a typical SVM analysis of fMRI signals, BOLD values from all brain voxels of each repetition time (TR) are contained in an M-dimensional (M: number of the brain voxels) input vector \mathbf{x} . SVM determines the output y from the input vector \mathbf{x} as follows:

$$y = \mathbf{w}^T \mathbf{x} + w_0$$

where the weight vector \mathbf{w} and the constant value w_0 , which are estimated by a SVM training algorithm from the training dataset, define a linear decision boundary, T is transpose of a vector.

Studies (LaConte, Strother et al. 2005; Mourao-Miranda, Bokde et al. 2005; Mourao-Miranda, Reynaud et al. 2006; Mourao-Miranda, Friston et al. 2007) so far have generated functional maps using the vector of weight values \mathbf{w} generated by the SVM classification process, or alternatively by mapping the correlation coefficient between the fMRI signal \mathbf{x} at each voxel and the output y determined by the SVM. However, these approaches are limited as they do not incorporate both the information involved in the SVM output of a brain state, namely, the BOLD activation at voxels and the degree of involvement of different voxels as indicated by their weight values. An important implication of the above point is that two different datasets of BOLD signals, presumably obtained from two different experiments, can potentially produce two identical hyperplanes (i.e., the identical \mathbf{w} and w_0) irrespective of their differences in data distribution. Yet, the two sets of

signal inputs could correspond to different functional maps. With this consideration, I proposed a new method called Effect Mapping (EM) based on SVM (Lee, Halder et al. 2010) for efficient multivariate analysis of fMRI data. Effect Mapping considers both the effect of the voxel activation to the output of a classifier, and the weight vector of the estimated SVM model. EM measures the effect of each voxel to the classifier output by computing Mutual Information (MI) between the voxel and the output (see Lee et al. 2010 for more detail).

To compare the capability of functional maps for detecting important spatial patterns of brain activity, brain voxels having the highest functional values in magnitude was selected from the functional maps. Then, the selected voxels in the data were used to assess classification performance in the data. In this comparison, EM showed similar performance to the other previous methods (see Lee et al. 2010 for more detail) while all the methods identified different sets of voxels as most informative. It might indicate that the performance in classification can be highly dominated by the voxels commonly selected. To better understand this result, in the second comparison, overlapping and non-overlapping activations from two maps were considered separately. Classification accuracies from overlapping areas of two competing functional maps and two non-overlapping areas of two competing functional maps were compared. This method of comparison gives insight into how well one functional map can identify discriminating voxels which the other functional map does not. In addition, this consideration is crucial when the competing methods identify different functional areas as most informative. In this comparison, the classification accuracy from overlapping area of two competing maps was similar to the classification accuracies of the voxels entirely selected from each map. However, in the performance comparison with non-overlapping area, EM outperforms the two previous methods (see Lee et al. 2010 for more detail).

2. 4. Multivariate analysis of Insula regulation

To investigate cerebral reorganization induced by real-time fMRI feedback training, I used the fMRI data from a previous experiment (Caria, Veit et al. 2007) from 6 healthy participants who underwent 5 training sessions of self-regulation of the anterior insular cortex (Caria, Veit et al. 2007). In this dataset, the first session

required participants to perform insula regulation without feedback. Subsequent feedback sessions consisted of regulation blocks and baseline blocks. During feedback sessions, normalized average BOLD signal from the right anterior insula was presented to the participants in real time in the regulation blocks as changing bars of a graphical thermometer (For more detail about experimental protocol and information about participants see Caria et al., 2007).

Spatial distribution of brain activity was investigated by applying EM in a group analysis. To observe the learning effect over the sessions with respect to the recruitment of distributed brain areas, it was assessed how many brain voxels are highly involved to discriminate regulation and baseline. It was measured by applying several thresholds in magnitude of effect values (EVs) and observing general characteristics of distribution of EVs. As the feedback training proceeds, more voxels acquire low EVs with a number of voxels keeping high EVs (see Fig. 2 in Lee et al. 2011). In a specified threshold, the session performed without feedback showed smaller clusters of voxels with higher and intermediate EVs spread over the whole brain. It can be interpreted as that a large number of brain areas involved in the task. However, in application of the same threshold in the feedback sessions many voxels having intermediate EVs disappeared, indicating probably the gradual disengagement of unnecessary connections (see Fig. 2 in Lee et al. 2011).

To extend the evidence, temporal interaction of ROIs selected from the E-maps was investigated. For this analysis, the Granger Causality (GC) model based on vector autoregressive models (VAR) was used (Goebel, Roebroeck et al. 2003). Particularly, after assessing the connectivity between ROIs, the Causal Density (CD; a measure of the number of connections between ROIs in a functional network (Seth 2005; Lee, Ruiz et al. accepted; Ruiz, Lee et al. submitted)) and the Averaged Connection Strength (ACS; Lee, Ruiz et al. accepted) were used (see Lee, Ruiz et al. accepted for more detail). They measure the extents of change of significant temporal interactions across sessions over the brain areas.

Session 1 (the session without feedback) shows the lowest CD of the functional network. As the learning proceeds with feedback, the causal density reaches a peak value in session 3 and decreases in the later sessions, indicating that the shape of temporal recruitment follows an 'inverted-U' curve, but the causal density in session 5 is still higher than in session 1 (see Fig. 4A in Lee et al. 2011 for more detail). An investigation of ACSs between ROIs indicates that the strength of temporal recruitment in some ROIs increases monotonically from session to session (see Fig. 4B in Lee et al. 2011 for more detail).

Multivariate pattern-based spatial analysis indicated that feedback training leads to more spatially focused recruitment of areas relevant for learning and emotions. Effective connectivity analysis revealed that initial training is associated with increase of network density, while further training "prunes" redundant connections but "strengthens" relevant connections. Based on the fundamental assumption that the brain operates as a system where spatially distributed brain areas interact with one another, this research shows evidence for brain reorganization at the system level during self-regulation training of the right anterior insula.

2. 5. An fMRI-Brain Computer Interface based on real-time spatial pattern classification

2. 5. 1. The methods

The real-time brain state classification system consists of image acquisition system and fMRI-BCI which performs image processing, brain state classification and visual feedback (see Fig.1). The image acquisition system uses the standard echo planar imaging (EPI) sequence with the modification, which allows functional image files to be stored in a specified directory in real time. The real-time sequence incorporated the following image acquisition parameters: repetition time (TR) = 1.5 s, echo time (TE) = 30 ms, matrix size = 64 x 64, 16 slices (voxel size = 3.3 x 3.3 x 5.0 mm³, slice gap = 1 mm). After acquisition of the functional brain images, they were transferred into another computer for the following processes. In the fMRI-BCI, Brain state classification is performed in the following steps (see Fig. 2): (1) signal preprocessing for online head-movement correction and spatial smoothing, (2) informative voxel selection by the method of Effect mapping (Lee,

Halder et al. 2010) resulting in the brain mask, (3) classifier retraining based on the brain mask obtained in step 2, and (4) real-time classifier testing on new data using the brain mask, (5) generation of visual feedback.

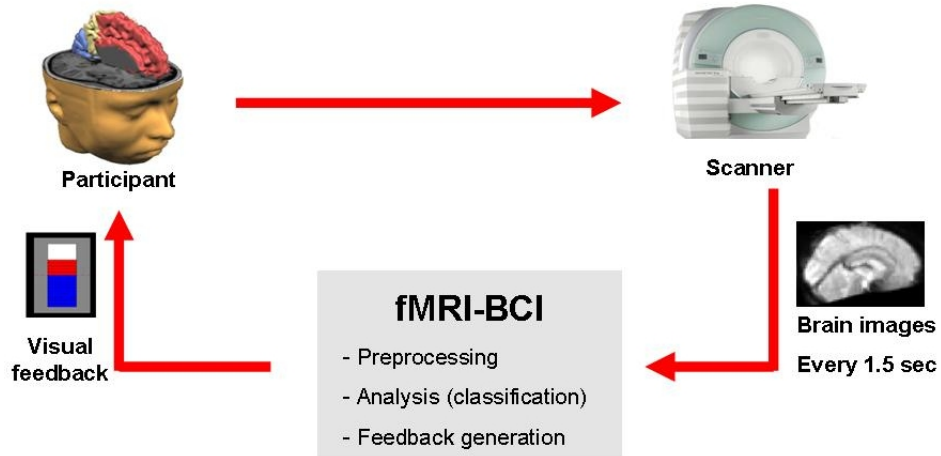


Fig. 1. The real-time fMRI brain state classification

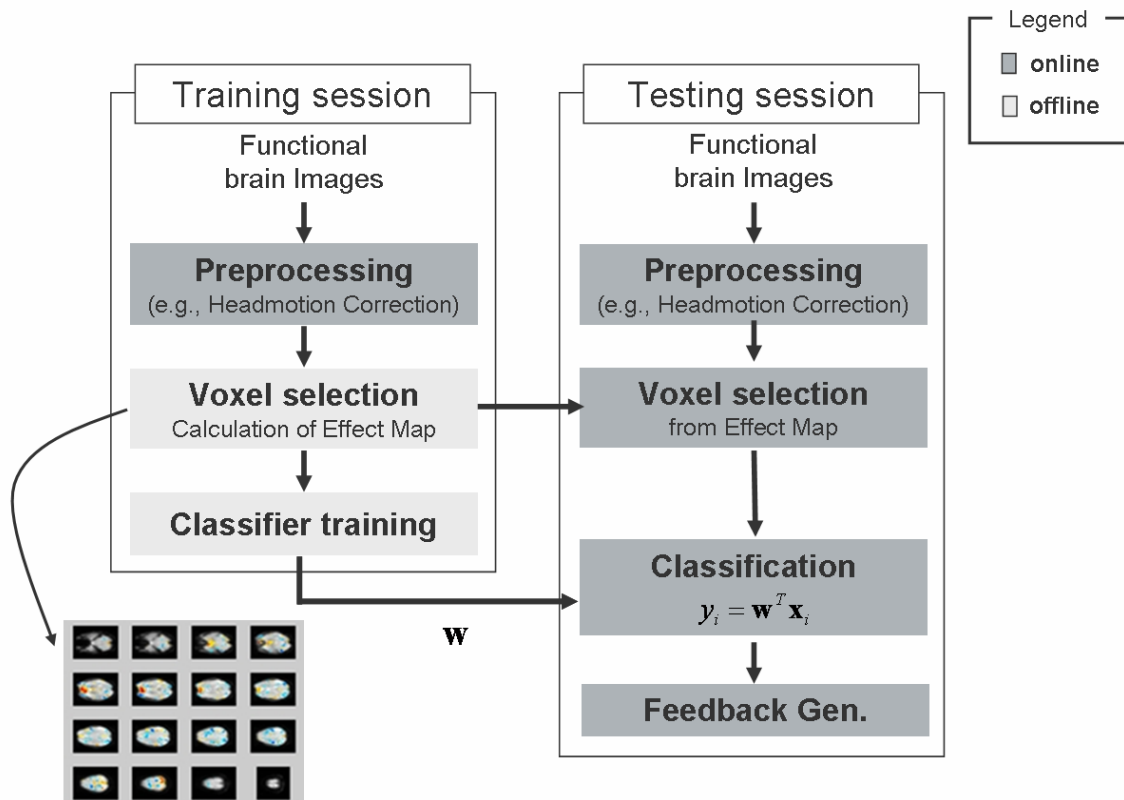


Fig. 2. Flow chart for fMRI signal preprocessing and classification.

2. 5. 2. Real-time classifications and feedback of multiple emotional brain states

The newly built fMRI-BCI system was applied to decode the brain state of a human during recalls of 3 different emotions (happiness, disgust, and sadness). The experiment includes three parts: experiment 1 for real-time binary classification of happiness and disgust, experiment 2 for multiclass classification of happiness, disgust, and sadness, and experiment 3 for assessing the effect of extended feedback training. Twelve healthy subjects participated in experiment 1, 4 subjects in experiment 2, and 2 subjects in experiment 3.

The experiment 1 included 4 succeeding stages: (1) data collection for classifier training, (2) training classifiers, (3) testing the classifier without feedback, and (4) testing the classifier with feedback. These two tests were performed to compare real-time classification performance in the absence and presence of feedback to evaluate the eventual application of real-time brain state decoding for BCI and clinical rehabilitation. This binary classification tests without and with feedback were successfully performed with $92\% \pm 6\%$ and $80\% \pm 13\%$, respectively (chance accuracy = 50%). In addition, 4 out of the 12 participants in experiment 1 participated in additional runs to test whether the classifier could robustly decode brain states even when participants do not use the same emotion inducing strategies that were used during training sessions. This classification test also showed high prediction accuracies without feedback ($80\% \pm 10\%$) and with feedback ($65\% \pm 18\%$).

The experiment 2 was designed to test real-time multiclass classification, both with and without feedback. In the experiment 2, three different emotion blocks (happiness vs. disgust vs. sadness) were alternating. Then, a 3-class classifier was used to decode brain states in the testing sessions. The classification accuracies in this test were $62\% \pm 14\%$ without feedback and $60\% \pm 16\%$ with feedback (chance accuracy=33%), indicating that multiple discrete emotion can be decode in the brain.

As shown in the experiment 1, one might argue why the performance with feedback is worse than without feedback. In the experiment 3, to assess whether feedback training helps participants to improve emotion recall, we recruited two more participants and trained them on two-class (happy vs. disgust) classification, for 3 sessions (in addition to 2 sessions for collection of training samples). To

adapt the classifier to the changes in the participant's brain activation induced by thermometer feedback and learned regulation, the classifier was retrained after each feedback session on the 2 latest runs of dataset in a style of moving window. In this experiment, both participants have gradual increasing performance (see Fig. 4C in Sitaram and Lee et al. 2010 for more detail).

To assess whether the classification accuracy was really determined by brain states induced by emotional recalls of participants (this is, to verify whether this classification was done with irrelevant information such as physiological confounds), an offline analysis is performed with the Effect-map (Lee et al., 2010) and further evaluation of classification accuracy of ROIs identified by the Effect-map. The analysis identified the medial frontal cortex (MFC), anterior cingulate cortex (ACC), insular cortex, superior temporal gyrus (STG), posterior cingulate cortex (PCC) and precuneus as the associated areas with the emotional task. This result is in line with the previous studies (Phan, Wager et al. 2002; Amodio and Frith 2006), and therefore support that participants performed emotion imagery and that the SVM results are reliable.

Based on these results it was demonstrated that emotional imagery can be successfully decoded in real-time, and an fMRI-BCI based on brain state classification can be used to provide visual feedback of emotional states for potential applications in the clinical treatment of dysfunctional affect.

2. 6. Decoding movement intention in the human brain

During scanning sessions, the experimental protocol was visually presented to the participants. The protocol consisted in an event-related design composed of 3 successive conditions, i.e., fixation, movement intention and imagery conditions. Duration of the fixation condition was integral multiples of 1 TR (TR=1.5sec). Movement-intention condition of 1TR was presented with a symbol indicating the type of the forthcoming motor imagery (i.e., left and right imagery). A participant was explicitly instructed not to imagine yet but to wait until the symbol indicating the onset of the imagery appeared on the screen. For an imagery condition of 3 TRs (4.5 seconds) a participant was asked to perform kinesthetic motor imagery of hand/arm movement (see Ruiz, Lee et al. in preparation for more detail).

To assess prediction accuracy of left and right conditions (i.e., left and right fixations, left and right intentions, and left and right imageries) in specific brain areas, brain masks of posterior parietal cortex (PPC), supplementary motor area (SMA), premotor cortex (PMC), and primary motor cortex (M1) were created. Here, fixation condition was artificially split into left and right fixation based on the following conditions (left or right intention) for the convenience of analysis. Then, classification accuracies in the areas across time points were measured through cross-validation of classification (leave-one-session-out cross-validation; see Ruiz, Lee et al. in preparation for more detail).

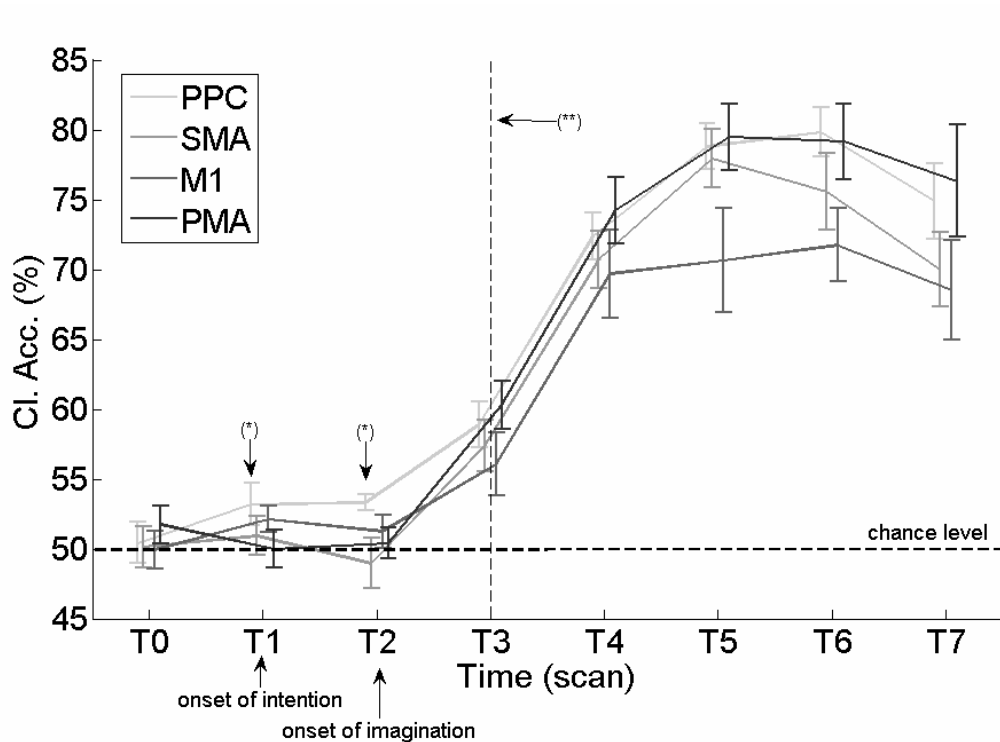


Fig. 3. Decoding accuracy of condition left vs. right hand/arm in the specified brain area (% mean accuracy \pm % standard error of the mean). At the point T0, classification accuracy is calculated from data obtained from left and right fixation. T1 and T2 indicate the onset of intention and imagery conditions. The (*) and (**) indicate the significant level $P < .01$ and $P < .0001$ in the one-sample t-test, respectively. Since T3 (the broken line), decoding accuracies are significant higher (with $P < .0001$) than the chance level in the one-sample t-test.

In this analysis, decoding accuracies in the classification of left and right conditions increased across time points (see Fig. 3). The performance in the PPC

is significantly different from the chance level at onset time of intention (T1; $P < .01$) and at onset time of imagination (T2; $P < .01$). With consideration of the hemodynamic delay (4.5s; 3 TR), classification accuracy at onset time of intention (T4) is greater than chance level for all regions of interest. In addition, the PPC, PMC and SMA show higher accuracies than M1 during movement intention and imagery.

This investigation showed that movement intention and the type of movement can be decoded in the PPC and premotor regions associated with motor planning. As movement intention was detectable before motor imagery in time, it might be feasible to predict the type of forthcoming motor execution in these areas. This finding suggests that the future BCI based on reading the movement intention from PPC can help patients who have frontal lesions or who have difficulties in performance of motor imagery.

2. 7. Discussion

In my PhD research, I proposed a new method for multivariate pattern analysis, called Effect Mapping, based on support vector machines. As this method incorporates both components (the weight vector and the input vector) to determine the output of SVM, and includes the consistency measure from distribution of the input vectors and the outputs of SVM, it could better identify informative voxels from the trained SVM. It was successfully applied in both online and offline analyses of fMRI signals.

An application of Effect Mapping in association with multivariate Granger causality model to the fMRI data obtained from the fMRI-BCI training provided an insight into how the brain reorganizes as self-regulation learning proceeds. Particularly, this study showed evidence for the functional cerebral reorganization induced by the fMRI-BCI learning for the first time. However, this analysis can at best cautious conclusion as the analysis could be performed only 6 participants, and the cognitive strategies used varied from person to person without any behavioral measurement. To better understand how the brain reorganizes functional activation and optimizes the use of resource with the respect to the energy consumption and the computation efficiency, future studies would need to

systematically control for confounding variables and incorporate monitoring of behavioral changes during self-regulation.

An additional part of my study involved the development of an online pattern classifier in an fMRI-BCI system. To construct a robust fMRI-BCI, my implementation contained preprocessing steps such as online head-movement correction, spatial smoothing, and voxel selection. However, as this system needs many sessions of fMRI data collection necessary to train classifiers, it set a practical limitation on its application to fMRI-BCI. Future studies should look to reduce time duration for classifier training. Or, it would be also a potential future direction to develop a subject independent system as it need not collect data to train classifier by using data collected from others. This subject independent system would be useful in clinical rehabilitation, where patients with brain abnormalities pertaining to motor, cognitive or emotion processing could be retrained to achieve normal level of functioning by providing feedback from a real-time pattern classifier that is trained on healthy subjects. Also, it would be promising to build an fMRI-BCI system based on spatial-temporal patterns as the present work ignored the temporal pattern of brain activity.

In real-time classification of brain states during emotional recall, binary classification between happiness and disgust, and multi-class classification among happiness, disgust, and sadness were successfully performed with high classification accuracy (Sitaram, Lee et al. in press). Particularly, robust prediction accuracies were observed even when participants were intentionally instructed to use different strategies between SVM training and testing runs, indicating the ability of the classifier to generalize across varied emotion imagery strategies and memory recall scenarios. Additionally, this study provided the first objective comparison, through pattern classification, of the degree of involvement of different brain regions in emotion imagery and regulation. This technique could be applied in a number of ways for research in affective neuroscience as well as in the treatment of emotional disorders. In the current work, we did not investigate changes in behavior through feedback training based on the brain states induced by recalling of emotional episodes, a topic that could be further scrutinized in future studies.

Lastly, it was showed that movement intention and the type of movement could be decoded from the human fronto-parietal cortex, opening up the possibility

for developing movement intention based prosthetics. However, it was difficult to clearly separate the intention period and the imagery period, and there was no additional monitoring of signals related to motor execution such as eye movement and electromyography (EMG) during performance of the task, issues that are currently being addressed in a continuation of the study.

2. 8. References

Amodio, D. M. and C. D. Frith (2006). "Meeting of minds: the medial frontal cortex and social cognition." *Nat Rev Neurosci* 7(4): 268-77.

Caria, A., R. Sitaram, et al. (2010). "Volitional Control of Anterior Insula Activity Modulates the Response to Aversive Stimuli. A Real-Time Functional Magnetic Resonance Imaging Study." *Biol Psychiatry*.

Caria, A., R. Veit, et al. (2007). "Regulation of anterior insular cortex activity using real-time fMRI." *Neuroimage* 35(3): 1238-46.

deCharms, R. C., F. Maeda, et al. (2005). "Control over brain activation and pain learned by using real-time functional MRI." *Proc Natl Acad Sci U S A*. 102(51): 18626-31. Epub 2005 Dec 13.

Haynes, J. D. and G. Rees (2006). "Decoding mental states from brain activity in humans." *Nat Rev Neurosci* 7(7): 523-34.

Haynes, J. D., K. Sakai, et al. (2007). "Reading hidden intentions in the human brain." *Curr Biol* 17(4): 323-8.

Kamitani, Y. and F. Tong (2005). "Decoding the visual and subjective contents of the human brain." *Nat Neurosci* 8(5): 679-85.

LaConte, S., S. Strother, et al. (2005). "Support vector machines for temporal classification of block design fMRI data." *Neuroimage* 26(2): 317-29.

LaConte, S. M., S. J. Peltier, et al. (2007). "Real-time fMRI using brain-state classification." *Hum Brain Mapp* 28(10): 1033-44.

Lee, S., S. Halder, et al. (2010). "Effective functional mapping of fMRI data with support-vector machines." *Hum Brain Mapp* 31: 1502-1511.

Lee, S., S. Ruiz, et al. (in press). "Detection of cerebral reorganization induced by real-time fMRI feedback training of insula activation." *Neurorehabilitation and Neural Repair*.

Mitchell, T. M., R. Hutchinson, et al. (2003). "Classifying instantaneous cognitive states from fMRI data." *AMIA Annu Symp Proc*: 465-9.

Mourao-Miranda, J., A. L. Bokde, et al. (2005). "Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data." *Neuroimage*. 28(4): 980-95. Epub 2005 Nov 4.

Mourao-Miranda, J., K. J. Friston, et al. (2007). "Dynamic discrimination analysis: a spatial-temporal SVM." *Neuroimage* 36(1): 88-99.

Mourao-Miranda, J., E. Reynaud, et al. (2006). "The impact of temporal compression and space selection on SVM analysis of single-subject and multi-subject fMRI data." *Neuroimage* 33(4): 1055-65.

Phan, K. L., T. Wager, et al. (2002). "Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI." *Neuroimage* 16(2): 331-48.

Posse, S., D. Fitzgerald, et al. (2003). "Real-time fMRI of temporolimbic regions detects amygdala activation during single-trial self-induced sadness." *Neuroimage* 18(3): 760-8.

Rota, G., R. Sitaram, et al. (2009). "Self-regulation of regional cortical activity using real-time fMRI: the right inferior frontal gyrus and linguistic processing." *Hum Brain Mapp* 30(5): 1605-14.

Ruiz, S., S. Lee, et al. (in preparation). "Multivariate prediction of movement intention in the human fronto-parietal cortex."

Ruiz, S., S. Lee, et al. (submitted). "Learned self-regulation of anterior insula in schizophrenia: effects on emotion recognition and neural connectivity " *Molecular Psychiatry*.

Schlkopf, B., C. J. C. Burges, et al. (1999). *Advances in kernel methods : support vector learning*. Cambridge, Mass., MIT Press.

Schlkopf, B. and A. J. Smola (2002). *Learning with kernels : support vector machines, regularization, optimization, and beyond*. Cambridge, Mass., MIT Press.

Sitaram, R., S. Lee, et al. (2011). *Real-Time Regulation and Detection of Brain States from fMRI Signals. Neurofeedback of Neuromodulation technique and applications*. R. COBEN and J. EVANS. London, elsevier: 227-254.

Sitaram, R., S. Lee, et al. (in press). "Real-time support vector classification and feedback of multiple emotional brain states." *Neuroimage*.

Soon, C. S., M. Brass, et al. (2008). "Unconscious determinants of free decisions in the human brain." *Nat Neurosci* 11(5): 543-5.

Vapnik, V. N. (1998). *Statistical learning theory*. New York, Wiley.

Weiskopf, N., R. Veit, et al. (2003). "Physiological self-regulation of regional brain activity using real-time functional magnetic resonance imaging (fMRI): methodology and exemplary data." *Neuroimage* 19(3): 577-86.

Yoo, S. S., T. Fairney, et al. (2004). "Brain-computer interface using fMRI: spatial navigation by thoughts." *Neuroreport* 15(10): 1591-5.

Yoo, S. S. and F. A. Jolesz (2002). "Functional MRI for neurofeedback: feasibility study on a hand motor task." *Neuroreport* 13(11): 1377-81.

Yoo, S. S., H. M. O'Leary, et al. (2006). "Increasing cortical activity in auditory areas through neurofeedback functional magnetic resonance imaging." *Neuroreport* 17(12): 1273-8.

3. Personal contributions to papers and manuscripts

<p>Sitaram, R., S. Lee, et al. (2011). Real-Time Regulation and Detection of Brain States from fMRI Signals. Neurofeedback of Neuromodulation technique and applications. R. COBEN and J. EVANS. London, elsevier: 227-254.</p> <p>For this book chapter I wrote the parts of historical development of Real-Time fMRI, overview of the fMRI-BCI system.</p>	1
<p>Lee, S.[†], S. Halder, et al. (2010). "Effective functional mapping of fMRI data with support-vector machines." Human Brain Mapping [Epub ahead of print] [†] Indicates the corresponding author.</p> <p>For this paper I contributed by conceiving the idea, designing the model, performing all the simulations, and writing the manuscript.</p>	2
<p>Lee, S.[†], S. Ruiz, et al. (in press). "Detection of cerebral reorganization induced by real-time fMRI feedback training of insula activation: A multivariate investigation." Neurorehabilitation and Neural Repair. [†] Indicates the corresponding author.</p> <p>For this paper, I formulated the hypothesis with author Sitaram, programmed all the scripts for analysis, and performed all the analysis. Finally, I wrote the manuscript with other authors.</p>	3
<p>Sitaram, R.*[†], S. Lee*[†], et al. (in press). "Real-time support vector classification and feedback of multiple emotional brain states." Neuroimage [Epub ahead of print]. * - Authors contributed equally. [†] Indicates the corresponding author.</p> <p>In this study, I built the fMRI-BCI system, designed the experimental protocols, conducted all the experiments, and wrote the manuscript with other authors.</p>	4
<p>Ruiz, S.*, S. Lee*, et al. (in preparation). "Multivariate prediction of movement intention in the human fronto-parietal cortex." * - Authors contributed equally.</p> <p>In this study, I designed the experiment with authors Ruiz and Sitaram, conducted all the experiments with author Ruiz, programmed all the scripts for analysis, and performed the analysis and wrote the manuscript with other authors.</p>	5

4. Published papers and manuscripts

4.1. Real-Time Regulation and Detection of Brain States from fMRI Signals

Real-Time Regulation and Detection of Brain States from fMRI Signals

Ranganatha Sitaram¹, Sangkyun Lee^{1,2}, Sergio Ruiz^{1,2,4}, and Niels Birbaumer^{1,3}

¹Institute of Medical Psychology and Behavioral Neurobiology, University of Tübingen, Tübingen, Germany

²Graduate School of Neural and Behavioral Sciences, International Max Planck Research School, Tübingen, Germany

³Ospedale San Camillo, Istituto di Ricovero e Cura a Carattere Scientifico, Venezia – Lido, Italy

⁴Department of Psychiatry, Faculty of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile

Contents

Introduction	227
Historical Development of Real-Time fMRI	228
Overview of the fMRI–BCI system	232
Signal Acquisition	232
Pre-processing	233
Signal Analysis	234
Feedback Generation	237
fMRI–BCI in Research and Clinical Treatment	238
fMRI–BCI Studies in Healthy Population	238
fMRI–BCI Studies in Clinical Populations	244
Conclusions	248
References	249

s0010 INTRODUCTION

p0010 Functional magnetic resonance imaging (fMRI) allows non-invasive assessment of brain function with high spatial resolution and whole brain coverage, by measuring changes of the blood oxygenation level-dependent (BOLD) signal. Although the BOLD response is an indirect measure of neural activity, there is accumulating evidence suggesting the close coupling between BOLD and electrical activity of the neurons

(Logothetis, 2008). It is postulated that the combined effect of increases and decreases in deoxygenated hemoglobin content resulting from changes in cerebral blood volume, cerebral blood flow, and oxygen metabolism following neural firing results in the BOLD signal (Buxton, Uludağ, Dubowitz, & Liu, 2004). fMRI data typically consist of time-series of several hundred 3D images across the brain over a period of time, with each image acquired every few seconds. fMRI usability and applications have been somewhat limited by the offline mode of data analysis, due to the large size of the data, very intensive computation involved in preprocessing and analysis of fMRI images. Fortunately, innovations in high-performance magnetic resonance scanners and computers, combined with developments in techniques for faster acquisition, processing and analysis of MR images, have burgeoned a fresh round of developments in fMRI methodology for scientific research and clinical treatment. Real time fMRI (rtfMRI) permits simultaneous measurement and observation of brain activity during an ongoing task. Online single subject preprocessing and statistical analysis of functional data is now possible within a single repetition time (TR), of 1.5–2 s. Novel applications based on rtfMRI have been developed in the last decade, including fMRI data quality assessment, neurosurgical monitoring, and neurofeedback for self-regulation of brain activity. Adding to the above developments are more recent advances in multivariate pattern classification of fMRI signals based on which brain activations in a whole neural network, rather than just a single brain region, could be potentially decoded and modulated (LaConte et al., 2007, Sitaram et al., 2010). Here, we first review historical developments in rtfMRI, followed by an overview of a fMRI Brain–Computer Interface (fMRI–BCI) system for enabling decoding and self-regulation of brain activity, and finally discuss the results of several studies on healthy individuals and patient populations for clinical treatment of neuropsychological disorders.

s0015 **HISTORICAL DEVELOPMENT OF REAL-TIME fMRI**

p0015 Functional imaging experiments typically follow a serial procedure in which fMRI images are first acquired from one or more participants performing the tasks under investigation, followed by an offline procedure of signal preprocessing of the images and statistical mapping that may take several days. This sequential processing approach has initially developed due to the large size of data generated in neuroimaging experiments,

whose processing and analyses incurred high computational costs. However, it is to be admitted that lack of online implementation of fMRI preprocessing and analysis could limit its applicability and usability in several ways.

p0020 First, experimenters cannot monitor data quality during scanning in the absence of online processing. Considering the high cost of MRI scanning, a system for monitoring of data quality is useful for saving the costs and efforts to obtain good quality data. If an investigator can detect artifacts correlated to the stimulus in real-time, such as head motion or other physiological artifacts, it enables him or her to re-acquire the data immediately. Furthermore, in the absence of online analysis, it is hard to modify an experimental design based on the physiological and behavioral changes of the participant. These limitations were first overcome by the pioneering work of Cox and colleagues.

p0025 Cox and colleagues (Cox, Jesmanowicz & Hyde, 1995) first proposed a method to analyze brain activation in real-time by computing correlation between a voxel's BOLD signal and a reference time series of the experimental design in a recursive manner to reduce the computational demands. Statistical maps of the brain were obtained in real-time by coloring voxels that show higher correlations than a user-specified threshold. However, this method did not yet correct for non-specific noise originating in the scanning instrument. Voyvodic (1999) reported an approach to improve the flexibility of the experimental paradigm with a software program for accurate, real-time paradigm control and online fMRI analysis. The paradigm control included simultaneous presentation of stimuli, automatic synchronization to an fMRI scanner, and monitoring of a variety of physiological and behavioral responses. The online analysis performed MR image reconstruction, head motion correction in the translational motion, and statistical tests for block or event-related design. Since then, many studies have improved the image acquisition process in terms of data quality, speed, and statistical power (Gao and Posse, 2003; Posse et al., 1999, 2001, 2003; Weiskopf et al., 2005; Yoo et al., 1999) and algorithms (Bagarinao et al., 2003; Cox & Savoy, 2003; Cox et al., 1995; Cox & Jesmanowicz, 1999; Gao & Posse 2003; Gembris, Taylor, Schor, Frings, Suter, & Posse, 2000; LaConte et al., 2007; Sitaram et al., 2010; Smyser et al., 2001; Voyvodic, 1999).

p0030 Further developments in multiecho echo-planar imaging (mEPI) increased the functional contrast-to-noise ratio (CNR) by sampling multiple echoes in a single shot of radio frequency (RF) pulse (Posse et al.,

1999). This method is suitable for rtfMRI as it allows the detection of small signal changes within each echo (of 30–40 ms in a 3T scanner). Adaptive multi-resolution EPI (Panych & Jolesz, 1994) achieved high spatial and/or temporal resolution in regions of functional activations distributed throughout the brain by selectively detecting those regions with RF encoding in multiple stages. Such a method can zoom into the regions of activations while ignoring quiescent regions. Yoo and his colleagues (Yoo et al., 1999) implemented a real-time adaptive acquisition system with a multi-resolution EPI to extract signals from cortex and allow more efficient data acquisition time.

p0035 Another notable improvement in imaging was in removing magnetic susceptibility artifacts. Due to the differences in magnetic susceptibilities of different imaged parts, such as air, bone, different brain tissues, the static magnetic field is not homogenous near borders of two brain regions. Particularly, in air–tissue interfaces such as the regions of the basal brain and frontal sinuses greater geometric distortions occur. When one radio-frequency excitation pulse is applied to a slice in single-shot imaging and a read-out time (T_{RO}) of about 10–40 ms is used to encode the slice, the inhomogeneities cause local shifts near the air–tissue interfaces (the misalignment of the functional images to wrong anatomical structure) in the resonance frequency ($1/T_{RO} = 25–100$ Hz). Several methods (Andersson, Hutton, Ashburner, Turner, & Friston, 2001; Jezzard & Balaban, 1995; Kybic et al., 2000; Studholme, Constable, & Duncan, 2000; Zaitsev, Hennig, & Speck, 2004) have been suggested to reduce the geometric distortion. However, not every one of these methods may be suitable for real-time applications as they require additional reference scans and computational time. Weiskopf and colleagues (Weiskopf et al., 2005) developed a real-time method to allow for simultaneous acquisition and distortion correction of functional images contributing further to the development of real-time fMRI.

p0040 Taken together, many different algorithms have been developed for the acquisition and real-time processing of fMRI signals. To improve the sensitivity of functional imaging, correction of head movement artifacts is a challenging problem. According to Cox and Jesmanowicz (1999), if two neighboring voxels differ in intrinsic brightness by 20%, then a motion of 10% of a voxel dimension can result in a 2% signal change – comparable to the BOLD signal change at 1.5T (Bandettini et al., 1992; Cox & Jesmanowicz, 1999). In addition, if movement correlates with a given task/stimuli, it can elicit false activations (Hajnal et al., 1994). If movement

is not correlated with the task/stimuli, signal changes due to the movement can change or reduce the actual activation in functional images (Cox & Jesmanowicz 1999). Cox and his colleagues (1999) reported an online method for head-movement correction by an algorithm for three-dimensional (3D) image rotation and shifting (a rigid body model; since the head moves as a whole, it is assumed that motion of the head can be estimated in three directions for translation and three directions for rotation) by generalizing the shears factorization directly to three dimensions. However, this approach is not suitable for real-time applications as the time needed for realignment of 80 images is estimated to be several minutes. Mathiak and Posse (2001) reported a real-time method of head motion correction in which rigid body (from six parameters including three for translation and three for rotation) transformation was applied in the interval between acquisitions of two functional images. Besides improving the performance of realignment of the functional images, an error (indicating potentially noise or artifact) added by the realignment process also would need to be considered. Mathiak and Posse (2001) proposed that at least three slices with an image matrix of 64×64 would be required to reduce the error in the estimated movement parameters to less than 1% of the voxel size.

p0045 Further concerning realignment, a variety of real-time pre-processing techniques for the corrections of respiration artifacts, spatial smoothing (Posse et al., 2003) and spatial normalization to stereotactic space (Gao & Posse 2003) have been developed. To identify significant voxels in online analysis, methods such as correlation (Cox et al., 1995), general linear model (Caria et al., 2007, 2010; Rota et al., 2009; Ruiz et al., 2008; Smyser et al., 2001; Weiskopf, Veit, Wilhelm, & Elena, 2003), and t-tests (Voyvodic, 1999) have been used.

p0050 The above technical advances in rtfMRI enabled the development of fMRI-BCI systems (see the following sections) for self-regulation of brain activity with neurofeedback to study plasticity and functional reorganization. Neurofeedback is based on the psychological theory of instrumental learning, i.e., training in the presence of contingent reward (Skinner, 1938; Weiskopf et al., 2004). Studies have reported different methods for the generation of reward. Yoo and Jolesz (2002) used the statistical map of brain activations as visual feedback. Posse et al. (2003) gave participants verbal feedback of the BOLD signal change in the amygdala at intervals of 60 s. Weiskopf et al. (2003) introduced real-time feedback by showing two time courses of the BOLD signal in two circumscribed brain regions,

namely, rostral–ventral and dorsal anterior cingulate cortex, which were updated at an interval of 2 s. DeCharms, Christoff, Glover, Pauly, Whitfield, & Gabrieli (2004) built an fMRI-BCI system to guide self-regulation of the somatomotor cortex with visual feedback of three different time courses, in the target ROI, in a background ROI (irrelevant to the task performance), and difference between the two ROIs. Recent studies (Caria et al. 2007, 2010; deCharms et al., 2004, 2005; Rota et al. 2009) have used visual feedback in the interval of 1–2 s.

s0020 **OVERVIEW OF THE fMRI–BCI SYSTEM**

p0055 fMRI–BCI could be defined as a closed loop system that extracts brain signals from regions of interest and/or classifies patterns of brain activity from a whole neural network in real-time, so that this information can be provided to the subject as contingent feedback to enable him to volitionally control the brain activity. In general, an fMRI–BCI system is comprised of the following subsystems: (1) the subject, (2) signal acquisition, (3) pre-processing, (4) signal analysis, and (5) feedback generation. Depending on the purpose of the experiment, the subject would be instructed and trained to perform mental tasks guided by the feedback information. Here, we describe the last four subsystems mainly based on the fMRI–BCI system built in the Institute of Medical Psychology and Behavioral Neurobiology, University of Tübingen, Germany. The word real-time in the following sections implies that signal processing is performed within a single TR (for example, 1.5 s).

s0025 **Signal Acquisition**

p0060 Experiments are conducted on a 3T whole body scanner using a standard 12 channel head coil (Siemens Magnetom Trio Tim, Siemens, Erlangen, Germany). In principle, scanners from other manufacturers and with other field strengths could be used for fMRI–BCI development. Whole brain images of the subject are acquired using an EPI pulse sequence which is modified to export functional images to the host computer of the scanner. Pulse sequence parameters used for signal acquisition in our experiments were as follows: repetition time $TR = 1.5$ seconds, echo time $TE = 45$ msec, flip angle = 70 degree, number of slices = 16, bandwidth = 1.3 KHz/pixel, $FOV_{PE/RO} = 210$, image matrix = 64×64 , voxel size = $3 \times 3 \times 5\text{mm}^3$. These parameters could be modified keeping in mind, however, the trade-off between signal-difference to noise ratio, also

called the contrast-to-noise ratio (CNR), and the spatio-temporal resolution. Intrinsic parameters that modify the inherent signal produced by a volumetric element (voxel) of the tissue, such as TR, TE, and flip angle affect the CNR. Extrinsic parameters do not affect the measured tissue, but influence the mechanics of the data collection, e.g., spatio-temporal resolution (Brown & Semelka, 1999). In the setup of the parameters for real-time signal acquisition, a suitable compromise must be made between spatial resolution, i.e., FOV, image matrix and slice thickness, number of slices, and temporal resolution, i.e., TR. In our fMRI-BCI experiments, the repetition time was reduced to 1.5 seconds (compared to 2–3 s in a conventional fMRI experiments) to increase the temporal resolution of BOLD signal, while the number of slices was reduced to 16 (from 25–30 slices in conventional fMRI measurement) and slice thickness was increased to 5 mm. With these parameters, adequate spatial and temporal resolutions for functional image acquisition were provided. However, the older MRI operating software provided by manufacturers (e.g., Syngo, version VB13 and before, Siemens Medical Solutions, Erlangen, Germany) did not have a provision for online export of functional images. To retrieve images in real-time for online processing, a generalized image reconstruction module was inserted into the conventional EPI sequence provided by Siemens (Caria et al., 2007, 2010; Rota et al., 2009; Sitaram et al., 2010; Weiskopf et al., 2003). This module receives EPI k-space raw data from the MRI scanner hardware, reconstructs whole brain images before start of the next volume of the brain, and stores them in a pre-specified directory that could be immediately accessed for another program to perform online preprocessing and analysis. Fortunately, recent versions of scanner operating software (e.g., Siemens Syngo version VB15, VB17 and VA30) have provided a standard option to enable real-time export of functional images, simplifying and standardizing the future development of real-time fMRI.

s0030 **Pre-processing**

p0065 After the acquisition of each volume of EPI images, various online pre-processing steps could be performed for artifact removal and noise reduction. To prevent artifactual signals caused by head movement, head padding and bite bars could be used. In addition, through a head-motion correction step, we could monitor how much the participants move in 6 directions (3 translations and 3 rotations) in real-time, and instruct them

to avoid such movement. Head movement can be inferred by monitoring the time courses of the head movement parameters. Patients could then be reminded to avoid moving if excessive head movement is observed. In addition to this motion correction, pre-processing also includes spatial smoothing to reduce the effect of noise, and de-trending to remove linear trends in the BOLD time-series.

s0035 **Signal Analysis**

p0070 After completion of pre-processing steps, whole-brain images are used for statistical analysis and generation of functional maps. For real-time statistical analysis, a variety of algorithms that perform subtraction of two different conditions, correlation analysis, or general linear model could be used. Real-time statistical analysis is usually performed either by analyzing recent time samples of data extracted from a sliding window or by incrementally analyzing all data acquired up to a given time point. The sliding-window method is superior in reflecting the current brain state as it uses the most recent information. However, this method is not statistically powerful because a limited number of samples are used in the statistical test. In contrast, the incremental method provides more robust information by using all the data acquired up to a given time point.

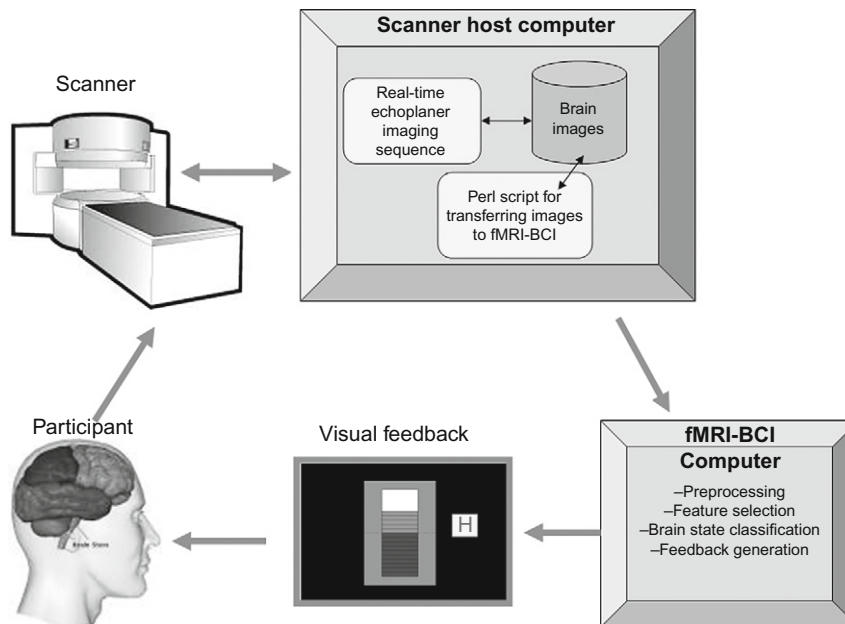
p0075 Studies (Caria et al., 2007, 2010; deCharms et al., 2004; Rota et al., 2009) have also used the subtraction method of determining activation maps where signals in the baseline condition are subtracted from the activation condition in a sliding window to provide feedback information in the ROIs. The correlation method (Cox et al., 1995; Gembris et al., 2000; Posse et al., 2001) is applied by computing the correlation coefficient between the time-series of the measured BOLD signal at each voxel and the reference (or design) time-series representing the change in the task conditions, and assessing the coefficients with a specified threshold. The correlation method can be used in either a sliding-window fashion or an incremental fashion.

p0080 General Linear Model (GLM) is now a standard method of analysis of functional images to estimate the parameters that fit the measured time-series of BOLD signal at each voxel, with a linear summation of multiple experimental and confounding effects weighted by the corresponding parameters. A GLM for two different conditions and one confounding effect produces three parameters to best fit the measured time-series of BOLD signal at a voxel. These parameters are computed independently

for all the voxels of the brain and then used for statistical tests such as the *t*-test and the F-test. As an example, the *t*-test can be applied on the first parameter values over all the brain voxels, where the parameter magnitudes correspond to involvement of each voxel in the first condition. These statistical tests identify the voxels that are significant (i.e., activated solely in the corresponding condition) in the tests. However, since the number of samples (i.e., scanned images) is limited in real-time applications, Bagarinao et al. (2003) developed a method of real-time GLM by updating the parameters in GLM as new data become available. Similarly, the commercially available real-time fMRI analysis software, Turbo Brain Voyager (TBV, Brain Innovations, Maastricht, Netherlands) used in our fMRI-BCI setup (Caria et al., 2007, 2010; Rota et al., 2009; Ruiz et al., 2008; Weiskopf et al. 2003) uses real-time GLM by applying the recursive least squares regression algorithm (Pollock, Green, & Nguyen, 1999) to update GLM estimation incrementally.

p0085 All of the above methods are univariate methods as all the statistical tests are separately performed at each voxel. In contrast, multivariate methods can recognize spatial and temporal patterns of activity from multiple distributed voxels in the brain. Multivariate methods accumulate weak information available at multiple locations to jointly decode cognitive states although information at any single location cannot differentiate between the states (Haynes & Rees 2006). Recent studies (Cox & Savoy 2003; Mitchell et al., 2003; Kamitani & Tong, 2005; Haynes & Rees 2006; Haynes, Sakai, Rees, Gilbert, Frith, & Passingham, 2007; Harrison & Tong 2009; Lee et al., 2010) have applied multivariate methods to increase sensitivity of fMRI analysis. Laconte and his colleagues (2007) developed an fMRI-BCI system by employing a multivariate pattern classification method called support vector machines (SVM). While most fMRI-BCI studies to date have investigated self-regulation of brain activity at one or two ROIs using univariate analysis, multivariate methods allow for real-time feedback of a whole network of brain activity pertaining to a task.

p0090 We have recently implemented a real-time classification method for automatically recognizing multiple emotional brain states from fMRI signals (Sitaram et al., 2010; see Figure 9.1). In our study, participants were instructed to recall two (happy and disgust) or three (happy, disgust, and sad) salient emotional episodes in a block design paradigm. While participants performed emotional imagery, whole brain images were acquired and pre-processed in real-time to correct for head-motion artifacts and



0010 **Figure 9.1** The Tübingen real-time fMRI brain state classification system is comprised of the following subsystems: an image acquisition subsystem, which is a modified version of the standard echo-planar imaging (EPI) sequence written in C and executed on the scanner host computer; and an fMRI-BCI subsystem, which performs image preprocessing, brain state classification and visual feedback, implemented in C and Matlab scripts (Mathworks, Natwick, MA) and executed on a 64-bit Windows desktop. A Perl-script on the scanner host transfers the acquired images after every scan (at an interval of 1.5s) to the fMRI-BCI computer. (Reproduced from *Neuroimage with permission.*)

AU:4

spatially smoothed to improve signal-to-noise ratio. As the number of voxels in the brain images is too large (tens of thousands, depending on the scanning parameters) for a pattern classifier to handle efficiently, a computational method called feature selection needs to be carried out to reduce the data size and improve the efficacy of classification. We have developed a novel method of feature selection called effect mapping (Lee et al., 2010). In practice, the emotion classifier is first trained for each participant on a set of initial data taken out of two sessions of the experiment. The trained classifier is then used to recognize online (at intervals of 1.5 s) the states of emotion, namely, happy, disgust or sad, as the participant engages in emotional imagery. In our study, participants were provided real-time visual feedback of the state of their brain based on the

classifier output. Our study, performed on 18 healthy individuals, showed that fMRI–BCIs built using a pattern classifiers can robustly decode multiple brain states (average classification accuracy >80%) in real time and provide feedback of the states.

s0040 **Feedback Generation**

p0095 As explained earlier, the results of real-time signal analysis in the ROI or in the network determined by the pattern classifier can be transformed to generate contingent feedback to the participant. Recent studies have shown the usefulness of feedback in learning to regulate the BOLD signal in one or more regions of the brain with fMRI–BCI (Caria et al., 2007, 2010; deCharms et al., 2004, 2005; Posse et al., 2003; Rota et al., 2009; Ruiz et al., 2008). In addition, Lee and his colleagues (2008) reported the effects of neurofeedback in self-regulation training by showing that insula-regulation training with feedback leads to cerebral reorganization in the brain regions relevant to emotion processing (Lee, Sitaram, Ruiz, & Birbaumer, 2008). In studies with rtfMRI (Caria et al., 2007, 2010; deCharms et al., 2004, 2005; Posse et al., 2003; Rota et al., 2009; Ruiz et al., 2008), different forms of feedback have been used. Except for one study that used verbal feedback (Posse et al., 2003), most studies to date (Caria et al., 2007; deCharms et al., 2004, 2005; Rota et al., 2009; Ruiz et al., 2008) have used visual feedback. Hinterberger et al. (2003) showed that visual feedback compared to auditory feedback leads to better learning in self-regulation of slow cortical potential (SCP; change of cortical potential below 1 Hz) using an EEG–BCI. As fMRI experiments generally have more acoustic background noise, visual feedback is presumably more effective. Visual feedback has been provided in a variety of forms such as time courses of BOLD activity in target areas updated in real-time, functional maps based on online statistical analysis, and virtual reality based animation. Using the conventional univariate method of analysis, feedback can be based on one or more ROIs by combining them in the form of additive or subtractive contrasts. The subtraction of BOLD signals of the reference area from that of the ROI can result in more robust feedback as the BOLD activity in the reference area reflects the global change of brain due to head movements, swallowing and systemic changes in BOLD. More sophisticated methods of feedback computation could potentially include correlational analysis between time-series of BOLD activations in the ROIs, bivariate or multivariate methods of functional connectivity computation using Granger causality modelling (GCM), and

multivariate pattern classification (Laconte et al., 2007; Sitaram et al., 2010). Future developments in the adaptation of existing brain signal analysis methods to real-time requirements will dictate how well these methods could be used in fMRI-BCI applications.

s0045 **fMRI–BCI IN RESEARCH AND CLINICAL TREATMENT**

p0100 In the last decades, neurofeedback based on electrical brain signals has been successfully applied to train subjects to self-regulate different components of the electroencephalogram, leading to measurable behavioral changes. EEG neurofeedback has been therapeutically applied to neurological and psychiatric disorders, such as intractable epilepsy, stroke, locked-in syndrome and amyotrophic lateral sclerosis (Birbaumer, 2006; Fuchs, Birbaumer, Lutzenberger, Gruzelier, & Kaiser, 2003; Kotchoubey et al., 2001; Kubler, Kotchoubey, Kaiser, Wolpaw, & Birbaumer, 2001; Murase, Duque, Mazzocchio, & Cohen, 2004; Strehl et al., 2006). For its part, the development of real-time fMRI and fMRI–BCI has been more recent. In the next sections, we will review the studies that have been conducted so far with this new technique on healthy subjects, and the attempts to implement this methodology on clinical populations.

s0050 **fMRI–BCI Studies in Healthy Population**

s0055 ***Regulation of Brain Regions of Emotion***

p0105 The modulation of brain areas related to emotional processing has been of particular interest for fMRI–BCI research. Posse and colleagues (Posse et al., 2003) used rtfMRI and feedback of amygdala activation to reinforce mood induction. Amygdala modulation can be of special importance due to its role in emotion processing and learning, and due to its involvement in several neuropsychiatric disorders (Buchanan, 2007; Lawrie, Whalley, Job, & Johnstone, 2003; Pause, Jungbluth, Adolph, Pietrowsky, & Dere, 2010). A group of six healthy subjects performed a paradigm of self-induction of mood to alternate between neutral and sad affective states while in the scanner. After each trial, subjects received verbal feedback of the signal change of BOLD activation in the amygdala. Subjects successfully achieved sad induction, and their self-mood ratings correlated with the level of activity in the amygdala. Therefore, this study showed that real-time fMRI could be used to monitor the activations of a particular brain area, and suggested that feedback may influence perceived mood. However, as both the task of self-induction and BCI feedback were

presented simultaneously, it was not possible to ascertain whether amygdala self-regulation was actually achieved due to the BCI feedback.

p0110 Weiskopf and colleagues from our group (Weiskopf et al., 2003) investigated whether fMRI–BCI could be applied to achieve self-regulation of the anterior cingulate cortex (ACC). The ACC is part of the limbic system and is subdivided into a dorsal “cognitive division” (ACcd), and a rostral-ventral “affective division” (Acad) (Bush, Luu, & Posner, 2000). A healthy subject was instructed to upregulate the activations of these areas, presented to him as a continuously updated visual feedback (delay <2 s), introducing the concept of “immediate feedback” of the BOLD signal. Using imagery of landscapes, sports, and social interaction in many training sessions, the subject achieved a significant increase of the BOLD signal between up-regulation and baseline blocks. Additionally, the signal in the affective ACC increased across the sessions of training, suggesting a “learning effect”. Although it was not possible to correlate the BOLD signal change with any behavioral measurement, the participant rated the valence and arousal of his affective state as more positive for regulation blocks compared to baseline blocks. Hence, this single subject study was one of the first that convincingly showed that BOLD signal of circumscribed brain regions can be self-regulated using fMRI–BCI, and that self-regulation by fMRI–BCI training might lead to affective modifications. Studies have shown (Caria et al., 2007, 2010) that a robust ability to self-regulate a brain region can be learned with contingent feedback, and not with sham feedback, nor with mental imagery without feedback.

p0115 Further to this study, our group investigated whether healthy subjects can voluntarily gain control over anterior insular activity using fMRI–BCI (Caria et al., 2007). The insula cortex is an anatomically complex mesocortical structure, part of the paralimbic system that plays a central role in sensory integration, emotion, and cognition, including such functions as olfaction, gustation, autonomic functions, temperature and pain perception, and self-awareness (for reviews see Augustine, 1996; Craig, 2009; Ture, Yasargil, Al-Mefty, & Yasargil, 1999). Modulation of the insular activity with fMRI–BCI training might be relevant for treatment of different psychiatric diseases as social phobia, antisocial behavior, schizophrenia, and addictive disorders (Nagai, Kishi, & Kato, 2007; Naqvi & Bechara, 2009). In this study, nine healthy subjects were trained to voluntarily control the BOLD signal of the anterior insular cortex, using fMRI–BCI in four feedback sessions. The visual feedback was the normalized and continuously updated average BOLD signal from the right anterior insula, presented to the subjects

by means of thermometer bars. All participants were able to successfully regulate the BOLD signal, and training resulted in a significantly increased activation cluster in the anterior portion of the right insula across sessions. However, self-regulation could not be achieved by a control group trained with sham feedback, suggesting that successful self-regulation is achieved by contingent fMRI feedback. This work was the first group study that showed that volitional control of emotionally relevant brain areas can be attained by fMRI–BCI training.

p0120 Given this, can learned self-regulation produce a measurable behavioral modification? In a later study, Caria and colleagues (Caria et al., 2010) explored the relationship between brain self-regulation and emotional behavior using fMRI–BCI. Healthy participants underwent four fMRI–BCI scanning sessions to modulate the BOLD response in the left anterior insula guided by visual feedback (as in the previous study). After each modulation block of self-regulation and baseline, participants were presented with either an emotionally negative or a neutral picture taken from the International Affective Picture System (IAPS; Lang, Bradley, & Cuthbert, 1997). Immediately after presentation, participants were required to rate the picture using the Self-Assessment Manikin (Lang, 1980). Participants learned to increase and decrease the BOLD response significantly in the anterior insula guided by contingent feedback, and behavioral data showed a significant difference of valence ratings of the aversive pictures in the last session. These results demonstrate that fMRI–BCI manipulation of paralimbic regions such as insula is possible and can modulate a specific emotional response.

p0125 A question that arises from these studies is whether, in addition to behavioral modification, does self-regulation of circumscribed brain areas also lead to cerebral reorganization. In a further analysis of our data reported in Caria et al. (2007), we investigated (Lee et al., 2008) changes in brain connections associated with anterior insula self-regulation. We used multivariate support pattern analysis and effective connectivity analysis with Granger causality modeling (Seth, 2009). Our analyses revealed changes in the neural network of emotion regulation represented by an inverted U-curve of connectivity densities across the sessions of self-regulation. Feedback training resulted in an initial increase of the density of the connections among regions such as the left and right insula, ACC, medial prefrontal cortex, dorsolateral prefrontal cortex, and amygdala. Further training seemed to indicate pruning of presumably redundant connections and strengthening of potentially relevant connections. This

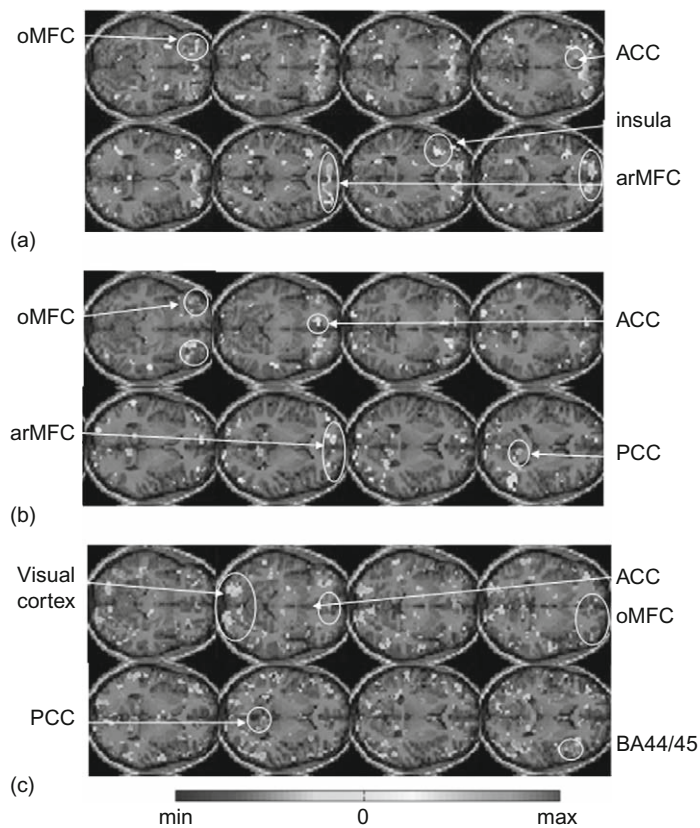
effect can be of special importance as it shows that BCI might be used to build a more efficient neural pathway, especially for conditions in which abnormal neural connectivity is implied (as in autism and schizophrenia).

p0130 In a very recent study, Hamilton and colleagues (Hamilton et al., 2010) tested whether healthy individuals can downregulate the activity of the subgenual anterior cingulate (sACC) cortex with fMRI–BCI. Studies have shown that downregulation of this region produced by deep brain stimulation led to a sustained antidepressive effect in patients with treatment-resistant depression, suggesting that endogenous modulation by BCI might be used as a therapeutic approach (Lozano et al., 2008; Mayberg et al., 2005). Using “positive affect strategies” and visual contingent feedback of the BOLD signal, eight healthy women were able to downregulate the BOLD signal from sACC. The learned downregulation, however, did not persist in a subsequent session where subjects were not provided with feedback. However, the study included a psychophysiological interaction (PPI) analysis of functional connectivity (Friston et al., 1997) which showed that BCI training was associated with a decreased correlation (connectivity) between sACC and posterior cingulate cortex. Similar to the study performed by Lee and colleagues (2008), this finding indicates that fMRI–BCI training leads to changes in the effective connectivity of a network subserving a task.

p0135 Prediction of emotional states from brain activity constitutes a major scope of affective neuroscience. It could solve several pressing clinical problems such as the assessment of affect in verbally incompetent people with dementia, minimally conscious state, and locked-in-syndrome, and the detection of deception. Recent advances in multivariate pattern classification of functional magnetic resonance imaging (fMRI) signals are especially important due to the high spatial resolution, whole brain coverage, and non-invasiveness of fMRI. As mentioned earlier, in a recent study (Sitaram et al., 2010) we showed that an online support vector machine (SVM) can be built to recognize two discrete emotional states, such as happiness and disgust, from fMRI signals in healthy individuals instructed to recall emotionally salient episodes from their lives. The classifier also showed robust prediction rates in decoding three discrete emotional states (happiness, disgust, and sadness) in an extended group of participants. Subjective reports collected from participants ascertained that they performed emotion imagery, and that the online classifier decoded emotions and not arbitrary states of the brain. This study also showed a relationship between the participants’ affect scores as measured by positive and

negative affect scores (PANAS) and the subjective ratings of their performance in the emotion imagery task, indicating that participants who report greater negative affect rate themselves relatively lower in their ability to perform the imagery task. Offline whole-brain classification as well as region-of-interest classification in 24 brain areas previously implicated in emotion processing revealed that the frontal cortex was critically involved in emotion induction by imagery. Finally, we demonstrated an adaptive pattern classifier-based real-time feedback system with which subjects were trained to enhance the functional network of emotion regulation by repeated training.

AU: 5



0015 **Figure 9.2** (See Color Plate section) Exemplary brain activation maps generated from a single subject whole brain SVM classification showing discriminating voxels for: (a) happy vs disgust classification; (b) happy vs sad classification and (c) disgust vs sad classification. Brain regions: oMFC, orbital medial frontal cortex; arMFC, anterior rostral MFC (based on Amodio et al., 2006); OFC, orbitofrontal cortex; ACC, anterior cingulate cortex; PCC, posterior cingulate cortex.

s0060 **Motor System**

p0140 In one of the earliest fMRI–BCI studies, Yoo and Jolesz (2002) tested whether subjects could self-regulate the activity of motor areas by fMRI neurofeedback. Brain activations from sensorimotor regions produced by a simple finger-tapping task were extracted by fMRI, in five healthy participants. Through sessions of regulation, participants were asked to adapt their hand motor strategies in order to expand the functional activations in the motor cortex, guided by the brain activations maps provided as visual feedback, at the end of each block of regulation. After a few trials of training, all participants were able to adapt their motor strategies to successfully expand their brain activations in the ROI.

p0145 DeCharms and colleagues (deCharms et al., 2004) used a hand motor imagery task in six participants who were instructed to optimize their strategies in order to increase the activations in the somatomotor cortex, while receiving ongoing real-time fMRI visual feedback of the level of activations in these brain regions. A significant monotonic activation increase in the ROI across training was also found. Furthermore, a control group of subjects who were trained in the scanner in an identical task, but without valid rtfMRI information (sham feedback), did not achieve self-regulation.

p0150 Other studies have attempted to use fMRI–BCI technology to translate brain activity into direct control of computers and robots. Yoo and colleagues (Yoo et al., 2004) used fMRI to decode brain activities associated with four distinct covert functional tasks (mental calculation, mental speech generation, and motor imagery of sequential finger position of right and left hand), and subsequently translated these activations into predetermined computer commands for moving four directional cursors. Three healthy participants were able to make a cursor navigate in a 2D maze demonstrating “spatial navigation by thought”. The same group (Lee, Ryu, Jolesz, Cho, & Yoo, 2009), later explored the use of fMRI–BCI to control a robotic arm. The BOLD signal extracted from primary motor areas (M1) of right and left hemispheres were used to adjust vertical and horizontal coordinates of this external device. Three healthy subjects attempted to move the robotic arm using motor imagery to activate M1 with the help of visual feedback to adjust the level of cortical activation. With different degrees of success, participants were able to gain voluntary control of two-dimensional movement of the robotic device. These preliminary studies can be of crucial importance for future therapeutic attempts in patients with motor dysfunctions due to stroke, brain or spinal injury, or degenerative disorders.

s0065 Auditory System

p0155 Yoo and colleagues tested the feasibility of using fMRI neurofeedback for the regulation of cortical activations related to auditory attention (Yoo et al., 2006). Eleven healthy participants passively received auditory stimulation in the scanner to determine the auditory areas as regions of interest. During the regulation sessions, participants were instructed to engage in an attentional task (listening to the auditory stimuli) in order to increase the volume of activation within the ROI. Between scanner sessions, verbal feedback of the activation of auditory areas was given to the subjects. The experimental group successfully increased the BOLD signal in left auditory areas and other extra-temporal areas. A control group that was not given feedback of the activations in the auditory cortex did not achieve consistent increase in activations in the region, indicating the importance of feedback in self-regulation.

s0070 Language Processing

p0160 In an experiment conducted by our team, fMRI–BCI was used to train subjects to achieve self-control of right inferior frontal gyrus (BA 45), and to measure whether this voluntary increase of the BOLD signal would modulate language processing (Rota et al., 2009). All seven subjects of the experimental group achieved voluntary self-regulation of the activation-level recorded in the target ROI, with a progressive increase of the level of activation in the right BA 45 across training sessions. Short-term behavioral effects of self-regulation with regard to language processing were explored by comparing accuracy levels and reaction times before and after feedback training in prosodic and syntactic tasks. During the self-regulation of BA 45, experimental subjects significantly improved their accuracy for the identification of affective prosodic stimuli (but not for syntactic stimuli), confirming the role of inferior frontal gyrus in the processing of emotional information. These results pointed out that self-regulation of prefrontal cortical areas by fMRI–BCI is possible, and could be explored as a means to normalize dysfunctional cortical networks to enhance cognitive and/or behavioral disturbances associated with clinical disorders.

s0075 fMRI–BCI Studies in Clinical Populations

p0165 The studies described above showed that fMRI–BCI can enable healthy subjects to achieve voluntary control of circumscribed brain areas. Some

of these studies have also shown behavioral changes and cerebral changes due to fMRI–BCI practice. The next paragraphs describe how far we have gone in the applications of fMRI–BCI with neurological and psychiatric disorders.

s0080 ***Pain Perception***

p0170 Chronic pain is a major health problem causing untold suffering for millions of patients, and economic burden worldwide. DeCharms and colleagues (deCharms et al., 2005) focused on the use of self-regulation of rostral ACC (rACC) to investigate the modulation of pain perception. This area, among others, is involved in mediating the conscious perception of pain (Mackey & Maeda, 2004; Petrovic & Ingvar, 2002; Wager et al., 2004). The results of fMRI–BCI training in a group of healthy subjects showed that it is possible to gain deliberate control of the rACC activation aided by contingent fMRI feedback, and that self-regulation is significantly associated with changes in the perception of pain caused by a noxious thermal stimulus. Going one step further, they demonstrated that similar self-regulation can be achieved by a group of chronic pain patients who reported decrease in the level of ongoing pain after the fMRI–BCI training. It remains to be seen if such behavioral changes persist in the long-term.

s0085 ***Stroke Rehabilitation***

p0175 In recent work by our group, Sitaram and colleagues (2010) assessed the feasibility of fMRI–BCI feedback training for enabling healthy individuals and stroke patients to regulate the ventral premotor cortex (PMv), an area involved in observation, imagery, and execution of movement (Grezes & Decety, 2001). The authors hypothesized that upregulation of the BOLD signal in the PMv would facilitate motor cortical output from primary motor cortex (M1). Each fMRI–BCI feedback session consisted of four runs of self-regulation training. Results showed that training enabled participants to learn to upregulate the BOLD response of PMv. To measure behavioral effects of self-regulation, the authors used paired pulses of transcranial magnetic stimulation (TMS) to induce intracortical inhibition and facilitation, and simultaneously measured motor evoked potential (MEP) on the participant's finger. Results showed evidence for reduction in intracortical inhibition after feedback training compared to the same situation before feedback training, and even further reduction during self-regulation of PMv.

s0090 **Tinnitus Treatment**

p0180 Tinnitus, the perception of sound in the absence of an auditory stimulus, is a common chronic disorder with limited treatment options that can adversely impact the quality of life of the patients. Current views of its neural basis include an over-activation of the auditory network (Eggermont, 2005; Kleinjung et al., 2005). Six patients with chronic tinnitus were examined by Haller, Birbaumer, & Veit (2010). After localizing the primary auditory cortex, patients were instructed to down-regulate the BOLD activity of this area aided by contingent visual feedback. Most of the patients learned to down-regulate their activations in the auditory ROI. Furthermore, a linear decrease in the auditory activations was detected over training sessions. After a single day of BCI training there was a decrease of the subjective report of tinnitus in two of six participants. This preliminary data suggests that fMRI–BCI potentially could produce beneficial effects for the treatment of this disorder.

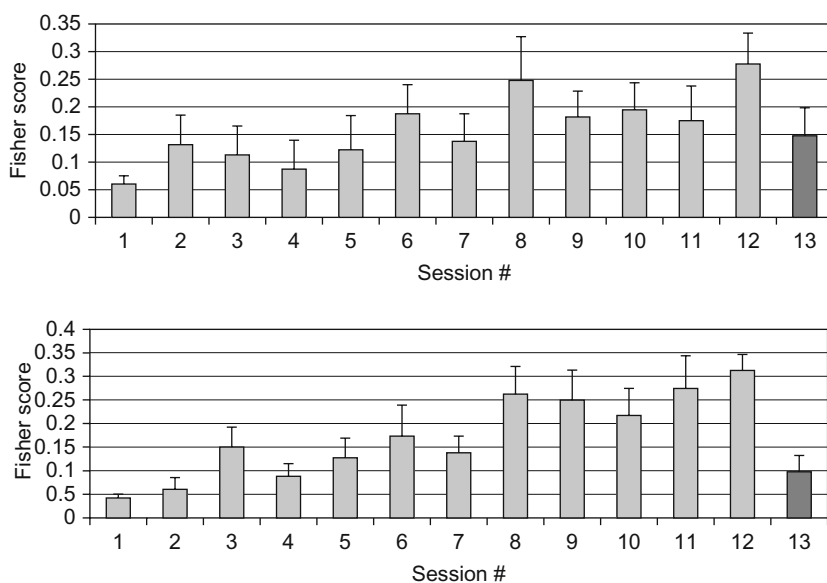
s0095 **Mental Disorders**

p0185 So far, two chronic and irreversible mental disorders have been the focus of interest in our group: psychopathy and schizophrenia. Psychopathy is a severe personality disorder often considered untreatable. Studies by our group have shown that persons diagnosed with psychopathy fail to activate prefrontal cortex and limbic regions (including insula cortex, cingulate cortex, left amygdale, and orbitofrontal cortex) during a fear-conditioning task (Birbaumer et al., 2005; Veit et al., 2002). Therefore, we hypothesized that upregulation of these areas could facilitate the acquisition of aversive conditioning, potentially modifying the behavioral manifestations of the disorder. Using a similar paradigm as Caria et al. (2007), we trained individuals with criminal psychopathy to self-regulate anterior insular cortex with fMRI–BCI. Preliminary results showed for the first time that such persons can learn self-regulation of left anterior insula. Furthermore, self-regulation led to an increase in the effective connectivity of the brain network involved in emotional processing.

p0190 In a second study, we made the first attempt to apply fMRI–BCI with schizophrenic patients (Ruiz et al., 2008; full article in press). The first aims of this study were to evaluate whether schizophrenic subjects can achieve volitional regulation of anterior insula cortex activity by fMRI–BCI training, and to explore the relationship between the capability to self-regulate and other aspects of their symptomatology. Insula cortex was chosen as the ROI based on the increasing evidence that insula

dysfunction might be critically involved in different aspects of schizophrenic psychopathology. Secondly, we explored whether self-regulation is associated with a behavioral modification of facial emotion recognition. Finally, we studied whether learned self-regulation can modulate the functional connectivity of the emotional brain network (measured by Granger causality modelling). Nine chronic schizophrenic patients, moderately symptomatic and under antipsychotic medication, were recruited. The training consisted of twelve sessions of fMRI-BCI during which patients were trained using online visual feedback of bilateral anterior insula activity. Our results showed that after a few sessions of training, patients learned to self-regulate the BOLD response in the insula cortex (see Figure 9.3). Self-regulation was not achieved, however, in a later session conducted without fMRI-feedback (transfer session).

p0195 The capability to self-regulate was negatively correlated with the severity of negative symptoms and the duration of the illness. After learned insula self-regulation, patients detected significantly more disgust faces, in line with the extensive evidence of the role of insula in face disgust recognition. However, for reasons that need more exploration,



#0020 **Figure 9.3** Group analysis of self-regulation in the left and right anterior insula. Black bars represent the Fisher score (as a measure of signal change in the ROI) and standard error across training sessions. The white bars represent the mean and standard error for the transfer session.

patients detected less happy faces during self-regulation. Volitional control of insula was also associated with a modulation of the perception of emotion intensity. Finally, volitional self-regulation led to a significant enhancement of the effective connectivity arising from the insula cortex on both hemispheres, and of the emotional network in general.

p0200 These results showed that with adequate training, these schizophrenic subjects were able to learn volitional regulation of the insula cortex by fMRI–BCI. Learned self-regulation led to changes in the perception of emotional faces, one of the hallmarks of schizophrenic dysfunction, thus providing evidence that behavioral modulation by this new technique in schizophrenia is possible. The enhancement of the connectivity in brain emotional network suggests that fMRI–BCI may be useful in “re-connecting” abnormal neural connections in schizophrenia.

s0100 CONCLUSIONS

p0205 New developments in computer and device technology, and signal processing have brought fresh enthusiasm and interest in real-time neuroimaging. Several new applications of this technique, most significantly led by research in brain–computer interfaces, are being developed and tested. rtfMRI has inspired a promising new approach to cognitive neuroscience. There is a growing evidence that learned control of the local brain activity through rtfMRI can be used as an independent variable to observe its effects on behavior. fMRI–BCI has enabled anatomically specific control of subcortical and cortical areas (some of them not accessible to electrophysiological methods), such as amygdala, insular regions, cingulate regions, and sensorimotor cortex. Encouraged by the behavioral modifications following self-regulation training, there have been several new attempts to apply this methodology to neuropsychiatric disorders.

p0210 The above studies and novel developments of BCI methodology can open up opportunities for studies in psychiatric populations and possible future therapeutic applications. However, before the full clinical application of fMRI–BCI, some important aspects have to be addressed and explored by further studies. fMRI–BCI is an expensive and difficult to implement technology. Furthermore, to date none of the mentioned studies has convincingly shown that self-regulation can be generalized “out of the scanner setting”, without the help of on-going contingent feedback. Whether the behavioral changes produced by fMRI–BCI are more than a short-term effect, has yet to be explored.

p0215 Finally, these findings have opened a fundamental question, that is, “How does learned regulation of the BOLD signal in the brain influence behavior?” A clearer understanding of the neural mechanisms underlying the BOLD response perhaps will lead us to answer this question. Perhaps fMRI–BCI will, itself, be employed to help understand the relationships between neural activity, the BOLD response and behavior.

REFERENCES

AU 1

- Amodio, D. M., & Frith, C. D. (2006). Meeting of minds: the medial frontal cortex and social cognition (Review). *Nature Reviews Neuroscience*, 7(4), 268–277.
- Andersson, J. L., Hutton, C., Ashburner, J., Turner, R., & Friston, K. J. (2001). Modeling geometric deformations in EPI time series. *Neuroimage*, 13(5), 903–919.
- Augustine, J. R. (1996). Circuitry and functional aspects of the insular lobe in primates including humans. *Brain Research Brain Research Reviews*, 22, 229–244.
- Bagarinao, E., Matsuo, K., Nakai, T., & Sato, S. (2003). Estimation of general linear model coefficients for real-time application. *Neuroimage*, 19(2 Pt 1), 422–429.
- Bandettini, P. A., Wong, E. C., Hinks, R. S., et al. (1992). Time course EPI of human brain function during task activation. *Magnetic Resonance in Medicine*, 25(2), 390–397.
- Birbaumer, N. (2006). Breaking the silence: Brain–computer interfaces (BCI) for communication and motor control. *Psychophysiology*, 43, 517–532.
- Birbaumer, N., Veit, R., Lotze, M., Erb, M., Hermann, C., Grodd, W., et al. (2005). Deficient fear conditioning in psychopathy: A functional magnetic resonance imaging study. *Archives of General Psychiatry*, 62, 799–805.
- Brown, M. A., & Semelka, R. C. (1999). *MRI basic principles and applications*. Hoboken, NJ: Wiley–Blackwell.
- Buchanan, T. W. (2007). Retrieval of emotional memories. *Psychology Bulletin*, 133, 761–779.
- Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences*, 4, 215–222.
- Buxton, R. B., Uludağ, K., Dubowitz, D. J., & Liu, T. T. (2004). Modeling the hemodynamic response to brain activation. *Neuroimage*, 23(Suppl 1), S220–S233.
- Caria, A., Sitaram, R., Veit, R., Begliomini, C., & Birbaumer, N. (2010). Volitional control of anterior insula activity modulates the response to aversive stimuli: A real-time functional magnetic resonance imaging study. *Biological Psychiatry*, in press.
- Caria, A., Veit, R., Sitaram, R., Lotze, M., Weiskopf, N., Grodd, W., et al. (2007). Regulation of anterior insular cortex activity using real-time fMRI. *Neuroimage*, 35, 1238–1246.
- Cox, D. D., & Savoy, R. L. (2003). Functional magnetic resonance imaging (fMRI) brain reading: Detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage*, 19(2 Pt 1), 261–270.
- Cox, R. W., & Jesmanowicz, A. (1999). Real-time 3D image registration for functional MRI. *Magnetic Resonance in Medicine*, 42(6), 1014–1018.
- Cox, R. W., Jesmanowicz, A., & Hyde, J. S. (1995). Real-time functional magnetic resonance imaging. *Magnetic Resonance in Medicine*, 33(2), 230–236.
- Craig, A. D. (2009). How do you feel – now? The anterior insula and human awareness. *Nature Reviews Neuroscience*, 10, 59–70.
- deCharms, R. C., Christoff, K., Glover, G. H., Pauly, J. M., Whitfield, S., & Gabrieli, J. D. (2004). Learned regulation of spatially localized brain activation using real-time fMRI. *Neuroimage*, 21, 436–443.

- deCharms, R. C., Maeda, F., Glover, G. H., Ludlow, D., Pauly, J. M., Soneji, D., et al. (2005). Control over brain activation and pain learned by using real-time functional MRI. *Proceedings of the National Academy of Sciences of the U S A*, *102*, 18 626–18 631.
- Eggermont, J. J. (2005). Tinnitus: Neurobiological substrates. *Drug Discovery Today*, *10*, 1283–1290.
- Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage*, *6*, 218–229.
- Fuchs, T., Birbaumer, N., Lutzenberger, W., Gruzelier, J. H., & Kaiser, J. (2003). Neurofeedback treatment for attention-deficit/hyperactivity disorder in children: A comparison with methylphenidate. *Applied Psychophysiology and Biofeedback*, *28*, 1–12.
- Gao, K., & Posse, S. (2003). TurboFIRE: Real-time fMRI with automated spatial normalization and Talairach Daemon database [Abstract]. 9th Annual Meeting of the OHBM, New York, USA.
- Gembris, D., Taylor, J. G., Schor, S., Frings, W., Suter, D., & Posse, S. (2000). Functional magnetic resonance imaging in real-time (FIRE), sliding-window correlation analysis and reference-vector optimization. *Magnetic Resonance in Medicine*, *43* (2), 259–268.
- Grezes, J., & Decety, J. (2001). Functional anatomy of execution, mental simulation, observation, and verb generation of actions: A meta-analysis. *Human Brain Mapping*, *12*, 1–19.
- Haller, S., Birbaumer, N., & Veit, R. (2010). Real-time fMRI feedback training may improve chronic tinnitus. *European Radiology*, *20*, 696–703.
- Hajnal, J. V., Myers, R., Oatridge, A., Schwieso, J. E., Young, I. R., & Bydder, G. M. (1994). Artifacts due to stimulus correlated motion in functional imaging of the brain. *Magnetic Resonance in Medicine*, *31*(3), 283–291.
- Hamilton, P., Glover, G. H., Hsu, J., & Johnson, R. (2010). Modulation of subgenual anterior cingulate cortex activity with real-time neurofeedback. *Human Brain Mapping*. **AU: 2**
- Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*(7238), 632–635.
- Haynes, J. D., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, *7*(7), 523–534.
- Haynes, J. D., Sakai, K., Rees, G., Gilbert, S., Frith, C., & Passingham, R. E. (2007). Reading hidden intentions in the human brain. *Current Biology*, *17*(4), 323–328.
- Hinterberger, T., Neumann, N., Pham, M., Kübler, A., Grether, A., Hofmayer, N., et al. (2004). A multimodal brain-based feedback and communication system. *Experimental Brain Research*, *154*(4), 521–526.
- Jezzard, P., & Balaban, R. S. (1995). Correction for geometric distortion in echo planar images from B0 field variations. *Magnetic Resonance in Medicine*, *34*(1), 65–73.
- Kamitani, Y., & Tong, F. (2005). Decoding the visual and subjective contents of the human brain. *Nature Neuroscience*, *8*(5), 679–685.
- Kleinjung, T., Eichhammer, P., Langguth, B., Jacob, P., Marienhagen, J., Hajak, G., et al. (2005). Long-term effects of repetitive transcranial magnetic stimulation (rTMS) in patients with chronic tinnitus. *Otolaryngology Head and Neck Surgery*, *132*, 566–569.
- Kotchoubey, B., Strehl, U., Uhlmann, C., Holzappel, S., König, M., Froscher, W., et al. (2001). Modification of slow cortical potentials in patients with refractory epilepsy: A controlled outcome study. *Epilepsia*, *42*, 406–416.
- Kubler, A., Kotchoubey, B., Kaiser, J., Wolpaw, J. R., & Birbaumer, N. (2001). Brain–computer communication: Unlocking the locked in. *Psychology Bulletin*, *127*, 358–375.

- Kybic, J., Thevenaz, P., Nirkko, A., & Unser, M. (2000). Unwarping of unidirectionally distorted EPI images. *IEEE Transactions on Medical Imaging*, 19(2), 80–93.
- LaConte, S. M., Peltier, S. J., & Maciuba, A. (2007). Real-time fMRI using brain-state classification. *Human Brain Mapping*, 28(10), 1033–1044.
- Lang, P. J. (Ed.), (1980). Behavior treatment and bio-behavioral assessment: Computer applications. *Technology in Mental Health Care Delivery Systems* Norwood, NJ: Ablex.
- Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (Eds.), (1997). *International Affective Picture System (IAPS): technical manual and effective ratings*. The Center for Research in Psychophysiology. Gainesville: University of Florida.
- Lawrie, S. M., Whalley, H. C., Job, D. E., & Johnstone, E. C. (2003). Structural and functional abnormalities of the amygdala in schizophrenia. *Annals of the New York Academy of Sciences*, 985, 445–460.
- Lee, S., Halder, S., et al. (2010). Effective functional mapping of fMRI data with support-vector machines. *Human Brain Mapping*, [Epub ahead of print].
- Lee, J. H., Ryu, J., Jolesz, F. A., Cho, Z. H., & Yoo, S. S. (2009). Brain-machine interface via real-time fMRI: Preliminary study on thought-controlled robotic arm. *Neuroscience Letters*, 450, 1–6.
- Lee, S., Sitaram, R., Ruiz, S., & Birbaumer, N. (2008). Measure of neurofeedback effects in an fMRI Brain-Computer Interface with support vector machine and Granger causality model. Xth International Conference on Cognitive Neuroscience, Bodrum, Turkey.
- Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI (Review). *Nature*, 453, 869–878.
- Lozano, A. M., Mayberg, H. S., Giacobbe, P., Hamani, C., Craddock, R. C., & Kennedy, S. H. (2008). Subcallosal cingulate gyrus deep brain stimulation for treatment-resistant depression. *Biological Psychiatry*, 64, 461–467.
- Mackey, S. C., & Maeda, F. (2004). Functional imaging and the neural systems of chronic pain. *Neurosurgery Clinics of North America*, 15, 269–288.
- Mathiak, K., & Posse, S. (2001). Evaluation of motion and realignment for functional magnetic resonance imaging in real time. *Magnetic Resonance Medicine*, 45(1), 167–171.
- Mayberg, H. S., Lozano, A. M., Voon, V., McNeely, H. E., Seminowicz, D., Hamani, C., et al. (2005). Deep brain stimulation for treatment-resistant depression. *Neuron*, 45, 651–660.
- Mitchell, T. M., Hutchinson, R., et al. (2003). Classifying instantaneous cognitive states from FMRI data. *AMIA Annual Symposium Proceedings*: 465–9.
- Murase, N., Duque, J., Mazzocchio, R., & Cohen, L. G. (2004). Influence of interhemispheric interactions on motor function in chronic stroke. *Annals of Neurology*, 55, 400–409.
- Nagai, M., Kishi, K., & Kato, S. (2007). Insular cortex and neuropsychiatric disorders: A review of recent literature. *European Psychiatry*, 22, 387–394.
- Naqvi, N. H., & Bechara, A. (2009). The hidden island of addiction: The insula. *Trends in Neuroscience*, 32, 56–67.
- Panych, L. P., & Jolesz, F. A. (1994). A dynamically adaptive imaging algorithm for wavelet-encoded MRI. *Magnetic Resonance in Medicine*, 32(6), 738–748.
- Pause, B. M., Jungbluth, C., Adolph, D., Pietrowsky, R., & Dere, E. (2010). Induction and measurement of episodic memories in healthy adults. *Journal of Neuroscientific Methods*, 189(1), 88–96.
- Petrovic, P., & Ingvar, M. (2002). Imaging cognitive modulation of pain processing. *Pain*, 95, 1–5.
- Pollock, D. S. G., Green, R. C., & Nguyen, T. (1999). *Handbook of Time Series Analysis, Signal Processing, and Dynamics* San Diego, CA: Academic Press.

- Posse, S., Binkofski, F., Schneider, F., Gembris, D., Frings, W., et al. (2001). A new approach to measure single-event related brain activity using real-time fMRI: Feasibility of sensory, motor, and higher cognitive tasks. *Human Brain Mapping*, *12*(1), 25–41.
- Posse, S., Fitzgerald, D., Gao, K., Habel, U., Rosenberg, D., Moore, G. J., et al. (2003). Real-time fMRI of temporolimbic regions detects amygdala activation during single-trial self-induced sadness. *Neuroimage*, *18*(3), 760–768.
- Posse, S., Wiese, S., Gembris, D., Mathiak, K., Kessler, C., Grosse-Ruyken, B., et al. (1999). Enhancement of BOLD-contrast sensitivity by single-shot multi-echo functional MR imaging. *Magnetic Resonance in Medicine*, *42*(1), 87–97.
- Rota, G., Sitaram, R., Veit, R., Erb, M., Weiskopf, N., Dogil, G., et al. (2009). Self-regulation of regional cortical activity using real-time fMRI: The right inferior frontal gyrus and linguistic processing. *Human Brain Mapping*, *30*, 1605–1614.
- Ruiz, S., Sitaram, R., Sangyun, L., Caria, A., Soekadar, S., Veit, R., et al. (2008). Learned control of insular activity and functional connectivity changes using a fMRI Brain Computer Interface in Schizophrenia. 38th annual meeting of the Society for Neuroscience. Washington. *Schizophrenia Research* 102/1-3, 92 [full print version in press].
- Seth, A. K. (2009). A MATLAB toolbox for Granger causal connectivity analysis. *Journal of Neuroscience Methods*. **AU: 3**
- Sitaram, R., Lee, S., et al. (2010). Real-time support vector classification and feedback of multiple emotional brain states. *Neuroimage*, in press.
- Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis* New York: Appleton-Century-Crofts.
- Smysler, C., Grabowski, T. J., Frank, R. J., Haller, J. W., & Bolinger, L. (2001). Real-time multiple linear regression for fMRI supported by time-aware acquisition and processing. *Magnetic Resonance in Medicine*, *45*(2), 289–298.
- Strehl, U., Leins, U., Goth, G., Klinger, C., Hinterberger, T., & Birbaumer, N. (2006). Self-regulation of slow cortical potentials: A new treatment for children with attention-deficit/hyperactivity disorder. *Pediatrics*, *118*, e1530–1540.
- Studholme, C., Constable, R. T., & Duncan, J. S. (2000). Accurate alignment of functional EPI data to anatomical MRI using a physics-based distortion model. *IEEE Trans Med Imaging*, *19*(11), 1115–1127.
- Ture, U., Yasargil, D. C., Al-Mefty, O., & Yasargil, M. G. (1999). Topographic anatomy of the insular region. *Journal of Neurosurgery*, *90*, 720–733.
- Veit, R., Flor, H., Erb, M., Hermann, C., Lotze, M., Grodd, W., et al. (2002). Brain circuits involved in emotional learning in antisocial behavior and social phobia in humans. *Neuroscience Letters*, *328*, 233–236.
- Voyvodic, J. T. (1999). Real-time fMRI paradigm control, physiology, and behavior combined with near real-time statistical analysis. *Neuroimage*, *10*(2), 91–106.
- Wager, T. D., Rilling, J. K., Smith, E. E., Sokolik, A., Casey, K. L., Davidson, R. J., et al. (2004). Placebo-induced changes in FMRI in the anticipation and experience of pain. *Science*, *303*, 1162–1167.
- Weiskopf, N., Klose, U., Birbaumer, N., & Mathiak, K. (2005). Single-shot compensation of image distortions and BOLD contrast optimization using multi-echo EPI for real-time fMRI. *Neuroimage*, *24*(4), 1068–1079.
- Weiskopf, N., Mathiak, K., Bock, S. W., Scharnowski, F., Veit, R., Grodd, W., et al. (2004). Principles of a brain-computer interface (BCI) based on real-time functional magnetic resonance imaging (fMRI). *IEEE Transactions on Biomedical Engineering*, *51*(6), 966–970.

- Weiskopf, N., Scharnowski, F., Veit, R., Goebel, R., & Birbaumer, N. (2004). Self-regulation of local brain activity using real-time functional magnetic resonance imaging (fMRI). *Journal of Physiology Paris*, *98*(4-6), 357–373.
- Weiskopf, N., Veit, R., Erb, M., Mathiak, K., Grodd, W., & Goebel, R. (2003). Physiological self-regulation of regional brain activity using real-time functional magnetic resonance imaging (fMRI), methodology and exemplary data. *Neuroimage*, *19*, 577–586.
- Yoo, S. S., Fairney, T., Chen, N. K., Choo, S. E., Panych, L. P., Park, H., et al. (2004). Brain-computer interface using fMRI: Spatial navigation by thoughts. *Neuroreport*, *15*, 1591–1595.
- Yoo, S. S., Guttman, C. R., Zhao, L., & Panych, L. P. (1999). Real-time adaptive functional MRI. *Neuroimage*, *10*(5), 596–606.
- Yoo, S. S., & Jolesz, F. A. (2002). Functional MRI for neurofeedback: Feasibility study on a hand motor task. *Neuroreport*, *13*, 1377–1381.
- Yoo, S. S., O'Leary, H. M., Fairney, T., Chen, N. K., Panych, L. P., Park, H., et al. (2006). Increasing cortical activity in auditory areas through neurofeedback functional magnetic resonance imaging. *Neuroreport*, *17*, 1273–1278.
- Zaitsev, M., Hennig, J., Speck, O., et al. (2004). Point spread function mapping with parallel imaging techniques and high acceleration factors: Fast, robust, and flexible method for echo-planar imaging distortion correction. *Magnetic Resonance in Medicine*, *52*(5), 1156–1166.

s0105 **UNCITED REFERENCES**

p0220 Amodio and Frith (2006); Hinterberger et al. (2004)

Author Queries

Location in Article	Query / remark
AU:1	Please complete missing reference details
AU:2	{details?}
AU:3	{vol and page nos}
AU:4	Give author-date source as well
AU:5	Please provide Figure citation 9.2

4. 2. Effective functional mapping of fMRI data with support-vector machines

Effective Functional Mapping of fMRI Data with Support-Vector Machines

Sangkyun Lee,^{1,2*} Sebastian Halder,¹ Andrea Kübler,^{1,3}
Niels Birbaumer,^{1,4,5} and Ranganatha Sitaram¹

¹Institute of Medical Psychology and Behavioral Neurobiology, University of Tübingen,
Gartenstr. 29, 72074 Tübingen, Germany

²Graduate School of Neural and Behavioral Sciences, International Max Planck Research School,
72074 Tübingen, Germany

³Department of Psychology I, University of Würzburg, Marcusstr. 9-11, 97070 Würzburg, Germany

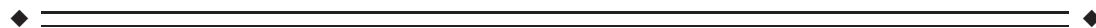
⁴Ospedale San Camillo, Istituto di Ricovero e Cura a Carattere Scientifico, Venezia—Lido, Italy

⁵Department of Psychology, Biological Psychology, Universidad de las Islas Baleares, Spain



Abstract: There is a growing interest in using support vector machines (SVMs) to classify and analyze fMRI signals, leading to a wide variety of applications ranging from brain state decoding to functional mapping of spatially and temporally distributed brain activations. Studies so far have generated functional maps using the vector of weight values generated by the SVM classification process, or alternatively by mapping the correlation coefficient between the fMRI signal at each voxel and the brain state determined by the SVM. However, these approaches are limited as they do not incorporate both the information involved in the SVM prediction of a brain state, namely, the BOLD activation at voxels and the degree of involvement of different voxels as indicated by their weight values. An important implication of the above point is that two different datasets of BOLD signals, presumably obtained from two different experiments, can potentially produce two identical hyperplanes irrespective of their differences in data distribution. Yet, the two sets of signal inputs could correspond to different functional maps. With this consideration, we propose a new method called *Effect Mapping* that is generated as a product of the weight vector and a newly computed vector of mutual information between BOLD activations at each voxel and the SVM output. By applying this method on neuroimaging data of overt motor execution in nine healthy volunteers, we demonstrate higher decoding accuracy indicating the greater efficacy of this method. *Hum Brain Mapp* 00:000–000, 2010. © 2010 Wiley-Liss, Inc.

Key words: fMRI; multivariate analysis; multivariate pattern analysis; support vector machine



INTRODUCTION

Pattern-based methods use sophisticated machine learning techniques, such as multilayer neural networks and

support vector machines to discriminate spatial, temporal, and spectral patterns in a system. Such methods have been successfully used in character recognition, speech

Contract grant sponsor: Deutsche Forschungsgemeinschaft (DFG); Contract grant numbers: DFG BI 195/59-1, DFG BI 195/56-1, KU 1453/3-1; Contract grant sponsor: SFB 437 (Kriegserfahrungen).

*Correspondence to: Sangkyun Lee, Institute of Medical Psychology and Behavioral Neurobiology, Eberhard-Karls-University of Tübingen, Gartenstr. 29, D-72074 Tübingen, Germany.
E-mail: sangkyun.lee@student.uni-tuebingen.de

Received for publication 7 July 2009; Revised 22 October 2009; Accepted 23 October 2009

DOI: 10.1002/hbm.20955

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

recognition, and image recognition applications [Jain et al., 2000]. Rapid progress in the application of data mining and statistical techniques and the growth of computing power has enabled the efficient manipulation and handling of large amounts of neuroimaging data, acquired from high resolution brain scans of several time points, for multivariate pattern analysis [Haynes and Rees, 2006; Norman et al., 2006]. Multivariate pattern classification and analysis methods have been used with great success in several neuroimaging studies, including unconscious antecedents of free decisions [Soon et al., 2008], lie detection [Davatzikos et al., 2005], visual processing [Haxby et al., 2001; Kamitani and Tong, 2005], and emotion [Mourao-Miranda et al., 2007].

In contradistinction to univariate analysis which evaluates each brain location separately although brain activity was measured from many thousands of locations simultaneously, multivariate analysis is based on the insight that multiple, spatially distributed regions act in consort during a task. Pattern analysis methods also provide an objective criterion for determining the importance of different brain regions in a given task by simply comparing the accuracy of decoding a task from signals extracted from individual brain regions or a group of brain regions [Haynes et al., 2007; Soon et al., 2008]. Pattern classification can be used not only to separate different task conditions or brain states, but also to test consistency of brain activation across tasks or sessions, and to track temporal transitions of brain states [Polyn et al., 2005].

Among different pattern classification methods, SVMs are one of the most widely used methods for fMRI signals [LaConte et al., 2005; Mourao-Miranda et al., 2005, 2006, 2007]. Support vector machines (SVMs) are a set of supervised learning methods used for classification and regression. By considering input data as two sets of vectors in an M -dimensional space, a linear SVM will construct a separating hyperplane in that space. A good separation is achieved by maximization of the *margin*, whose boundary is the distance to the separating hyperplane from the input vectors (i.e., support vectors) closest to it [Schölkopf and Smola, 2002; Schölkopf et al., 1999; Vapnik, 1995]. The criterion can be denoted as a quadratic optimization problem and the best solution can therefore be found by applying optimization theory. The advantage of SVMs in real-world applications is their superior performance in classification accuracy with small sample sizes and high dimensional inputs.

Some applications of SVM to fMRI signals generated functional maps by displaying the weight value at each voxel [Mourao-Miranda et al., 2005, 2006, 2007]. The studies maintained that the weight vector can identify the most discriminating voxels by multivariate analysis since the weight vector is the direction along which the input vectors from two conditions differ most. These studies considered only one (i.e., the weight vector) of the two factors that determining an SVM output, namely, the weight

vector and the input vector. It needs to be highlighted that SVMs are trained on the input samples to minimize the classification-error rates by computing the weight vector solely from support vectors that reside near the border of the hyperplane. It follows from this consideration that the weight vector so computed is not completely influenced by the statistical distribution of input vectors. As such, using the weight vector alone to generate functional maps is equivalent to presenting only a part of the information.

LaConte et al. used the pair-wise correlation between the BOLD signal at each voxel and distance from the margin [LaConte et al., 2005]. This approach derives from the consideration that the distance from the separating hyperplane is related to ease of discrimination, and based on the intuition that the sample closest to the hyperplane is most difficult to classify. Although the distance is related to discrimination, it does not follow that the farthest sample is the most important for discrimination as it may contain very little information about data distributions of the tasks (see Theory for more detail). Additionally, this approach does not exploit the advantage of multivariate analysis of the SVM due to its univariate measure.

In this study, we propose a new functional-mapping method to identify the voxels more closely related to the actual importance in classification. The method incorporates information from both the weight vector and the input vectors that together determine the SVM output. Toward this end, we first derive the formula for a new quantity called the effect value (EV). Effect value for a single voxel is defined as the statistical relation (mutual information; see Theory for detail) between the voxel and the SVM output, multiplied by the corresponding weight value of the SVM. Subsequently, we compare the proposed method with the functional maps generated by the previous methods. The comparison is done quantitatively by evaluating classification performance from the voxels identified as informative in competing functional mapping approaches generated based on data acquired during overt motor execution. We chose fMRI data from the overt motor execution task for this investigation, as this task is easier to execute consistently across runs and among all healthy volunteers, so that differences in classification accuracy is more readily attributable to differences in the performance of the classifiers *per se* than due to artifactual effects of unreliable data.

THEORY

First, we summarize the basic concept of SVM and the procedure for applying SVM to fMRI data. Next, based on considerations of the theoretical basis of SVM and the limitations of the conventional interpretation of SVM, we propose the new method.

Support Vector Machine (SVM)

In a typical SVM analysis of fMRI signals, BOLD values from all brain voxels of each repetition time (TR) are contained in the M -dimensional (M : number of all the brain voxels) input vector \mathbf{x}^i (notation: $\mathbf{X}^{\text{index of sample}}$; bold font indicates a column vector, $x_{\text{index of voxel}}^{\text{index of sample}}$). SVM determines a scalar class label L^i from \mathbf{x}^i as follows:

$$L^i = \text{sgn}(y^i = \mathbf{w}^T \mathbf{x}^i + b), \quad (1)$$

$$i = 1, \dots, N, (N : \text{number of input vectors})$$

where the weight vector \mathbf{w} and the constant value b , which are estimated by a SVM training algorithm from the training dataset, define a linear decision boundary, T is transpose of a vector, $\text{sgn}(\cdot)$ is a sign function, $\text{sgn}(x) = +1, 0, -1$ if $x > 0, x = 0, x < 0$, respectively.

When the input vectors \mathbf{x}^i and the design labels L_D^i (if the input vector comes from a condition of interest, then $L_D^i = 1$; on the other hand if the input vector comes from a rest condition or a control condition, then $L_D^i = -1$) are taken from the training dataset, the linear SVM algorithm attempts to find a separating hyperplane $y = \mathbf{w}^T \mathbf{x} + b = 0$ in the feature space. The weight vector \mathbf{w} of a linear SVM is obtained by minimizing objective function of Eq. (2) with constraints Eqs. (3) and (4),

$$\frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N \xi^i, \quad (2)$$

$$\text{with } L_D^i (\mathbf{w}^T \mathbf{x}^i + b) \geq 1 - \xi^i, \quad (3)$$

$$\text{and } \xi^i \geq 0 \quad (4)$$

where the slack variable ξ^i is introduced to describe a non-separable case (i.e., data that cannot be separated without classification error), C denotes the weighting on the slack variable (i.e., the extent to which misclassification is allowed). The minimization of Eq. (2) is originated from concept of the maximization of the *margin* (length of the $\text{margin} = 2/\|\mathbf{w}\|^2$), whose boundaries are defined as $y = \mathbf{w}^T \mathbf{x} + b = \pm 1$ built from support vectors (i.e., $\text{SV} = \{\mathbf{x}^i | y^i = \mathbf{w}^T \mathbf{x}^i + b = \pm 1\}$) in each class.

The main objective function Eq. (2) and constraint terms Eqs. (3) and (4) can be combined into one nonconstraint form by the introduction of a Lagrange multiplier. From the formula, the weighted vector \mathbf{w} can be obtained:

$$\mathbf{w} = \sum_{i=1}^N \alpha^i L_D^i \mathbf{x}^i \quad (5)$$

Here, α^i is the Lagrange multiplier and its value determines whether the input vector \mathbf{x}^i is a support vector or

not. When α^i is nonzero, the corresponding input vector is the support vector.

Interpretation of SVM Results

In fMRI studies using SVM, an intuitive way to analyze the results of SVM training might be to overlay the weight vector onto brain images [Mourao-Miranda et al., 2005]. Although this method describes which weight value produces the larger effects, the method is limited due to the fact that the output y is not estimated from the weight vector alone, but also from the input vector containing the BOLD values at each voxel. In addition, this method has disadvantages as different datasets pertaining to different tasks may generate identical hyperplanes (see Fig. 1A).

LaConte et al. used feature space weighting (FSW) to generate the functional maps from the SVM results [LaConte et al., 2005]. This approach used a distance measure from the separating hyperplane to the estimated output of SVM, leading to a weighted average contrast function (i.e., the contrast value at each time weighted by distance measure from the margin). Although they reported a relationship between the distance and discriminability, theoretically, the distance of the input vectors from the separating hyperplane is not completely representative of the given task (see Fig. 1B). In addition, generation of the functional maps by using pixel-wise correlation between single voxel activation and the contrast function might not accurately represent the distinction between the classified brain states because it does not take into account the multivariate contribution of the weight vector to the SVM output.

Typically, SVM maximizes the distance between two hyperplanes composed from support vectors of each class without considering the data distribution of all the input vectors. As seen in Eq. (5), the weight vector is a weighted sum of only the support vectors and does not incorporate any information from the nonsupport vectors. In addition, when the input vectors of one condition (+1 or -1) are considered, one single activation pattern would represent the class. That is, many repetitions of one task lead to similar activation patterns in the brain regions associated with the task (i.e., small variance in elements of a class or higher signal-to-noise ratio (SNR)). In this view, important or essential elements of the input vectors from a class would be expected to show higher consistency or lower variance of activations across several samples. However, Figure 1C illustrates that the elements of the input vectors corresponding to the higher weight values do not necessarily repeat with higher consistency. In Figure 1C, three different input vectors result in similar SVM output with the same weight vector. This shows that consideration of only a single component, i.e., the weight vector, is not enough to determine the effect of individual elements of the input vector on the SVM output. This argument calls for the

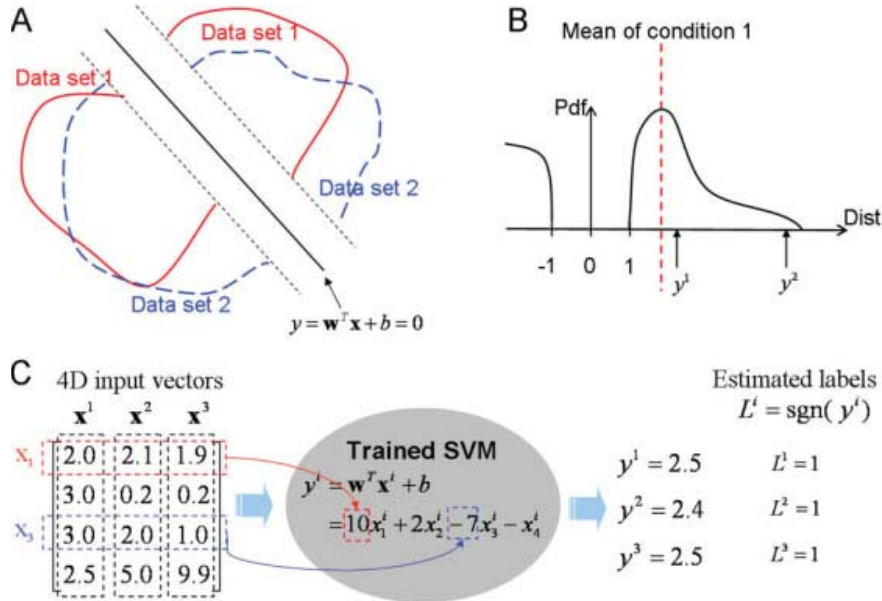


Figure 1.

Illustrations of characteristics of an SVM. **(A)** Two different datasets (red and blue) having the same separating hyperplane ($y = \mathbf{w}^T \mathbf{x} + b = 0$). Even though the two datasets have different distributions (a red line: distribution of dataset 1, a blue broken line: distribution of dataset 2), SVM trained with each dataset separately can determine the same separating hyperplane (i.e., the same weight vector). **(B)** Distribution of the SVM outputs. Distance of an input vector from the separating hyperplane is proportional to the SVM output, greater distance indicating greater separation between the conditions. In condition 1, SVM output y^1 is closer to mean or the center of distribution (a red broken line) than SVM output y^2 . Therefore, it is likely that there are more samples close to \mathbf{x}^1 than to \mathbf{x}^2 . However, in terms of data distribution, \mathbf{x}^1 is more important than \mathbf{x}^2 even though \mathbf{x}^1 is closer to the separating hyperplane than \mathbf{x}^2 . **(C)** Illustration of the determination of the SVM output with the weight vector and the input vectors. In (C), for the trained SVM model $y = 10x_1 + 2x_2 - 7x_3 - x_4$, there are three input vectors, particularly nonsupport vec-

tors, classified to the same class (+1). When the importance of each element of the input vectors is considered based on the magnitude of the corresponding weight value, the 1st (10) and 3rd (-7) elements are the two most important ones [Mourao-Miranda et al., 2005]. If the brain state of one class (+1 or -1) can be represented as one spatial pattern, importance of each element of the input vectors could be considered based on consistency of the elements. The values (2.0, 2.1, and 1.9) of the 1st element (x_1) are consistent across the input vectors, but the values (3.0, 2.0, and 1.0) of the 3rd element (x_3) do not show any consistency even if the SVM output are consistent, and the corresponding weight value is large. This shows that the importance of an element of the input vectors is not simply proportional to the magnitude of the corresponding weight value alone. Rather, the effective importance of an element depends on both the weight value and the sample data distribution considered together. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

combined application of the weight vector and the input vector in obtaining legitimate functional activations.

Effect Mapping (EM)

Effect mapping considers both the effect of the voxel activation to the output of a classifier, and the weight vector of the estimated SVM model. EM measures the effect of each voxel to the classifier output by computing mutual

information (MI) between the voxel and the output. MI is defined as the amount of information that one random variable contains about another random variable [Cover and Thomas, 1991]. That is, when two random variables X and Y occur with a joint probability mass function $p(x,y)$ and marginal probability function $p(x)$ and $p(y)$, the entro-

pies of the two random variables and the joint probability are given respectively by:

$$H(X) = \sum_{x \in X} -p(x) \log p(x),$$

$$H(Y) = \sum_{y \in Y} -p(y) \log p(y),$$

and $H(X, Y) = \sum_{x \in X} \sum_{y \in Y} -p(x, y) \log p(x, y).$

MI, $I(X;Y)$, is the relative entropy between the joint distribution and the product distribution, i.e.

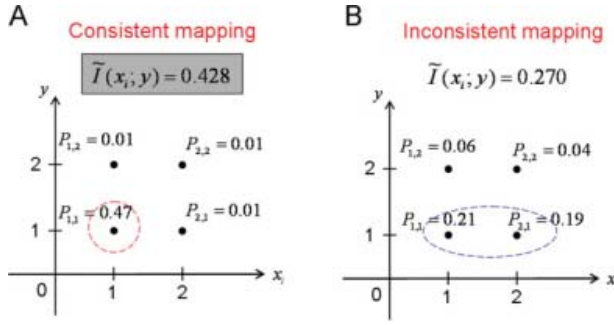


Figure 2.

Illustration of mutual information in consistent (i.e., most of x_i s are mapped into $y = 1$; a red broken line of circle of Figure **A**), and inconsistent (i.e., mapping of x_i s are inconsistent without showing a dominant probability in a point; a blue broken line of ellipse of Figure **B**) mappings. A probability of each mapping point is denoted as $P_{a,b} = p(x_i = a, y = b)$. In all the figures, data distribution is shown in one condition ($y > 0$) with the assumption that all the input vectors are correctly classified, and $p(y)$ decreases with increase of y (i.e., $p(y = 1) > p(y = 2)$) due to the fact that a SVM model in such high dimensional data as fMRI data has many support vectors, and many of SVM outputs reside in a region close to hyperplanes, $y = \mathbf{w}^T \mathbf{x} + b = \pm 1$, composed of support vectors. Figures A and B show the normalized mutual information between x_i and y . With regard to mapping of x_i s into the SVM outputs y , A shows higher probability in one point reflecting higher consistency than B, and higher mutual information than B. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

$$I(X; Y) = H(X) + H(Y) - H(X, Y) \\ = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log \frac{p(x, y)}{p(x)p(y)} \quad (6)$$

Because of the variance of mutual information based on entropies $H(X)$ and $H(Y)$, normalized mutual information is defined as [Maes et al., 1997]:

$$\tilde{I}(X; Y) = \frac{I(X; Y)}{H(X) + H(Y)} \quad (7)$$

As shown in Figure 2, $\tilde{I}(x_i; y)$ can take into consideration consistency of x_i s based on data distribution of y on the assumption that the probability density of the SVM outputs decrease with increase of y . Distribution of the SVM outputs, y , in our analysis also reflects this assumption well. In addition, $\tilde{I}(x_i; y)$ reflects the nonlinear dependencies between distribution of x_i and y better than a simple correlation method.

Hence, the effect value (EV) E_k of a voxel k , designed to take into consideration the above points, is defined as:

$$E_k = w_k \tilde{I}(x_k; y), \quad k = 1, \dots, M \quad (M : \text{number of voxels}) \quad (8)$$

where y is the SVM output after excluding the sign function, w_k and x_k are weight value and activation in voxel k , respectively.

MATERIALS AND METHODS

Participants and Experimental Protocol

We analyzed fMRI data from nine right-handed healthy college students (age: 26.4 ± 5.2). None of the participants had any history of neurological or psychiatric disorders. The study was approved by the ethics committee of the Faculty of Medicine of the University of Tuebingen. Stimuli were presented in a block design. There were two active conditions (left-hand (LH) and right-hand (RH) movements) and a resting condition. During active conditions, each participant was instructed to move his/her palm and fingers freely. Participants were asked to restrict movement above the wrist, for example, in the elbows and shoulders. Each active and inactive condition (rest state) lasted 30 s (15 scans). In our analysis, 12 active condition blocks (6 LHs + 6 RHs) from 2 runs of fMRI measurement were used to train SVM and make comparisons.

Data Acquisition

Functional images were acquired on a 3.0 T whole body scanner, with a standard 12-channel head coil (Siemens Magnetom Trio Tim, Siemens, Erlangen, Germany). A standard echo-planar imaging sequence was used (EPI; TR = 2 s, TE = 30 ms, flip angle $\alpha = 78^\circ$, bandwidth = 2.232 kHz/pixel). Thirty-two slices (voxel size = $3 \times 3 \times 3.75$ mm³, slice gap = 1 mm), AC/PC aligned in axial orientation were acquired.

Preprocessing and Classification

Preprocessing was performed with SPM5 (Wellcome Department of Imaging Neuroscience, London) and classification was performed using MATLAB (The Mathworks, Natick, MA) scripts.

We performed realignment of functional images, coregistration between functional images and structure image, and normalization of functional and structure images onto the Montreal Neurological Institute (MNI) space. After selecting nonbrain areas with a brain mask (i.e., a brain mask file "mask.img") generated with SPM5, z-normalizations (z-value: $(x - \text{mean}(x)) / \text{standard deviation}(x)$, x : samples) were applied across all the time-series of each voxel on each run of each participant data separately to correct for the variance of BOLD signals of different runs and different participants. Whole brain images from each TR were used to generate input vectors to the SVM classifier, and individual scans were classified.

The freely available SVM software SVMlight [Joachims, 1999] was used to implement the classifier. Linear SVMs were trained with a fixed regularization parameter $C = 10^5$ (i.e., hard margin SVM) to remove variability of

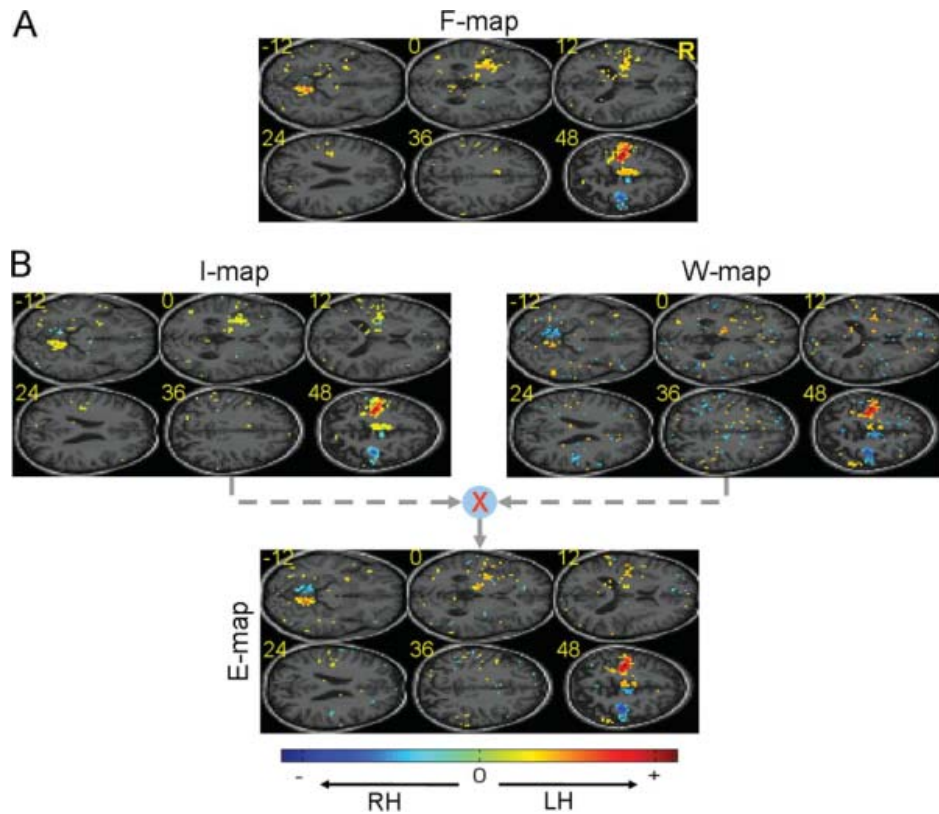


Figure 3.

Functional maps from a group analysis of non-smoothed data. In functional maps (F(SW)-, W(eight)-, and E(ffect)-maps) computed from the SVM outputs, clusters with positive value (red) correspond to the positive SVM weight values, while clusters with negative values (blue) correspond to the negative SVM weight values. In **B**, E-map is drawn after rescaling the EVs of Eq. (8) for the purpose of display (to make extreme values to be smaller) with the following formula: $|E_k| = \text{sgn}(E_k) \log(1 + |E_k|/\text{std}(|E|))$ where $\text{std}(|E|)$ is the standard deviation of all E_k . These maps show six horizontal slices every 12 mm obtained from the whole brain in

MNI space. The functional maps are drawn by selecting 5% of voxels (for the purpose of display) having the highest values in magnitude. **(A)** F-map. **(B)** I-, W-, and E- maps. The I-map shows a distributed pattern of mutual information between the input vector and the SVM output (without the sign function; $y = \mathbf{w}^T \mathbf{x} + b$). In principle, mutual information is zero or positive, but for the purpose of comparison with other methods, the values are multiplied by the sign of the SVM weight values. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

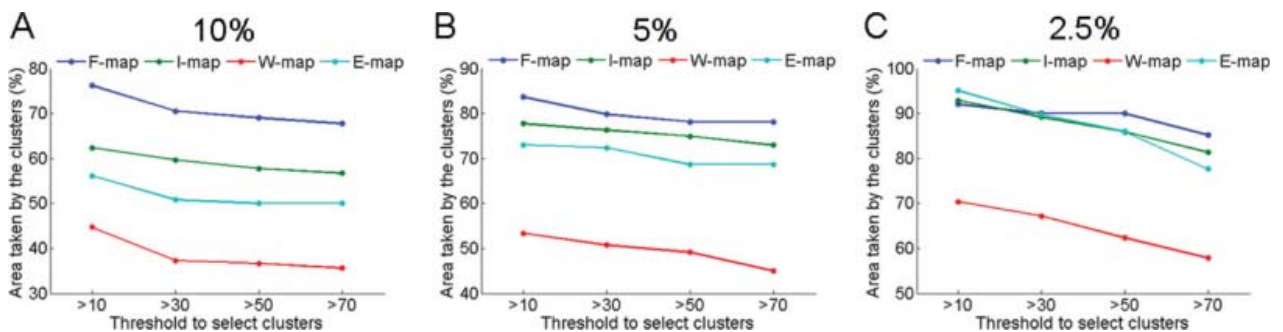


Figure 4.

Ratio of area of clusters remaining after applying the second-level threshold to total area of voxels remaining after applying the first-level threshold. The first-level thresholds 10, 5, and 2.5%, are used in **A–C**, respectively. The second-level thresholds {10, 30, 50, 70} are used in all the figures. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

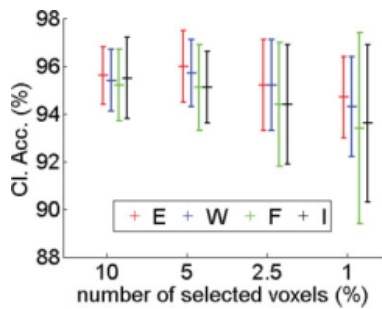


Figure 5.

Comparison of cross-validation results (i.e., classification accuracies (mean classification accuracy rate (%) \pm standard error of the mean (%))) from voxels selected by the functional maps. All figures are drawn with nine-fold CV on non-smoothed data. After selecting top 10, 5, 2.5, and 1% brain voxels in magnitude in each functional map, the selected voxels were used to evaluate classification accuracy, respectively. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

classification performance dependent on the regularization parameter C . In the classification procedure, the LH and RH were given 1 and -1 as design labels for the SVMs. Nine-fold cross validation (CV) [Hastie et al., 2001] was applied in group data of nine participants. In each fold, the data of eight participants were used to train an SVM classifier, and then the data of one remaining participant were used to test the classifier.

Computation of Probability

In the computation of mutual information, joint distribution and marginal distributions were performed with two-dimensional joint histogram and one-dimensional histogram (<http://www.cs.rug.nl/~rudymatlab/>). In our analysis, numbers of bins of the histogram were defined as $N^{1/3}$ (N : number of samples; $(8 \times 180)^{1/3} = 11$ bins).

Comparison of Functional Maps

To prevent the double dipping problem [Kriegeskorte et al., 2009], comparisons were made with nine-fold CV from group data of nine participants. In each fold, the data of eight participants were used to train an SVM, and to compute functional maps (i.e., F(SW)-, I-, W(eight)-, E(ffect)-maps) (see Fig. 3). To compare the capability of functional maps for detecting important spatial patterns of brain activity, the top $x\%$ of brain voxels (e.g., 5%) having the highest functional values in magnitude was selected from the functional maps. Then, the selected voxels in the data of eight participants were used to train a linear SVM

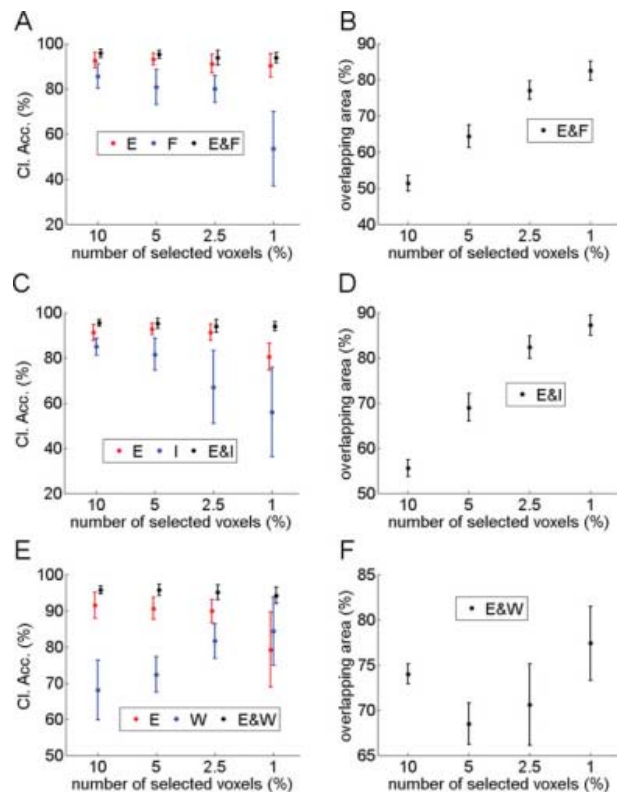


Figure 6.

Comparison of cross-validation results (i.e., classification accuracies; mean classification accuracy rate (%) \pm standard error of the mean (%)) for overlapping and non-overlapping areas of two different types of functional map. All figures are drawn with nine-fold CV on nonsmoothed data. In **A**, **C**, and **E**, after selecting top 10, 5, 2.5, and 1% brain voxels in magnitude from two functional maps, overlapping and nonoverlapping area (based on the concept of “relative complement”) were used to evaluate classification accuracy, respectively. In **B**, **D**, and **F**, after selecting top 10, 5, 2.5, and 1% brain voxels in magnitude from two functional maps, ratios of area taken by the voxels commonly selected from the two functional maps to the total area taken by all the selected voxels were shown. In each figure, legends from one functional map (i.e., E, F, I, and W), and legends from two functional maps (i.e., E and F, E and I, and E and W) indicate the results from the exclusive selected voxels from one functional map and commonly selected voxels from two functional maps, respectively. (A) Performance comparison between E- and F-maps. (B) Ratio of overlapping area of E- and F-maps. (C) Performance comparison between E- and I-maps. (D) Ratio of overlapping area of E- and I-maps. (E) Performance comparison between E- and W-maps. (F) Ratio of overlapping area of E- and W-maps. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

with $C = 10^5$, and the same voxels in the data of one remaining participant were used to classify.

Additionally, when comparing between two functional maps generated by two competing methods, it is possible that the decoding accuracies of the two maps are similar due to common voxels selected by these methods, yet the two methods might indicate different sets of voxels as most informative. This consideration is critical if differences of activation patterns from two functional maps are obtained in different functional areas or on the border of two functional areas. As an example, let us assume that two patterns of voxels in the visual cortex, encompassing brain areas such as V1, V2, V3, and V4, are generated by two different functional mapping methods. It is conceivable that the two methods provide similar decoding accuracies, but one method (e.g., E-map) identifies voxels from the V1 and V3 areas as most informative, while the other (e.g., W-map) identifies brain areas from the V1 and V4 areas as most informative. Given this, the open question is: which functional area between the V3 and V4 areas is more involved in the given task, acknowledging that V1 is the common area of activation. To be able to answer this question in our study, overlapping and nonoverlapping activations from two competing maps were considered separately in the voxel selection process. One SVM classifier for overlapping areas of the two competing functional maps, and two separate SVM classifiers for two nonoverlapping areas of the two competing functional maps were trained from the BOLD activations from the voxels selected. Then, the trained SVMs were tested to classify the data of one remaining participant. For instance, in comparing between nonoverlapping areas of E- and W-maps, voxels only generated by E-map but not by W-map were used to train an SVM and classify the test dataset after selecting the top $x\%$ of brain voxels from both of E- and W-maps, and vice versa. This method of comparison gives insight into how well one functional map can identify discriminating voxels which the other functional map does not. In addition, if voxels selected from the two different maps show similar decoding accuracy, and decoding accuracies from commonly selected voxels are also comparable to the decoding accuracy from all the voxels selected from each method, it is likely that the classification performance is more dependent on the overlapping voxels than on the nonoverlapping voxels. To evaluate the performance of functional maps more clearly, therefore, it is useful to test the decoding accuracy from nonoverlapping (i.e., exclusively selected voxels) areas. In addition, information such as the amount of overlap and how much decoding accuracy from the overlapping voxels is close to decoding accuracy from all the selected voxels provides greater insight for the purpose of comparison.

Clustering Functional Maps

To compare functional maps in terms of the degree of spread of local patterns or the degree of focus, areas occupied by different clusters was calculated in each map. The top $x\%$ of voxels (e.g., 5%), hereafter called first-level threshold, having the highest functional values (i.e., FSW value in the F-map, I value in I-map, weight value in W-map, and EV in E-map) in magnitude were selected. The selected voxels were clustered by the surface connection method [Thurfjell et al., 1992]. Then, the ratio of the area of clusters consisting of at least " x " voxels (where $x = 10, 30, 50, \text{ and } 70$), hereafter called second-level threshold, to total area of the selected voxels was computed. Increasing the minimum size of clusters to be selected indicates that the selected clusters get larger and area occupied by all the selected clusters remains the same or becomes smaller.

RESULTS

Classification

We first evaluated whether SVM could consistently classify the fMRI group data between left- and right-hand movement conditions. Our analysis indicated that SVM could discriminate between LH and RH conditions with an average accuracy of over 90% ($95.8\% \pm 1.2\%$).

Comparison of Functional Maps

Characteristics of activation clusters of each functional map are presented by showing ratios of the area occupied by selected clusters after second-level thresholding to total area of voxels selected with first-level threshold (see Fig. 4). The selection of voxels of Figure 4 was performed with the same method as Figure 3, but with more number of thresholds, i.e., 10, 5, and 2.5%. The graphs of Figure 4B show the characteristics of clusters in Figure 3. The F-map illustrates that larger clusters occupy most of the selected area (see Fig. 4). In contrast, the W-map shows more wide-spread clusters (see Figs. 3 and 4) with higher ratios of the area taken by small clusters. The E-map displays an intermediate degree of spread between the I-map and the W-map. However, characteristics of cluster area from the E-map obtained after decreasing the 1st-level threshold (i.e., from 10 to 5% and 2.5%) resembles those from F-/I-maps.

Figure 5 shows comparison of decoding performance in the selected voxels by four functional maps (F-/W-/I-/E-maps). The decoding accuracy of E-map is similar to or slightly higher in the numbers of selected voxels (10, 5, 2.5, and 1%) than the others.

In Figure 6, classification accuracies of overlapping, and nonoverlapping, exclusively selected voxels are shown for the top $x\%$ (10, 5, 2.5, and 1%) of brain voxels in each type of functional map. These comparisons were performed on nonsmoothed data, on two types of

functional maps at a time. Commonly selected voxels from the two functional maps provide higher classification accuracy in all the thresholds than exclusively selected voxels. Also, the decoding accuracies from the overlapping area are almost same as those of all the selected voxels from each functional map even when ratio of the overlapping area is not much high (i.e., below 70%) (see Figs. 5 and 6). Exclusively selected voxels from E- map provides higher classification accuracy in all the thresholds than F-/I-maps (see Fig. 6A,C). In comparison between E- and W-maps, E-map provides higher prediction accuracies in thresholds 10, 5, and 2.5%, while W-map shows better performance in 1% of brain voxels (see Fig. 6E).

DISCUSSION

The present study demonstrates that effect-mapping could be an alternative method in the multivariate analysis of fMRI data by considering both discriminability and data distribution. As shown in Figures 3 and 4, the E-map shows intermediate level of sparseness or focus in comparison to the mutual information method, and the weight-vector method. This could be explained by the process of derivation of EV, and thus effect mapping can merge the effect of each voxel on the SVM output in a multivariate way as a result of multiplying the SVM weight vector with the mutual information. In addition, E-mapping combines the univariate (I-map) and previous multivariate approaches [W-map; Mourao-Miranda et al., 2005, 2006, 2007] as a hybrid of the two methods. It can also be shown that decrease of the weight value at each voxel based on the normalized MI value in effect-value is regarded as a sort of constraint method of multivariate analysis, and the multivariate characteristics, therefore, remain in the Effect-map. We showed that E-mapping can identify the voxels more closely related to classification than the other methods previously used as it includes both the contributing factors that determine the output of SVM: the differing importance of spatially distributed brain voxels as represented by the weight vector, and the statistical distribution of the brain activations in the voxels as represented by the mutual information between each voxel activation and the SVM output. Figure 1A indicated that the SVM weight vector is a linear combination of the support vectors and does not represent the statistical distribution of the input patterns, and thus SVMs trained from different datasets potentially may have the same weight vector. Although the separating hyperplane is optimal in terms of generalization performance, it is a discriminative function in the dimension of the given input vectors. In Figure 1B, it was shown that greater distance of the sample from the separating hyperplane need not necessarily indicate more importance in data distribution in all the cases. The difference between the I-map and W-map depicted in Figure 3B reiterates the previous theoretical illustration of Figure 1C.

With these points of view, we have evaluated the efficacy of different methods of multivariate functional mapping: weight vector mapping [Mourao-Miranda et al., 2005, 2006, 2007], feature-space-weighting [LaConte et al., 2005], mutual information mapping, and the proposed effect mapping, by comparing the classification accuracy of the brain states based on each method for increasing levels of threshold of voxel selection. The comparison was performed in two ways: (1) by comparing the performance from all the voxels selected by each method (see Fig. 5), (2) by comparing the performance from overlapping and nonoverlapping voxels between the methods (see Fig. 6). The first approach showed a small but consistent improvement in prediction by the effect mapping method over the other two methods. However, the second approach gives more insight into the workings of the three methods. It shows that the overlapping voxels or common voxels selected by the two different methods provides most of the information necessary for prediction of the brain states indicated by the highest classification accuracy (around 96%). However, since the maps identify the nonoverlapping areas as well as the overlapping areas as informative, it is still necessary to evaluate the areas exclusively identified as informative from the two different methods. Voxels exclusively identified as informative by the FSW or the weight vector method did not perform as well in pattern classification as the voxels exclusively identified as informative voxels by the EM method. Additionally, according to Op de Beeck et al. [2008], "The word "map" is generally used to refer to a gradient of selectivities along the cortical sheet. By contrast, "module"—in the context of brain function—refers to the clustering of selectivities in discrete regions, with clear selectivity discontinuities at the boundaries of these regions." At the current state of art, since it is not clear which brain functions are maps and which others are modules, it is premature to select a fixed number of voxels for the purpose of optimal thresholding to provide the highest decoding accuracy. When one observes the classification accuracies at different threshold levels (see Figs. 5 and 6), it is apparent that EM produces consistent prediction accuracies at most thresholds in comparison to other mapping methods. This is a significant advantage for both brain-state decoding and functional mapping applications, as one need not conduct a comprehensive search for the best threshold.

In conclusion, our theoretical explications and empirical analysis indicate that the new technique of effect mapping could enhance the identification of brain activation patterns in various perceptive and cognitive tasks. Although our proposed method partially overcomes the limitations in the previous methods, there are still disadvantages with computing the mutual information in a univariate way. In addition, the influence and the relationship between overlapping areas and nonoverlapping areas could be considered to understand the SVM output and obtain more accurate functional maps in neuroimaging.

ACKNOWLEDGMENTS


Author S. L. is grateful to DAAD (German Academic Exchange Service) for supporting this research.

REFERENCES

- Cover TM, Thomas JA (1991): *Elements of Information Theory*. New York: Wiley. 542 p.
- Davatzikos C, Ruparel K, Fan Y, Shen DG, Acharyya M, Loughhead JW, Gur RC, Langleben DD (2005): Classifying spatial patterns of brain activity with machine learning methods: Application to lie detection. *Neuroimage* 28:663–668.
- Hastie T, Tibshirani R, Friedman JH (2001): *The Elements of Statistical Learning: Data Mining, Inference, and Prediction: With 200 Full-Color Illustrations*. New York: Springer. 533 p.
- 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 (2006): Decoding mental states from brain activity in humans. *Nat Rev Neurosci* 7:523–534.
- Haynes JD, Sakai K, Rees G, Gilbert S, Frith C, Passingham RE (2007): Reading hidden intentions in the human brain. *Curr Biol* 17:323–328.
- Jain K, Duin APW, JianchangMao R (2000): Statistical pattern recognition: A review. *IEEE Trans Pattern Anal* 22:4–37.
- Joachims T (1999): Making large-scale SVM learning practical. In: Schölkopf B, Burges C, Smola A, editors. *Advances in Kernel Methods - Support Vector Learning*: MIT-Press.
- Kamitani Y, Tong F (2005): Decoding the visual and subjective contents of the human brain. *Nat Neurosci* 8:679–685.
- Kriegeskorte N, Simmons WK, Bellgowan PS, Baker CI (2009): Circular analysis in systems neuroscience: The dangers of double dipping. *Nat Neurosci* 12:535–540.
- LaConte S, Strother S, Cherkassky V, Anderson J, Hu X (2005): Support vector machines for temporal classification of block design fMRI data. *Neuroimage* 26:317–329.
- Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P (1997): Multimodality image registration by maximization of mutual information. *IEEE Trans Med Imaging* 16:187–198.
- Mourao-Miranda J, Bokde AL, 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.
- Mourao-Miranda J, Reynaud E, McGlone F, Calvert G, Brammer M (2006): The impact of temporal compression and space selection on SVM analysis of single-subject and multi-subject fMRI data. *Neuroimage* 33:1055–1065.
- Mourao-Miranda J, Friston KJ, Brammer M (2007): Dynamic discrimination analysis: A spatial-temporal SVM. *Neuroimage* 36:88–99.
- Norman KA, Polyn SM, Detre GJ, Haxby JV (2006): Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends Cogn Sci* 10:424–430.
- Op de Beeck HP, Haushofer J, Kanwisher NG (2008): Interpreting fMRI data: Maps, modules and dimensions. *Nat Rev Neurosci* 9:123–135.
- Polyn SM, Natu VS, Cohen JD, Norman KA (2005): Category-specific cortical activity precedes retrieval during memory search. *Science* 310:1963–1966.
- Schölkopf B, Smola AJ (2002): *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. Cambridge, Mass.: MIT Press. 626 p.
- Schölkopf B, Burges CJC, Smola AJ (1999): *Advances in Kernel Methods: Support Vector Learning*. Cambridge, Mass.: MIT Press. 376 p.
- Soon CS, Brass M, Heinze HJ, Haynes JD (2008): Unconscious determinants of free decisions in the human brain. *Nat Neurosci* 11:543–545.
- Thurfjell L, Bengtsson E, Nordin B (1992): A new three-dimensional connected components labeling algorithm with simultaneous object feature extraction capability. *CVGIP: Graph Models Image Process* 54:357–364.
- Vapnik VN (1995): *The Nature of Statistical Learning Theory*. New York: Springer. 188 p.

4. 3. Detection of cerebral reorganization induced by real-time fMRI feedback training of insula activation: A multivariate investigation

Detection of Cerebral Reorganization Induced by Real-Time fMRI Feedback Training of Insula Activation: A Multivariate Investigation

Neurorehabilitation and
Neural Repair
XX(X) 1–9
© The Author(s) 2010
Reprints and permission: <http://www.sagepub.com/journalsPermissions.nav>
DOI: 10.1177/1545968310385128
<http://nnr.sagepub.com>


Sangkyun Lee, MSc^{1,2}, Sergio Ruiz, MD^{1,2,3}, Andrea Caria, PhD^{1,4},
Ralf Veit, PhD¹, Niels Birbaumer, PhD^{1,5}, and Ranganatha Sitaram, PhD¹

Abstract

Background. Studies with real-time functional magnetic resonance imaging (fMRI) demonstrate that humans volitionally regulate hemodynamic signals from circumscribed regions of the brain, leading to area-specific behavioral consequences. Methods to better determine the nature of dynamic functional interactions between different brain regions and plasticity due to self-regulation training are still in development. **Objective.** The authors investigated changes in brain states while training 6 healthy participants to self-regulate insular cortex by real-time fMRI feedback. **Methods.** The authors used multivariate pattern analysis to observe spatial pattern changes and a multivariate Granger causality model to show changes in temporal interactions in multiple brain areas over the course of 5 repeated scans per subject during positive and negative emotional imagery with feedback about the level of insular activation. **Results.** Feedback training leads to more spatially focused recruitment of areas relevant for learning and emotion. Effective connectivity analysis reveals that initial training is associated with an increase in network density; further training “prunes” presumably redundant connections and “strengthens” relevant connections. **Conclusions.** The authors demonstrate the application of multivariate methods for assessing cerebral reorganization during the learning of volitional control of local brain activity. The findings provide insight into mechanisms of training-induced learning techniques for rehabilitation. The authors anticipate that future studies, specifically designed with this hypothesis in mind, may be able to construct a universal index of cerebral reorganization during skill learning based on multiple similar criteria across various skilled tasks. These techniques may be able to discern recovery from compensation, dose–response curves related to training, and ways to determine whether rehabilitation training is actively engaging necessary networks.

Keywords

cerebral reorganization, brain–computer interface, functional magnetic resonance, multivariate analysis

Introduction

Advances in magnetic resonance data acquisition and processing techniques have enabled the implementation of real-time functional magnetic resonance imaging (rtfMRI) by online processing of fMRI signals.¹ With rtfMRI the blood oxygen level dependent (BOLD) signal can be used to provide participants with information about the actual level of activity in circumscribed brain regions.^{1–3} Studies have reported that healthy participants as well as patients are able to learn to control activity in different brain areas^{4–8} and such modulation may lead to behavioral changes,^{9,10} thus anticipating the development of treatment methods for neurological disorders. In spite of these exciting developments, it is still not clear how cortical and subcortical regions reorganize during self-regulation training.

Studies have shown that the activity of the insular cortex is correlated with participants’ perception of emotional states.^{11–13} In a meta-analysis of positron emission tomography and fMRI studies of emotional tasks, Phan et al¹⁴ revealed that anterior cingulate and insula are positively correlated

¹University of Tübingen, Tübingen, Germany

²International Max Planck Research School, Tübingen, Germany

³Pontificia Universidad Católica de Chile, Santiago, Chile

⁴University of Trento, Trento, Italy

⁵Istituto di Ricovero e Cura a Carattere Scientifico, Venezia, Lido, Italy

Corresponding Author:

Sangkyun Lee, Institute of Medical Psychology and Behavioural Neurobiology, Eberhard-Karls-University of Tübingen, Gartenstr 29, D-72074 Tübingen, Germany
Email: lee.sangkyun@gmail.com

with emotional recall and emotional tasks with high cognitive demand. With consideration of the above, in a previous study, we trained participants with rtfMRI neurofeedback to modulate their BOLD activity in the right anterior insula⁴ and then focused more on the task-related increases and decreases in brain activation in selected regions of interest (ROIs), such as the left and right anterior insula.

Related to the recent studies⁴⁻⁸ on self-regulation learning with the aid of neurofeedback information, studies on skill learning¹⁵⁻¹⁸ with neuroimaging techniques are of particular interest. In these studies, 2 main ideas were proposed in the extant literature to explain the mechanism behind skill acquisition.¹⁵ The first proposes that learning can take place through the efficient use of specific neuronal circuits (eg, Hebbian learning). The second idea emphasizes that skill acquisition goes through stages,^{16,17} that is, different processes are associated with the different levels of skill and complexity of the task.¹⁸ It is then proposed that different stages of skill acquisition may use different neural substrates. Petersen et al¹⁵ maintained that although these 2 mechanisms may appear different, they may not necessarily be exclusive, emphasizing further that changes in neural circuit efficiency and differences in processing are part of skill acquisition in almost all situations and that changes in activation across brain areas may be a common theme in learning paradigms. With this reasoning, more recent studies have investigated skill acquisition on a range of motor, visuomotor, perceptual, and cognitive tasks.^{15,19-21} In the majority of these studies, effects were reported as increases or decreases of activation in brain areas related to task performance, but it has not been shown how the interaction between spatially distributed brain regions can change as a system due to learning and how the brain optimizes the function. Based on the aforementioned considerations, we hypothesized that if learning takes place in stages and each stage of learning leads to more efficient use of neuronal circuits, then transition from one stage of learning to the next could be observed as changes in network connectivity, in addition to the changes in the activations of brain regions being connected.

To this end, we applied 2 different multivariate methods to fMRI data to reveal changes in the brain activity during self-regulation training. First, we used the effect mapping method²² based on support vector machines (SVMs) to observe changes in the spatial activation patterns in the brain across the training sessions. Second, we implemented multivariate Granger causality modeling (GCM) to compute directed causal influences between spatially distributed voxels of the brain. From this model, we calculated the causal density (CD²³; the density of network connection) and the average of causal strength (ACS; the average strength of causal connections in the network) to reveal temporal changes in the network across learning sessions. We expected that certain measures of the 2 methods, namely, quantification of changes of spatial activation patterns and CD and ACS, could together indicate the

changes in network connectivity during learning to self-regulate insula.

Materials and Methods

Participants and Experimental Protocol

Our study was performed on fMRI data collected from 6 healthy participants who underwent 5 fMRI scanning sessions in a day in the previous study.⁴ The first session required participants to perform insula regulation without feedback. Subsequent feedback sessions consisted of 4 regulation blocks (22.5 seconds each) during which participants had to learn to increase insular activity, alternating with 5 baseline blocks (22.5 seconds each) during which they were instructed to return to baseline-level activity. During feedback sessions, the normalized average BOLD signal from the right anterior insula was presented to the participants in real time as changing bars of a graphical thermometer. For more details about the experimental protocol and information about participants, see Caria et al.⁴

Data Acquisition

Functional images were acquired on a 3.0-T whole body scanner, and a standard 12-channel coil was used as a head coil (Siemens Magnetom Trio Tim, Siemens, Erlangen, Germany). A standard echo-planar imaging sequence was used (TR = 1.5 seconds, matrix size = 64 × 64, effective echo time TE = 30 ms, flip angle $\alpha = 70^\circ$, bandwidth = 1.954 kHz/pixel). Sixteen oblique axial slices (voxel size = 3.3 × 3.3 × 5.0 mm³, slice gap = 1 mm) were acquired.

Preprocessing and Classification

Preprocessing was performed with SPM5 (Wellcome Department of Imaging Neuroscience, London, England), and classification was performed using MATLAB (The Mathworks, Natick, MA) scripts.

We performed realignment, coregistration, and normalization onto the Montreal Neurological Institute space. Finally, the data were smoothed spatially with a Gaussian kernel of 8 mm full width at half maximum. For the preprocessed data, a mask of nonbrain areas was created by removing voxels below a specified intensity value in a mean image of whole scans. The aforementioned mask was applied to all the functional images of all the participants, and z normalizations (z value: $(x - \text{mean}(x))/\text{standard deviation}(x)$, x : samples) were applied across all the time series on each single-participant data separately because variance of BOLD signals of different participants should be considered. All brain voxels from each single scan are collected into an input vector, whose design label is given based on its condition (1 for the regulation condition and -1 for the rest condition). To account for the

hemodynamic delay, the design label was shifted by 4 scans (6 s), and then the data from the completed 8 blocks (4 blocks from the regulation + 4 blocks from the baseline) of all the 6 participants were used to construct the input vectors of classifiers. To classify the data of the 2 given conditions (ie, regulation vs baseline) across sessions, SVM (linear SVM with the regularization parameter $C = 1$, using SVMlight²⁴) were used in each separate session. Classification performance from data within a session was evaluated through 6-fold cross-validation.²⁵ In each fold, the data of 5 participants were used to train an SVM classifier, and then the data of 1 remaining participant were used to test the classifier. This process was repeated across all the folds without the testing sets overlapping across folds (ie, leave-one-subject-out cross-validation).

Multivariate Spatial Analysis

To analyze spatial patterns of brain activity, we used a multivariate method called effect mapping (EM).²² It reveals changes in the spatial patterns of activity across training sessions. EM was shown to be stronger when compared with other methods in both theoretical and empirical points of view in our previous study.²²

To identify informative voxels from the SVM model, which separates 2 conditions by finding the separating hyperplane, the EM measures the effect of each voxel in multivoxel space to the SVM output by considering both the factors (ie, the input vectors and the weight vector; $y = \mathbf{w}^T \mathbf{x} + w_0$, where y is the SVM output, \mathbf{w} is the weight vector, and \mathbf{x} the input vector), which determine the SVM output. The effect of each voxel to the classifier output is measured by computing normalized mutual information (NMI)²⁶ between the voxel and the SVM output.

Hence, the effect value (EV) E_k of a voxel k is defined as (see Lee et al²² for more details):

$$E_k = w_k \tilde{I}(x_k; y), k = 1, \dots, M \quad (M: \text{number of voxels}), \quad (1)$$

where $\tilde{I}(x_k; y)$ is the NMI between the voxel and the output, w_k and x_k are the SVM weight value and activation in voxel k , respectively.

After normalizing the absolute value of E_k from Equation (1), we obtain the following relation:

$$nE_k = \text{sgn}(E_k) \log(1 + |E_k| / \text{std}(|E|)), k = 1, \dots, M, \quad (2)$$

where $\text{sgn}(\cdot)$ is a sign function and $\text{std}(|E|)$ is the standard deviation of all E_k . In the present study, Equation (2) (nE_k) was used to compute the EV at each voxel to make E(ffect)-maps from different sessions comparable. Hereafter, the EV indicates nE_k .

To compare quantitatively the effect of learning on changes in the recruitment of spatially distributed brain areas, we applied thresholds of $EV \geq Th \times E_{\max}$ and $EV \leq Th \times E_{\min}$ in

each session separately. Here, Th (where $Th = 0.1, \dots, 0.7$) is the threshold values, and E_{\max} , E_{\min} , and EV are a maximum and minimum value of effect values and an effect value at a single voxel, respectively. The application of these thresholds is equivalent to the situation of removing voxels that are not significantly important for the learning and retaining those that are directly involved in it.

BOLD Signal Difference

To evaluate BOLD signal changes of the selected brain region in all the sessions, the BOLD signal difference was defined as $(\text{BOLD}_{\text{reg}} - \text{BOLD}_{\text{base}}) / \text{BOLD}_{\text{base}} * 100$.

Effective Connectivity Analysis

In our temporal analysis, Granger causality (GC) models were computed for all sessions based on vector autoregressive models (VARs).²⁷ When ROIs were selected from the E-map generated from multivariate SVM analysis, the time-series vector $\mathbf{b}(n)$ of BOLD activations of the ROIs is defined under the VAR process as

$$\mathbf{b}(n) = - \sum_{i=1}^p \mathbf{A}(i) \mathbf{b}(n-i) + \mathbf{e}(n), \quad (3)$$

where p is the maximum number of past values considered for estimating the current vector, $\mathbf{A}(i)$ contains the estimated coefficients of i th delayed time samples, and residual vector $\mathbf{e}(n)$ in sample point n . In this VAR process, the causality from $b_j(n)$ (the time series of the corrected BOLD activations at voxel j) to $b_k(n)$ (the time series of the corrected BOLD activations at voxel k) is evaluated as follows. If $e_k(n)$ is increased by excluding $b_j(n)$, it implies that $b_k(n)$ Granger-causes $b_j(n)$. This can be tested by using the F test of the null hypothesis.

Based on our hypothesis that self-regulation learning leads to more efficient use of neuronal circuits, the areas having higher EVs in the last session could be regarded as part of a core network related to regulating activation of the insular cortex, and hence assessment of the connectivity of this network over time would allow us to understand the changes in the network during the learning. To this end, local maxima having the highest effect values were selected from the E-map of the last session of the group analysis.

Then, preprocessing steps (realignment, time reslicing, normalization, and smoothing) were performed. The time series of BOLD values from 27 voxels forming a cube of $3 \times 3 \times 3$ voxels, comprising the local maximum and 26 adjacent voxels, were obtained based on our observation that a single cluster may have multiple local maxima, each one of these maxima potentially showing different temporal dynamics. Given this, time series of a single voxel might not be robust against noise or other artifacts.

In the Granger connectivity analysis, the greater the number of samples the greater is the reliability of the results.

From this it follows that a small number of brain images from a session of a single participant do not allow the effective connectivity analysis with many selected ROIs. Therefore, in our analysis, rather than performing the connectivity analysis in a single participant, data from a session of all the participants were used together in the analysis. Although this approach might have limitation in the provision of strong evidences with statistical significance, it could nevertheless be an alternative method to overcome the problem of limited number of samples commonly available in fMRI time series.

To determine effective connectivity in the regulation periods, only signals corresponding to the regulation conditions from the group data were used in the analysis. To correct for intersubject variations in BOLD values, the mean value in each subject was subtracted, and then the data from each subject were concatenated to form a set of group data. The GC analysis was carried out by adopting the Causal Connectivity Toolbox²³ to work with fMRI signals. In the estimation process of $\mathbf{A}(i)$ in Equation (3), the end part of time courses of each regulation block was excluded in the right-hand side of Equation (3) to prevent parameter values to be wrongly estimated due to discontinuities between the concatenated regulation blocks. Consider the time series $\mathbf{b}[1, \dots, n, \dots, N]$, where N is the number of samples. Let us assume that a discontinuity occurs at the index n , the discontinuity represented by $\mathbf{b}(n)$, where n is the start index of a new block. Let the model order be 2. In our analysis, we have not used $\mathbf{b}(n-1)$ and $\mathbf{b}(n-2)$ in the right-hand side of this formula to estimate $\mathbf{A}(i)$ because $\mathbf{b}(n)$ needs $\mathbf{b}(n-1)$ and $\mathbf{b}(n-2)$, that is, $\mathbf{b}(n) \approx -\mathbf{A}(1)\mathbf{b}(n-1) - \mathbf{A}(2)\mathbf{b}(n-2)$, and $\mathbf{b}(n+1)$ needs $\mathbf{b}(n)$ and $\mathbf{b}(n-1)$, that is, $\mathbf{b}(n+1) \approx -\mathbf{A}(1)\mathbf{b}(n) - \mathbf{A}(2)\mathbf{b}(n-1)$. As the discontinuity occurs at n and parameter matrix \mathbf{A} computed from the relation between $\mathbf{b}(n)$ and $\mathbf{b}(n-1)$ would be wrong, we excluded such points, that is, $\mathbf{b}(n) \approx -\mathbf{A}(1)\mathbf{b}(n-1) - \mathbf{A}(2)\mathbf{b}(n-2)$ and $\mathbf{b}(n+1) \approx -\mathbf{A}(1)\mathbf{b}(n) - \mathbf{A}(2)\mathbf{b}(n-1)$, in the estimation of \mathbf{A} .

To find the model order p in Equation (3), Bayesian information criteria^{23,25} were used. After estimating the parameters, the F test was used to evaluate the connection strength between ROIs. To reveal changes in the temporal interactions among different brain areas due to feedback training, CD and ACS were computed for all the training sessions. For our analysis, CD and ACS should reflect the global trend of temporal dynamics at the system level. For this purpose, GCM is a more appropriate method than dynamical causal model²⁸ or psychophysiological interaction,²⁹ as GCM can be used to analyze changes in the effective connectivity of multiple brain regions.²³

The CD²³ reflects the degree of interaction among the ROIs as defined by the following equation:

$$CD = \frac{GC}{2N(N-1)}, \quad (4)$$

where GC is the total number of significant causal connections observed and N the number the ROIs. In addition, to

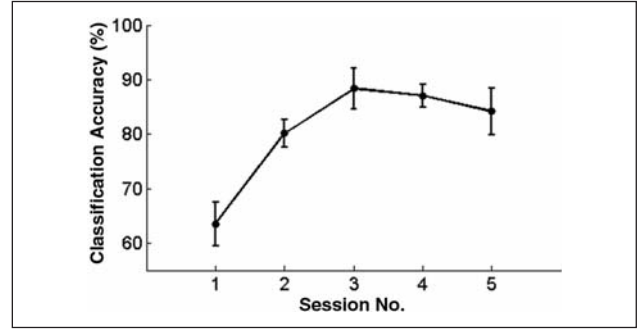


Figure 1. Classification accuracy of group data in the 6-fold cross validation (leave-one-subject-out approach; mean accuracy [%] \pm standard error of the mean [%])

investigate the change of connection strength across sessions, the ACS in each causal network is newly defined as follows:

$$ACS = \frac{\sum_{i,j=1,i \neq j}^N \log(F_{ij})}{GC}, \quad (5)$$

where F_{ij} is the F value corresponding to connection from ROI j to ROI i .

Results

SVM Classification

To investigate our hypotheses, we first evaluated whether SVM could consistently classify the fMRI signals between regulation and baseline conditions. Our analysis indicated that SVM could discriminate between regulation and baseline conditions in the feedback sessions ($T[23] = 20.9$, $P < 10^{-10}$) with an average accuracy of more than 80% in most of the cases, whereas the average accuracy of classification in the first session without feedback ($T[5] = 3.4$, $P < .02$) was lower than with feedback, indicating the importance of contingent feedback in successful regulation of brain activity (see Figure 1). In addition, the high classification accuracies during the sessions with feedback represent that the E-maps from all the sessions are reliable for subsequent analysis and interpretation.

Cerebral Reorganization

During Learning of Self-Regulation

Spatial interaction of brain activity. Spatial distribution of brain activity was investigated by applying EM in a group analysis by using fMRI signals from 6 participants as input to the pattern classifier (see Figure 2). When $Th = 0.2$, Figure 2A and B show characteristics of the E-maps across sessions. Application of the threshold to the first session (performed without feedback) resulted in smaller clusters of voxels with higher

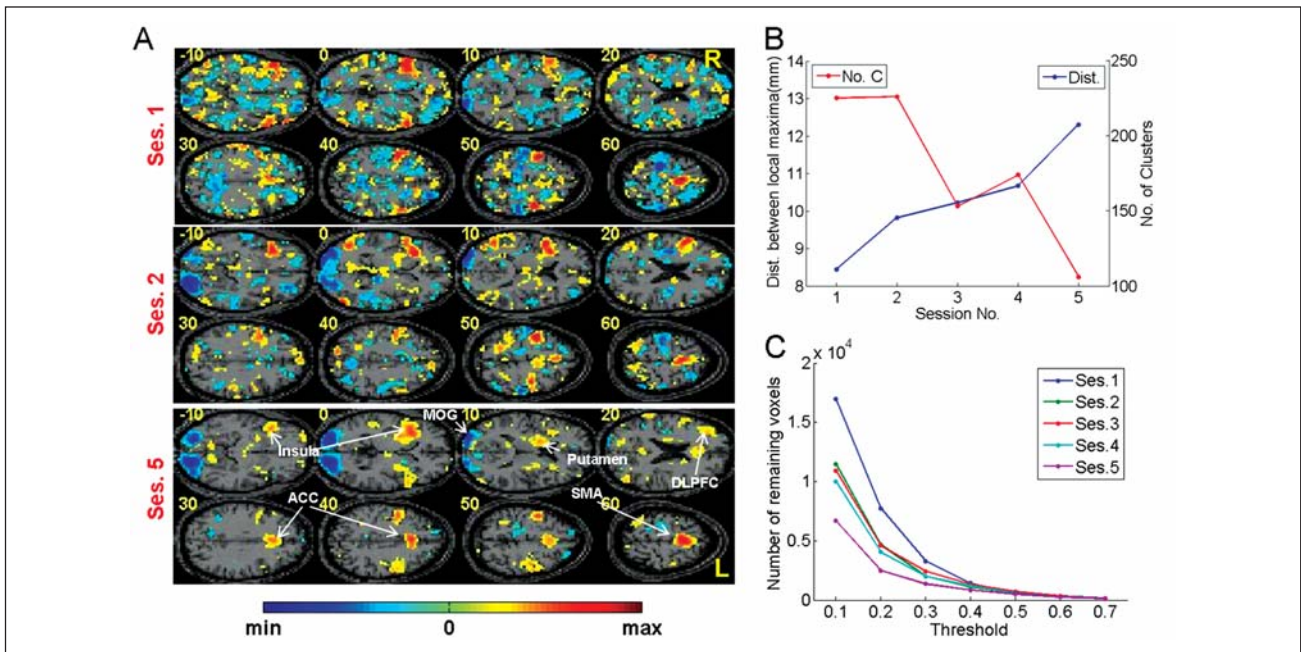


Figure 2. Multivariate analyses of group data over training sessions. (A) E-maps in session 1 (without feedback), and sessions 2 and 5. E-maps are shown after applying a threshold given by the relation: $EV \geq 0.2E_{\max}$ and $EV \leq 0.2E_{\min}$. Application of this threshold removes voxels that are not significantly important for regulation and retains those that are directly involved. In the functional maps, clusters with positive EVs (red) correspond to the positive support vector machine (SVM) weight values, whereas clusters with negative EVs (blue) correspond to the negative SVM weight values. When compared with sessions 1 and 2, the E(effect)-map of session 5 shows more focused activations, indicating the effect of learning on functional reorganization of the brain. (B) From the aforementioned voxels selected in each session, mean values of minimum distances between local maxima and numbers of clusters are plotted. (C) Numbers of remaining voxels after thresholding EVs with $EV \geq Th E_{\max}$ and $EV \leq Th E_{\min}$ ($Th = 0.1, 0.2, \dots, 0.7$; eg. “A” and “B” are obtained with $Th = 0.2$) in all the sessions. This method shows the distributions of EVs across sessions. The E-maps of the 5 sessions have similar number of voxels having high EVs (eg. threshold > 0.5). Whereas the E-map from session 5 shows many voxels having low EVs, the E-map from session 1 shows many voxels having intermediate EVs. The E-maps from sessions 2, 3, and 4 show similar characteristics, but the number of the voxels having lower EVs increases gradually as the feedback training proceeds. This suggests that, as the feedback training proceeds, more voxels acquire low EVs (ie, voxels irrelevant to the task performance), indicating that fewer voxels are involved in the task

and intermediate EVs spread over the whole brain, suggesting the involvement of a large number of brain areas. However, application of the threshold in the feedback sessions resulted in focused activations in a few brain regions, indicating probably the gradual disengagement of unnecessary connections. Particularly, in session 5, the distribution is more focused and localized in comparatively fewer areas, such as the lingual gyrus, middle occipital gyrus, supplementary motor area, anterior cingulate, right anterior insula, putamen, and dorsolateral prefrontal cortex (see Figure 2A and Table 1). In comparing clusters composed of the discriminating voxels of each session, the number of clusters and the average of the minimum distances between clusters in all the sessions were computed. As the learning proceeds, the number of clusters decreases, and the average of the minimum distance between local maxima increases, indicating focused activation (see Figure 2B). In addition, the application of several thresholds shows more general characteristics of distribution of EVs across sessions. As the feedback training proceeds, more

voxels acquire low EVs with a number of voxels keeping high EVs (see Figure 2C).

In addition, BOLD signal changes of areas having higher EVs in session 5, that is, right dorsolateral prefrontal cortex (DLPFC), left inferior frontal gyrus (IFG), right Putamen, and anterior cingulate cortex (ACC), were investigated in a further analysis (see Figure 3). In this analysis, the BOLD signals of the areas during the feedback sessions are higher than in the nonfeedback session. However, the signals gradually decrease with learning in the feedback sessions but are still higher in the last session than in the first session (no feedback session), suggesting “inverted-U curves” of BOLD signal control.

Temporal interaction in brain activity. To extend the evidence for the cerebral spatial distribution of brain activity, temporal interaction was also investigated. After selecting ROIs based on the aforementioned spatial analysis (see Table 1), temporal interactions between time series of BOLD signals in the ROIs were investigated using the multivariate Granger causality analysis from the estimation of vector autoregressive models.²⁷

Table 1. Maxima and Minima in the Final Session in a Group Analysis

No.	Region of (De)activation	BA	MNI Coordinates			EV
			x	y	z	
1	Middle occipital gyrus L		-17	-89	-5	-3.06
2	Middle occipital gyrus R		26	-92	5	-2.67
3	Lingual gyrus R		20	-86	-5	-2.61
4	SMA R	6	3	7	60	2.39
5	Middle occipital gyrus L		-36	-86	-5	-2.27
6	Anterior cingulate R (ACC)		7	17	40	1.99
7	Insula R	47	33	17	0	1.97
8	Precentral R (SMA)	6	46	-3	50	1.96
9	Insula R		43	20	-5	1.91
10	Anterior cingulate R	32	7	26	25	1.59
11	Lentiform nucleus R (putamen R)		20	0	10	1.56
12	Anterior cingulate L		-7	17	40	1.56
13	Precentral gyrus L (SMA)		-50	-7	45	1.47
14	Putamen R		30	3	0	1.46
15	Precentral gyrus L	6	-40	-10	55	1.41
16	Inferior frontal gyrus L (IFG L)		-56	13	0	1.35
17	Middle frontal gyrus R (DLPFC R)	46	33	40	20	1.32
18	Middle frontal gyrus L (DLPFC L)	6	-53	3	45	1.28
19	Middle temporal gyrus L		-53	-73	20	1.28
20	Angular R		46	-50	25	-1.2
21	Pallidum R (putamen R)		20	3	-5	1.19
22	Precentral gyrus R		30	-26	65	-1.19
23	Superior temporal gyrus R (STG R)	22	66	-43	20	1.18

Abbreviations: ACC, anterior cingulate cortex; BA, Brodman area; DLPFC, dorsolateral prefrontal cortex; EV, effect value; MNI, Montreal Neurological Institute; SMA, supplementary motor area; STG, superior temporal gyrus.

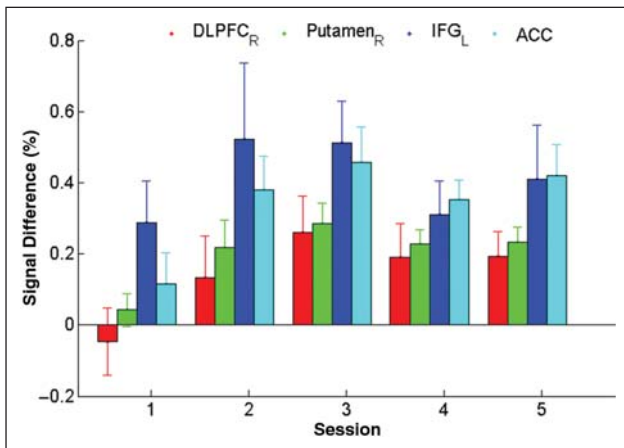


Figure 3. Blood oxygen level dependent (BOLD) signal differences in selected brain regions in the first (nonfeedback) feedback sessions in right dorsolateral prefrontal cortex (DLPFC), right Putamen, left inferior frontal gyrus (IFG), anterior cingulate cortex (ACC) suggest “inverted-U” curves of activation patterns. Each bar shows mean difference (%) \pm standard error of the mean (%)

To evaluate how temporal recruitment of brain regions in all sessions changes, the CD,²³ a measure of the number of

connections between selected ROIs in a functional network, was used after finding significant connections between the ROIs. The investigation was done with F test for the connections between the ROIs (ie, the estimated model parameters) with $P < .05$ (see Figure 4A). Session 1 shows the lowest CD of the functional network. As the training proceeds with feedback, the causal density reaches a peak value in session 3 and decreases in sessions 4 and 5, indicating that the shape of temporal recruitment follows an “inverted-U” curve. After performing an F test for the connections between the ROIs with $P < .0001$, ACSs between connections during all the sessions were investigated (see Figure 4B). An investigation of ACSs between ROIs indicates that the strength of temporal recruitment in some ROIs increases monotonically from session to session.

Discussion

The performances in classification accuracy (see Figure 1) and BOLD signal activations in right anterior insula (paired samples T test between session 5 and session 1, $t = 4.86$, $P < .005$; see Figure 5) with and without contingent feedback confirmed that the participants were able to perform the given task successfully.

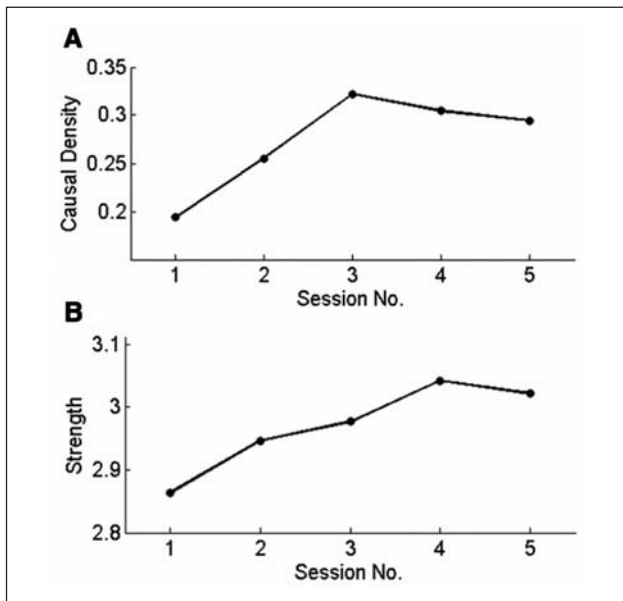


Figure 4. Functional interaction of brain regions in the group data. The analyses were done after F test for the connections between the regions of interest (ROIs) with $P < .05$ in (A) and $P < .0001$ in (B). (A) Causal density in all the sessions. (B) Average connection strength between connections. The causal density of the functional network decreases in session 5 indicating substantial “pruning” (A) yet “strengthening” (B) of the connections between ROIs

From Figures 1 and 2A, one might argue that the visual activations could solely improve the classification during the feedback sessions compared with the nonfeedback sessions. However, higher classification in session 3 than in session 2 clearly indicates that something more than just visual activation is helping increase the classification. Furthermore, there are other brain areas (see Figure 3) to show greater differences between regulation and baseline in the feedback sessions than in the nonfeedback session. In particular, since activations in the target area (insula) are significantly different in the feedback sessions compared with the nonfeedback sessions, the classification accuracy would reflect these differences.

In the multivariate spatial analysis with the E-mapping method, decrease of redundancy and more focused brain activation might indicate that training results in more economic use of brain resources during volitional regulation of the region. Since the aim of our multivariate spatial analysis is to show how the brain reorganizes across training sessions as a system rather than to identify the brain areas related to the task at a significant statistical level, we have decided to show the characteristics of spatial patterns over a broad range of thresholds. Figure 2C clearly shows that the number of voxels having lower EVs (ie, indicating probably voxels not so important to the task) increases gradually as the training proceeds. In addition, due to the present limitations of the statistical

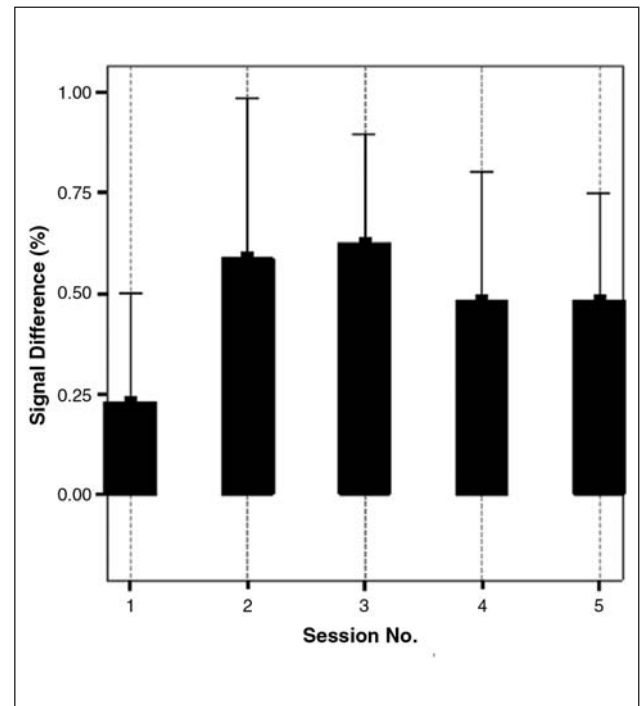


Figure 5. Blood oxygen level dependent (BOLD) signal difference in right anterior insular cortex. BOLD signal differences in all the feedback sessions are higher than in the nonfeedback session. Each bar shows mean difference (%) \pm standard error of the mean (%)

significance tests (eg, difficulty to determine what level is the best to assess a given hypothesis), we believe that showing the spatial characteristics over a wide range of threshold values would be comprehensive and complete. The changes of BOLD signal difference of the selected regions (see Figure 3), particularly the BOLD signal increase in the early part of training and the decrease in the late part of training, would further support our hypothesis.

In the effective connectivity analysis, the CD and the ACS might indicate global trends of neural processing during learning. Since the formation and maintenance of neural activation and communication among multiple, distant brain areas in such a mental task requires blood and oxygen supply to the sites,^{30,31} denser networks might expend more energy. With more training in the presence of contingent feedback, participants reported only 1 or 2 of the most effective imagery strategies that consistently affected the thermometer bars. This focusing of strategies to a few effective ones might explain the gradual reduction in the causal density of the functional network with the high levels of the average connection strength. Even though our connectivity analysis could not provide any summary statistics in the group level due to limitation in the time samples of brain images, the general trends of both of CDs over sessions (see Figure 4A) and the BOLD signal changes in the multiple areas (see Figure 3)

might be in line with previous research on skill learning, which tends to show evidence for an inverted-U-shaped learning curve due to reallocation of cognitive resources in language expertise,³² motor skill learning,³³ and attentional expertise in meditation.³⁴

The self-regulation mainly involves the prefrontal cortex, especially the DLPFC, presumably responsible for retrieval of emotional memories^{35,36}; the ACC, implicated in attentional processing¹⁴; the right insula as a site for the meta-representation of interoceptive signals^{37,38}; the left IFG for verbalization of emotional episodes³⁹; and the putamen, engaged in reinforcement learning and feedback/reward processing.^{40,41} The repeated appearance of SMA may indicate motor response elements in the images.⁴²⁻⁴⁵ We hypothesize that if the imagery of positive or negative emotions is effective (as seen in the last session of the feedback training), it may elicit activations in the SMA, indicating neural disposition toward action (eg, approach, withdrawal).

Our results provide supporting evidence for the “scaffolding-storage” theory.¹⁵ In this framework, a set of supporting brain regions form a scaffold to manage a novel demand in unskilled performance. After sufficient practice, neural processes are performed more efficiently, decreasing activation in the scaffolding regions but increasing activation in those areas underlying the critical task-specific processes. Based on this framework, in our study, right DLPFC, left IFG, right Putamen, and ACC increasing in BOLD activation levels (see Figure 3) during early feedback sessions and decreasing in later feedback sessions may represent the scaffolds. However, it is not entirely clear which areas and connections are directly involved in emotion processing per se and which are related to supporting elements in task learning.

To investigate the learning effects of rtfMRI feedback training, positive or negative imagery of emotion were used in the present study. A limitation of the current work is that different strategies to induce emotion were not systematically classified, and thus the indirect measure used here does not indicate to what extent different strategies affect the brain through practice.

Based on the fundamental assumption that the brain operates as a system where spatially distributed brain areas interact with one another, the present study provides evidence for brain reorganization at the system level in terms of transitions in different measures of network connectivity. However, it is to be noted that the present study is a reanalysis of the previously collected data of self-regulation training.⁴ Consequently, due to the limitations of the experimental design and analysis technique, and the absence of behavioral measures of learning, we could not completely establish the different stages of reorganization pertaining to different stages of learning. Nevertheless, we believe that the multivariate network measures proposed and demonstrated here, such as, changes in the spatial patterns of activity (including number of clusters, distance between local maxima, and number of

voxels having lower EVs), causal density, and average causal strength, would be valuable in assessing and understanding cerebral reorganization at different stages of learning. We anticipate that future studies, specifically designed with this hypothesis in mind, may be able to construct a universal index of cerebral reorganization during skill learning based on the measurement of multiple criteria (changes of spatial patterns, causal density, causal strength, etc) in multiple types of skilled tasks.

Authors' Note

Sangkyun Lee and Ranganatha Sitaram contributed equally to this work.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research and/or authorship of this article: This work was supported by DFG (the Deutsche Forschungsgemeinschaft) grants BI 195/56-1 and BI 195/59-1 to NB and RS, respectively. SL is grateful to DAAD (German Academic Exchange Service) for supporting this research.

References

1. Weiskopf N, Sitaram R, Josephs O, et al. Real-time functional magnetic resonance imaging: methods and applications. *Magn Reson Imaging*. 2007;25:989-1003.
2. Sitaram R, Caria A, Veit R, et al. fMRI brain-computer interface: a tool for neuroscientific research and treatment. *Comput Intell Neurosci*. 2007;25487. doi:10.1155/2007/25487
3. Sitaram R, Weiskopf N, Caria A, et al. fMRI brain-computer interfaces: a tutorial on methods and applications. *IEEE Signal Process Mag*. 2008;25:95-106.
4. Caria A, Veit R, Sitaram R, et al. Regulation of anterior insular cortex activity using real-time fMRI. *Neuroimage*. 2007; 35:1238-1246.
5. deCharms RC, Christoff K, Glover GH, Pauly JM, Whitfield S, Gabrieli JD. Learned regulation of spatially localized brain activation using real-time fMRI. *Neuroimage*. 2004;21:436-443.
6. Posse S, Fitzgerald D, Gao K, et al. Real-time fMRI of temporolimbic regions detects amygdala activation during single-trial self-induced sadness. *Neuroimage*. 2003;18:760-768.
7. Weiskopf N, Mathiak K, Bock SW, et al. Principles of a brain-computer interface (BCI) based on real-time functional magnetic resonance imaging (fMRI). *IEEE Trans Biomed Eng*. 2004;51:966-970.
8. Yoo SS, Fairney T, Chen NK, et al. Brain-computer interface using fMRI: spatial navigation by thoughts. *Neuroreport*. 2004;15:1591-1595.
9. deCharms RC, Maeda F, Glover GH, et al. Control over brain activation and pain learned by using real-time functional MRI. *Proc Natl Acad Sci U S A*. 2005;102:18626-18631.

10. Rota G, Sitaram R, Veit R, et al. Self-regulation of regional cortical activity using real-time fMRI: the right inferior frontal gyrus and linguistic processing. *Hum Brain Mapp.* 2009;30:1605-1614.
11. Damasio AR, Grabowski TJ, Bechara A, et al. Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nat Neurosci.* 2000;3:1049-1056.
12. Lane RD, Fink GR, Chau PM, Dolan RJ. Neural activation during selective attention to subjective emotional responses. *Neuroreport.* 1997;8:3969-3972.
13. Reiman EM, Lane RD, Ahern GL, et al. Neuroanatomical correlates of externally and internally generated human emotion. *Am J Psychiatry.* 1997;154:918-925.
14. Phan KL, Wager T, Taylor SF, Liberzon I. Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage.* 2002;16:331-348.
15. Petersen SE, van Mier H, Fiez JA, Raichle ME. The effects of practice on the functional anatomy of task performance. *Proc Natl Acad Sci U S A.* 1998;95:853-860.
16. Fitts PM. Perceptual-motor skills learning. In: Melton AW, ed. *Categories of Human Learning.* New York, NY: Academic Press; 1964:243-285.
17. Posner MI, Keele SW. On the genesis of abstract ideas. *J Exp Psychol.* 1968;77:353-363.
18. Rosenbaum DA. Hierarchical organization of motor programs. In: Wise SP, ed. *Higher Brain Functions: Recent Explorations of the Brain's Emergent Properties.* New York, NY: Wiley; 1987:45-66.
19. Sakai K, Hikosaka O, Miyauchi S, Takino R, Sasaki Y, Pütz B. Transition of brain activation from frontal to parietal areas in visuomotor sequence learning. *J Neurosci.* 1998;18:1827-1840.
20. Olesen PJ, Westerberg H, Klingberg T. Increased prefrontal and parietal activity after training of working memory. *Nat Neurosci.* 2004;7:75-79.
21. Kelly AM, Garavan H. Human functional neuroimaging of brain changes associated with practice. *Cereb Cortex.* 2005;15:1089-1102.
22. Lee S, Halder S, Kubler A, Birbaumer N, Sitaram R. Effective functional mapping of fMRI data with support-vector machines. *Hum Brain Mapp.* 2010;31:1502-1511.
23. Seth AK. Causal connectivity of evolved neural networks during behavior. *Network.* 2005;16:35-54.
24. Joachims T. Making large-scale SVM learning practical. In: Schölkopf B, Burges C, Smola A, eds. *Advances in Kernel Methods: Support Vector Learning.* Boston, MA: MIT Press; 1999:169-184.
25. Hastie T, Tibshirani R, Friedman JH. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* New York, NY: Springer; 2001.
26. Maes F, Collignon A, Vandermeulen D, Marchal G, Suetens P. Multimodality image registration by maximization of mutual information. *IEEE Trans Med Imaging.* 1997;16:187-198.
27. Goebel R, Roebroeck A, Kim DS, Formisano E. Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modeling and Granger causality mapping. *Magn Reson Imaging.* 2003;21:1251-1261.
28. Friston KJ, Harrison L, Penny W. Dynamic causal modeling. *Neuroimage.* 2003;19:1273-1302.
29. Friston KJ, Buechel C, Fink GR, Morris J, Rolls E, Dolan RJ. Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage.* 1997;6:218-229.
30. Moore CI, Cao R. The hemo-neural hypothesis: on the role of blood flow in information processing. *J Neurophysiol.* 2008;99:2035-2047.
31. Logothetis NK. What we can do and what we cannot do with fMRI. *Nature.* 2008;453:869-878.
32. Sakai KL. Language acquisition and brain development. *Science.* 2005;310:815-819.
33. Doyon J, Song AW, Karni A, Lalonde F, Adams MM, Ungerleider LD. Experience-dependent changes in cerebellar contributions to motor sequence learning. *Proc Natl Acad Sci U S A.* 2002;99:1017-1022.
34. Slagter HA, Lutz A, Greischar LL, et al. Mental training affects distribution of limited brain resources. *PLoS Biol.* 2007;5:e138.
35. Ochsner KN, Gross JJ. The cognitive control of emotion. *Trends Cogn Sci.* 2005;9:242-249.
36. Ochsner KN, Ray RD, Cooper JC, et al. For better or for worse: neural systems supporting the cognitive down- and up-regulation of negative emotion. *Neuroimage.* 2004;23:483-499.
37. Craig AD. How do you feel? Interoception: the sense of the physiological condition of the body. *Nat Rev Neurosci.* 2002;3:655-666.
38. Craig AD. Interoception: the sense of the physiological condition of the body. *Curr Opin Neurobiol.* 2003;13:500-505.
39. Burianova H, Grady CL. Common and unique neural activations in autobiographical, episodic, and semantic retrieval. *J Cogn Neurosci.* 2007;19:1520-1534.
40. O'Doherty JP. Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Curr Opin Neurobiol.* 2004;14:769-776.
41. Haruno M, Kawato M. Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus-action-reward association learning. *J Neurophysiol.* 2006;95:948-959.
42. Sabatinelli D, Lang PJ, Bradley MM, Flaisch T. The neural basis of narrative imagery: emotion and action. *Prog Brain Res.* 2006;156:93-103.
43. Leslie KR, Johnson-Frey SH, Grafton ST. Functional imaging of face and hand imitation: towards a motor theory of empathy. *Neuroimage.* 2004;21:601-607.
44. Oliveri M, Babiloni C, Filippi MM, et al. Influence of the supplementary motor area on primary motor cortex excitability during movements triggered by neutral or emotionally unpleasant visual cues. *Exp Brain Res.* 2003;149:214-221.
45. Warren JE, Sauter DA, Eisner F, et al. Positive emotions preferentially engage an auditory-motor "mirror" system. *J Neurosci.* 2006;26:13067-13075.

4. 4. Real-time support vector classification and feedback of multiple emotional brain states

Title: Multivariate prediction of movement intention in the human fronto-parietal cortex.

Author affiliation

Sergio Ruiz^{a,b,c,*}, Sangkyun Lee^{a,c,*}, Claudia Schneider^c, Niels Birbaumer^{a,d,e}, Ranganatha Sitaram^a

* These authors contributed equally to this work.

^aInstitute of Medical Psychology and Behavioral Neurobiology, University of Tübingen, Gartenstr. 29, 72074 Tübingen, Germany

^bDepartment of Psychiatry, Faculty of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile

^cGraduate School of Neural & Behavioural Sciences, International Max Planck Research School, 72074 Tübingen, Germany

^dOspedale San Camillo, Istituto di Ricovero e Cura a Carattere Scientifico, Venezia – Lido, Italy

^eDepartment of Psychology, Biological Psychology, Universidad de las Islas Baleares, Spain

Corresponding author

Ranganatha Sitaram

Research Scientist

Institute of Medical Psychology and Behavioural Neurobiology,

Eberhard-Karls-University of Tübingen, Gartenstr. 29, D-72074 Tuebingen, Germany

Email: sitaram.ranganatha@uni-tuebingen.de

Phone: +49-7071-29-73263

Fax: +49-7071-29-5956

Abstract

Growing evidence indicates the critical role of posterior parietal cortex (PPC) in coding movement intention. Recently, brain computer interfaces (BCIs) and neuroprosthetics have emerged as promising tools for rehabilitation of patients suffering from stroke and other movement disabilities. The detection of brain activity related to movement intention could be used to activate neuroprostheses in patients with frontal lesions or with difficulties performing motor imagery. The present study examines whether a multivariate pattern classifier can decode movement intention and the type of movement, from PPC in an event related functional magnetic resonance (fMRI) paradigm. Univariate statistical analysis was first performed to ascertain the role of PPC in movement intention, and then a spatio-temporal support vector machine (SVM) was used to predict left vs. right hand movement intention. Our results show that movement intention can be detected by patterns of brain activations in the PPC, supplementary motor area (SMA) and premotor cortex (PMC) with above chance accuracy. Furthermore, high prediction accuracies could be obtained while discriminating between left and right hand movement intention and imagery. Our results indicate that movement intention and the type of movement could be decoded from the human fronto-parietal cortex, opening up the possibility for developing movement intention based prosthetics.

Introduction

Neuroprosthetic devices and Brain Computer Interfaces (BCIs) have recently emerged as potential tools for the treatment of motor disabilities (Birbaumer and Cohen, 2007; Buch et al., 2008; Wolpaw et al., 2002). Experiments have shown that it is possible to use brain signals to decode mental states to operate an external device. Majority of current approaches to developing neuroprosthetics in humans relay on the capability of the patients to perform motor imagery, despite the fact that motor imagery impairments have often been found after brain lesions (Gonzalez et al., 2005; Malouin et al., 2004; Sabate et al., 2007). Hence, it could be useful to decode movement information in the cognitive steps that precede motor execution or imagery. It has been suggested that BCIs that detect “movement intention” and use this information to operate a neuro-prosthetic device could better benefit the rehabilitation of these patients (Andersen and Cui, 2009; Mulliken et al., 2008; Musallam et al., 2004). Movement intention is a high-level cognitive function needed for early movement planning that can specify both the type of movement and the target for that movement (Andersen and Buneo, 2002). The use of brain activations during movement intention for neuro-prosthetics could have a temporal advantage due to the fact that the prosthetic device would receive a decoded signal before the performance of motor execution or motor imagery.

Evidence has been accumulating in the last years about the role of parietal cortex in action planning. Several studies in primates have shown that the posterior parietal cortex (PPC) contains anatomically segregated regions, i.e. “intentional maps” that code for the planning of different movements (Andersen and Buneo, 2002; Fogassi et al., 2005; Mazzoni et al., 1996; Quian Quiroga et al., 2006; Sakata et al., 1995; Snyder et al., 1997, 2000). Neuroimaging and lesion studies in humans have also shown the involvement of fronto-parietal areas in motor intention and awareness (Berti et al., 2005; Haggard, 2005, 2008; Lau et al., 2004; Sirigu et al., 2004). In this sense, a recent study in brain tumour patients demonstrated that electrical stimulation of the inferior parietal cortex (IPC) causes a strong intention and desire to move, while higher intensities of stimulation in this area led to illusory movement awareness, thus providing insight that both motor intention and motor awareness emerge from activations of parietal regions, before motor execution (Desmurget et al., 2009).

Therefore, if a multivariate pattern classifier could be developed to detect movement intention and the type of movement from signals originating in PPC, such a system would be beneficial for rehabilitation of patients with movement disabilities. This study would furthermore point to an alternative region for operating neuroprosthetics, away from sensorimotor areas, which are often damaged in patients suffering from motor disability.

It has been pointed out (Haynes et al., 2007) that multivariate approaches are more sensitive in decoding brain states as they integrate spatial and temporal information from different regions of the brain, in contrast to univariate statistical parametric mapping (SPM) which analyses each brain location in isolation. Multivariate pattern classification of neuroimaging data is establishing itself as a powerful approach for decoding cognitive, affective and perceptual states, represented by the following recent studies: neural antecedents of voluntary movement (Soon et al., 2008), unconsciously perceived sensory stimuli (Haynes and Rees, 2006; Haynes et al., 2007), visual processing (Kamitani and Tong, 2005), information-based mapping (Kriegeskorte et al., 2006), memory recall (Polyn et al., 2005), detection of deception (Davatzikos et al., 2005) and emotion perception (Pessoa and Padmala, 2005).

In the present study, we hypothesised that different patterns of blood-oxygen-level dependent (BOLD) activations in PPC will be elicited by movement intention and that these different activations will be specific for the type of movement to be performed (left vs. right hand/arm). We further hypothesised that a multivariate pattern classifier could be used to decode movement intention and the type of movement based on the BOLD activity of PPC. Towards this end, we used an event-related fMRI paradigm in which movement intention is elicited in participants by informing them about the upcoming kinaesthetic motor imagery that they have to perform (left or right hand/arm), but explicitly instructing them to suppress the execution of imagery until indicated. This movement suppression paradigm requires participants to withhold an intended movement imagery, knowing the upcoming direction of movement. We also extended the analysis of multivariate classification to motor imagery in order to explore if it is possible to decode kinaesthetic imagery from PPC. Finally, we also compared the prediction accuracy of the multivariate pattern classifier for decoding movement intention and motor imagery between PPC, and other areas involved in motor planning and execution, namely

supplementary motor area (SMA), premotor cortex (PMC) and primary motor cortex (M1).

To investigate if differential patterns of activation are elicited for movement intention and motor imagery for left or right hand/arm, first we performed the traditional univariate SPM analysis, and to later demonstrate that such patterns could reliably be predicted using classification algorithms, we employed a spatio-temporal support vector machine (SVM).

Materials and Methods

Participants

Ten healthy participants (5 males/5 females, age range 21-26 years) were recruited. They were right handed as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971) and emmetropic. All participants gave informed consent to participate in the study that has been approved by the local ethics committee of the Faculty of Medicine of the University of Tübingen.

FMRI data acquisition

Experiments were conducted using a 3-Tesla MR Trio system (Siemens, Erlangen, Germany) with a standard 12 channel head coil. Functional image acquisition spanned the whole brain with 16 slices (voxel size=3.3×3.3×5.0 mm³, slice gap=1 mm). Slices were AC/PC aligned in axial orientation. A standard echo-planar imaging (EPI) sequence was used (TR=1.5 seconds, matrix size=64×64, effective echo time TE=30 ms, flip angle $\alpha=70^\circ$, bandwidth=1.954 kHz/pixel). For superposition of functional maps upon brain anatomy a high-resolution T1-weighted structural scan of the whole brain was collected from each participant (MPRAGE, matrix size=256×256, 160 partitions, 1 mm³ isotropic voxels, TR=2300 ms, TE=3.93 ms, TI=1100 ms, $\alpha=8^\circ$). In order to reduce movements two foam cushions immobilized the participant's head. Participants were instructed not to move, to remain relaxed and breathe regularly to avoid potential BOLD artefacts due to changes in the vasculature irrelevant to the task.

Experimental protocol

During the scanning sessions, the experimental protocol was visually presented to the participants by a mirror attached to the head-coil and using the Presentation software (Neurobehavioral Systems, Inc., CA, USA). The protocol consisted in an event-related design of successive runs composed of fixation, movement intention and imagery blocks, in that order, with varying fixation durations (see Fig. 1). Fixation block durations were pseudo-randomized to integral multiples of 1 TR (1.5 seconds), between 1 and 5 TR. Movement intention blocks of 1 TR were presented with an arrow (left or right) indicating the side of the forthcoming motor imagery. Participants were explicitly instructed not to imagine yet but to wait until the symbol indicating the onset of the imagery appeared on the screen. For the 4.5 seconds long imagery blocks participants were told to perform kinaesthetic motor imagery of hand/arm movement. The movement that had to be imagined incorporated a sequence of three sub-movements: reaching an imagery tool placed about 10 centimetres in front of the hand, grabbing it, and flexing the arm lifting the imagined object towards the ipsilateral shoulder. This sequence of movements was used because it has been shown that complex imagined movements produce stronger brain activations (Lotze et al., 1999; Wolbers et al., 2003). Participants mentally conducted these movements with right or left hand/arm. Prior to scanning, participants underwent an initial training task on the patient table of the scanner in the same position as during the experiment. The training consisted of a few trials of actual movement execution with each hand/arm to facilitate and to enhance the vividness of subsequent imagery (Thyryon and Roll, 2009)

[Fig.1 about here]

During the scanning, participants were instructed to keep their gaze steady at the center of the screen where the visual stimuli were presented. Stimuli were adjusted in visual attributes (color, size, general shape). For each participant four scanning sessions were conducted with each session being identical in its paradigm. One session consisted of 20 runs of fixation, movement intention, and imagery blocks, for each hand/arm, resulting in 40 trials in total. Left and right hand/arm trials were pseudo-randomized. Participants familiarized themselves with the protocol before the beginning of the scanning session.

Off-line data analysis

fMRI data analysis was performed using statistical parametric mapping 5 (SPM5) (<http://www.fil.ion.ucl.ac.uk/spm/software/spm5/>, Wellcome Department of Imaging Neuroscience) running on Matlab (MathWorks, Inc., Natick, MA). For spatial preprocessing and data analysis each of the four time-series per participant were realigned and resliced (3x3x3 mm) after unwarping in phase encoding direction (anterior/posterior) to account for magnetic susceptibility artefacts. The anatomical images were corrected for intensity inhomogeneities and then normalized to MNI space. A segmentation into white and gray matter was performed. The echoplanar images were then coregistered with the anatomical images. To achieve this, the mean of the realigned functional images was taken as a reference image. Finally, the realigned functional images were spatially normalized to the Montreal Neurological Institute (MNI) template. The resulting images were smoothed with an 8 mm (full width at half maximum (FWHM)) Gaussian filter. FMRI data were high-pass filtered (cut off period 128 seconds). For each individual participant, a general linear model (GLM) was applied to the time-course of each voxel (Worsley and Friston, 1995). The movement parameters estimated during the realignment procedure were introduced as covariates of no-interest into the general linear model.

Support Vector Classification

Preprocessing was performed with SPM5, and classification was performed using MATLAB (The Mathworks, Natick, MA) scripts.

To evaluate decoding accuracy from data of specific brain areas, brain masks of PPC, SMA, PMC, and M1 were created with WFU PickAtlas Toolbox. The masks were created by using Brodmann area (BA) as follows: BA 5 + 7 + 39 + 40 for PPC, mesial part of BA 6 for SMA, the remaining part of BA 6 for PMC, and BA 4 for M1. Particularly, for separation of the mask of the PMC from the mask of BA6, the mask of the SMA (in the aal labels of WFU PickAtlas Toolbox) was subtracted from that of BA6.

Input vectors containing BOLD activations in the masks for SVM training were prepared in the following manner. For intention and imagery conditions, BOLD values

from each individual brain area (PPC, SMA, PMC, and M1) of each single scan were collected into the input vector. For the fixation condition, BOLD values from each individual brain area were collected into an input vector as a mean scan of all the scans. In our investigation, the input vectors were derived from normalized MNI space without spatial and temporal smoothing. To classify the data between various conditions namely, intention (including left and right) vs. fixation, and left vs. right (i.e., left and right fixations, left and right intentions, and left and right imageries) across time points, SVM (linear SVM with the regularization parameter $C=1$, based on libSVM (<http://www.csie.ntu.edu.tw/~wcjlin/libsvm>)) was implemented. The fixation condition was artificially split into two separated conditions, i.e., left and right fixation, for the convenience of analysis. The design labels of the SVM classifier are defined, in general, as follows: input vectors from the condition A and B were taken as 1 and -1, respectively. The classification performance from data was evaluated through 4-fold cross validation (CV) (Hastie et al., 2001). In each fold, data of 3 sessions were used to train an SVM classifier, and then the data of 1 remaining session were used to test the classifier.

The pattern analysis accounted for the delay in the homodynamic response with respect to the stimulus onset by introducing an equivalent delay in the input data set. This was achieved by right-shifting the design labels by a factor of 3 TR (4.5 s) and by incorporating an additional 4.5 s of subsequent activation values at all the selected voxels.

Multivariate spatial analysis with Effect map

Based on the parameters of the trained SVM model, we analyzed the fMRI data with the Effect Mapping (EM) (Lee et al., in press). To identify informative voxels from the SVM model, the EM measures the effect of each voxel in multi-voxel space to the SVM output by considering two factors, namely, the input vectors and the weight vector; $y = \mathbf{w}^T \mathbf{x} + w_0$, where y is the SVM output, \mathbf{w} is the weight vector, and \mathbf{x} the input vector which determine the SVM output. The effect of each voxel on the classifier output is measured by computing normalized Mutual Information (NMI) between the voxel and the SVM output.

MI is defined as the amount of information that one random variable contains about another random variable (Cover and Thomas, 1991). That is, when two random variables X and Y occur with a joint probability mass function $p(x, y)$ and marginal probability function $p(x)$ and $p(y)$, the entropies of the two random variables and the joint probability are given respectively by:

$$\begin{aligned} H(X) &= \sum_{x \in X} -p(x) \log p(x), & H(Y) &= \sum_{y \in Y} -p(y) \log p(y), \\ \text{and } H(X, Y) &= \sum_{x \in X} \sum_{y \in Y} -p(x, y) \log p(x, y) \end{aligned} \quad (1)$$

MI, $I(X; Y)$, is the relative entropy between the joint distribution and the product distribution, i.e.,

$$I(X; Y) = H(X) + H(Y) - H(X, Y) \quad (2)$$

To correct for variance of mutual information based on entropies $H(X)$ and $H(Y)$, normalized mutual information is defined as (Maes et al., 1997):

$$\tilde{I}(X; Y) = \frac{I(X; Y)}{H(X) + H(Y)} \quad (3)$$

Hence, the effect value (EV) E_k of a voxel k is defined as:

$$E_k = w_k \tilde{I}(x_k; y), \quad k = 1, \dots, M \quad (\text{M: number of voxels}) \quad (4)$$

where y is the SVM output after excluding the sign function, w_k and x_k are the SVM weight value and activation in voxel k , respectively.

After normalizing the absolute value of E_k from Eq. (4), we obtain the relation:

$$nE_k = \text{sgn}(E_k) \log(1 + |E_k| / \text{std}(|E|)), \quad k = 1, \dots, M \quad (5)$$

where $\text{sgn}(\cdot)$ is a sign function, and $\text{std}(|E|)$ is the standard deviation of all E_k . In the present study, Eq. (5) (nE_k) was used to compute the EV at each voxel to make E(ffect)-maps from different subjects and fold of CV be comparable.

With different contrasts, i.e., intention vs. fixation, and left vs. right over time points (left and right fixations, left and right intentions, and left and right imaginations), E-maps were separately obtained from data taken together from all the 4 areas, (i.e., PPC, PMC,

SMA, and M1) of the participants. The E-maps from a contrast for 4-fold CV of all subjects (i.e., 40 E-maps; 4 E-maps from 4-fold CV and 10 participants) were averaged into an E-map for a group analysis, and then the averaged map was smoothed spatially with 5 mm Fixed Width Half Maximum (FWHM) to minimize distortion of the map for ease of interpretation. In the interpretation of the E-map, positive and negative EVs were related the design labels, 1 and -1, of the SVM classifier, respectively. That is, if the design labels of two conditions are exchanged, the sign of EVs are also reversed.

Results

Univariate analysis

First, we investigated the brain activations elicited during movement intention. A group level analysis showed that movement intention elicited high activations in SMA and IPC (angular gyrus and supramarginal gyrus) for both left and right hand/arm (see Table 1)

[Table1 about here]

A subtraction analysis performed to check whether differential activation patterns for left and right hand/arm movement intention are statistically significant revealed a lateralization effect for fronto - parietal areas (see Fig. 2).

[Fig. 2 about here]

Multivariate Analysis

Classification accuracies between intention and fixation were assessed to check whether movement intention for left and right hand/arm can be discriminated from fixation. The analysis was performed separately with and without the hemodynamic delay. Even without the inclusion of the hemodynamic delay, the decoding accuracies in all the areas (PPC, SMA, PMC, and M1) were greater than 50% (see Fig. 3A). In addition, E-maps in the 2 time points, namely at zero delay and at 3 scan delay, showed clear difference in activation patterns of the SMA. The E-map without hemodynamic delay shows negative

EVs in SMA, while the E-map with hemodynamic delay showed positive EVs in this area (see Fig. 3B).

[Fig.3 about here]

To reveal whether functional activation patterns between left and right conditions in the 4 areas can be discriminated, decoding accuracies of the 4 areas across time points were evaluated. In this analysis, decoding accuracies increased across time points after the onset time of motor imagery (see Fig. 4A). With consideration of the hemodynamic lag (4.5s; 3 TR), decoding classification accuracy at onset time of intention (T4) is greater than chance level for all regions of interest. In addition, the PPC, PMC and SMA show higher accuracies than M1 during movement intention and imagery.

[Fig.4 about here]

Discussion

The main aim of this study was to examine whether a multivariate pattern classifier can be used to decode movement intention from the human PPC. In order to elicit a repeatable movement intention, we designed a paradigm in which participants were required to suppress movement, with the knowledge of the movement direction, and the knowledge that they will be required to move in a short while. Univariate methodology was used to visualize brain activations elicited by movement intention. Supporting our hypothesis, SPM analysis showed that PPC (particularly inferior parietal lobule) was highly activated prior to the execution of motor imagery, indicating involvement of this area in movement intention. The importance of PPC in movement has also been emphasised by several studies performed in the last years that have suggested that PPC acts as “sensorimotor interface”, converting sensory information into motor commands (“inverse transformation model”) and integrating sensory input with previous and ongoing motor commands (“forward model”), thus playing a crucial role in movement planning, online movement, and maybe future movement plans (for a review see: (Buneo and Andersen, 2006).

SMA was also highly activated during movement intention blocks. This pattern of activation is concordant with electrophysiological and neuroimaging studies that have

shown the involvement of SMA (and pre-SMA) in overt movement preparation (Deiber et al., 1996; Michelon et al., 2006; Simon et al., 2002) and in motor imagery preparation (Caldara et al., 2004; Cunnington et al., 2005; Jankelowitz and Colebatch, 2002). The above studies supported the idea that both motor execution and motor imagery involve similar pre-movement brain processes and that SMA codes motor information for the “readiness for movement”, for overt or covert execution (Cunnington et al., 2005). Furthermore, electrical stimulation of this area can elicit an “urge to move”, without overt motor activity (Fried et al., 1991). Hence, it has been proposed that SMA could also play a direct role in motor intention (Haggard, 2005, 2008; Lau et al., 2004).

Our results also showed that movement intention is lateralized in the PPC and SMA, pointing to the potential value of movement intention decoding for BCI and neuro-prosthetics.

Two main analyses were performed with the multivariate method. First, we examined whether it was possible to discriminate movement intention vs. fixation. For all four target areas (PPC, SMA, PMC and M1) prediction accuracies were above chance level. As we expected for the role of PPC in movement intention, this area had especially higher prediction values, although much variation was seen among subjects. The high prediction accuracy obtained for movement intention for PMC is congruent with its documented role in motor preparation and planning (for reviews see: (Abe and Hanakawa, 2009; Hoshi and Tanji, 2004, 2007). The fact that this finding was not reflected as PMC activations in univariate analysis, points to SVM's ability to pickup discriminating voxels which SPM misses.

If the hemodynamic delay is considered, the prediction accuracy for the discrimination between intention to move vs. fixation in the target areas improves, reaching values in the range 75- 85% for PPC, PMC and SMA. The significance of including the hemodynamic delay is reflected best in the classification performance and discriminating voxels in the SMA. When no hemodynamic delay is applied (zero delay), SMA shows a prominent deactivation (see Fig. 3b), barely detecting the intention condition. However, as the delay is gradually increased to 3 TR, SMA shows an increasing activation, and improved discrimination for the intention condition.

The different results obtained with and without the application of the hemodynamic delay can have important implications for BCI classification. The values reported for prediction accuracy even without adjusting for the sluggish BOLD response were above chance, showing that an online BCI or a prosthetic based on movement intention can be feasible. However, according to our results, the multivariate pattern classifier can only decode with higher accuracies if the classifier is remodelled in consideration of the delay. A BCI based on electrical brain activity, measured by electroencephalography (EEG), magnetoencephalography (MEG), electrocorticograph (EcoG) or implanted electrodes, which are not affected by the hemodynamic delay but has adequate resolution to extract signals from PPC or SMA, may overcome the above limitations and hence could be more suitable for clinical applications.

The second aspect examined by SVM analysis was whether the multivariate classifier can discriminate between right and left hand/arm movement intention. In general, the mean prediction accuracy of PPC showed above chance values across different time points. With the incorporation of the hemodynamic delay, and considering T4 as the onset of movement intention, prediction accuracy for left and right movement intention improves for all target areas reaching a range of 70-80% in PPC. Also, after incorporating the hemodynamic delay, high classification accuracy values (> 70%) for left and right hand/arm motor imagery was seen for PPC, SMA and PMC. This observation is again congruent with the prior literature that has shown that premotor regions and PPC are actively involved in motor imagery (Fleming et al., 2009; Gerardin et al., 2000; Hanakawa et al., 2003; Lotze et al., 1999).

Another aspect to consider is that for both movement intention vs. fixation and left vs. right movement intention/imagery classification, a general tendency of higher prediction accuracies for PPC, PMC and SMA compared to that for M1 was found across almost every time point. These findings are concordant with the idea that premotor and parietal regions play a predominant role in different aspects of action planning (Andersen and Cui, 2009; Hanakawa et al., 2008) and with the observation that primary motor cortex is inconsistently activated during motor imagery, and

usually in a lower degree than during motor execution (Dechent et al., 2004; Gerardin et al., 2000; Hanakawa et al., 2008), and for a review see:(Munzert et al., 2009).

Our work has limitations. Although participants were explicitly instructed at the beginning of each new run of the experiment protocol that they should avoid movement or imagery during the intention block, we could not ascertain whether they followed the instructions by any physiological measurement, such as electromyography (EMG). With regard to the inclusion of hemodynamic delay in our multivariate analysis, we applied a fixed value of the delay for the different brain areas that were investigated, although recent evidence suggests that the hemodynamic delay may not have the same properties across different regions of the brain (Logothetis, 2008). The brain masks for PPC, SMA, PMC and M1 were generated from an anatomical template and were not derived from subject specific activations, i.e., by a functional localizer. The unspecific nature of these masks could have adversely affected the decoding performance to some extent.

What implications do these findings have for BCI and motor rehabilitation?

Our study has shown that it is possible to decode movement intention and the type of movement with high prediction accuracy from PPC and premotor regions involved in motor planning. As movement intention occurs before motor execution and motor imagery, our results show that it is feasible to use signals from these areas in anticipation of motor imagery. In patients with frontal lesions or with difficulties to perform motor imagery, decoding movement intention from PPC can have practical benefits. If a paradigm similar to the one used here is implemented, a portable BCI based on EEG or near infrared spectroscopy (NIRS) (Birbaumer, 2006; Sitaram et al., 2007) could be built to operate a neuro-prosthetic device as an “intention BCI”.

Acknowledgement

This work was supported by grants from the SFB437 “Kriegserfahrungen”, of the Deutsche Forschungsgemeinschaft, DFG BI 195/59-1, DFG BI 195/56-1 and DFG KU 1453/3-1. Author S.R is thankful to the Faculty of Medicine, Pontificia Universidad Católica de Chile, and Conycit for supporting this research. Author S. L. is grateful to DAAD (German Academic Exchange Service) for its support.

References

- Abe, M., Hanakawa, T., 2009. Functional coupling underlying motor and cognitive functions of the dorsal premotor cortex. *Behav Brain Res* 198, 13-23.
- Andersen, R.A., Buneo, C.A., 2002. Intentional maps in posterior parietal cortex. *Annu Rev Neurosci* 25, 189-220.
- Andersen, R.A., Cui, H., 2009. Intention, action planning, and decision making in parietal-frontal circuits. *Neuron* 63, 568-583.
- Berti, A., Bottini, G., Gandola, M., Pia, L., Smania, N., Stracciari, A., Castiglioni, I., Vallar, G., Paulesu, E., 2005. Shared cortical anatomy for motor awareness and motor control. *Science* 309, 488-491.
- Birbaumer, N., 2006. Breaking the silence: brain-computer interfaces (BCI) for communication and motor control. *Psychophysiology* 43, 517-532.
- Birbaumer, N., Cohen, L.G., 2007. Brain-computer interfaces: communication and restoration of movement in paralysis. *J Physiol* 579, 621-636.
- Buch, E., Weber, C., Cohen, L.G., Braun, C., Dimyan, M.A., Ard, T., Mellinger, J., Caria, A., Soekadar, S., Fourkas, A., Birbaumer, N., 2008. Think to move: a neuromagnetic brain-computer interface (BCI) system for chronic stroke. *Stroke* 39, 910-917.
- Buneo, C.A., Andersen, R.A., 2006. The posterior parietal cortex: sensorimotor interface for the planning and online control of visually guided movements. *Neuropsychologia* 44, 2594-2606.
- Caldara, R., Deiber, M.P., Andrey, C., Michel, C.M., Thut, G., Hauert, C.A., 2004. Actual and mental motor preparation and execution: a spatiotemporal ERP study. *Exp Brain Res* 159, 389-399.
- Cover, T.M., Thomas, J.A., 1991. *Elements of information theory*. Wiley, New York.
- Cunnington, R., Windischberger, C., Moser, E., 2005. Premovement activity of the pre-supplementary motor area and the readiness for action: studies of time-resolved event-related functional MRI. *Hum Mov Sci* 24, 644-656.
- Davatzikos, C., Ruparel, K., Fan, Y., Shen, D.G., Acharyya, M., Loughead, J.W., Gur, R.C., Langleben, D.D., 2005. Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. *Neuroimage* 28, 663-668.

- Dechent, P., Merboldt, K.D., Frahm, J., 2004. Is the human primary motor cortex involved in motor imagery? *Brain Res Cogn Brain Res* 19, 138-144.
- Deiber, M.P., Ibanez, V., Sadato, N., Hallett, M., 1996. Cerebral structures participating in motor preparation in humans: a positron emission tomography study. *J Neurophysiol* 75, 233-247.
- Desmurget, M., Reilly, K.T., Richard, N., Szathmari, A., Mottolese, C., Sirigu, A., 2009. Movement intention after parietal cortex stimulation in humans. *Science* 324, 811-813.
- Fleming, M.K., Stinear, C.M., Byblow, W.D., 2009. Bilateral parietal cortex function during motor imagery. *Exp Brain Res*.
- Fogassi, L., Ferrari, P.F., Gesierich, B., Rozzi, S., Chersi, F., Rizzolatti, G., 2005. Parietal lobe: from action organization to intention understanding. *Science* 308, 662-667.
- Fried, I., Katz, A., McCarthy, G., Sass, K.J., Williamson, P., Spencer, S.S., Spencer, D.D., 1991. Functional organization of human supplementary motor cortex studied by electrical stimulation. *J Neurosci* 11, 3656-3666.
- Gerardin, E., Sirigu, A., Lehericy, S., Poline, J.B., Gaymard, B., Marsault, C., Agid, Y., Le Bihan, D., 2000. Partially overlapping neural networks for real and imagined hand movements. *Cereb Cortex* 10, 1093-1104.
- Gonzalez, B., Rodriguez, M., Ramirez, C., Sabate, M., 2005. Disturbance of motor imagery after cerebellar stroke. *Behav Neurosci* 119, 622-626.
- Haggard, P., 2005. Conscious intention and motor cognition. *Trends Cogn Sci* 9, 290-295.
- Haggard, P., 2008. Human volition: towards a neuroscience of will. *Nat Rev Neurosci* 9, 934-946.
- Hanakawa, T., Dimyan, M.A., Hallett, M., 2008. Motor planning, imagery, and execution in the distributed motor network: a time-course study with functional MRI. *Cereb Cortex* 18, 2775-2788.
- Hanakawa, T., Immisch, I., Toma, K., Dimyan, M.A., Van Gelderen, P., Hallett, M., 2003. Functional properties of brain areas associated with motor execution and imagery. *J Neurophysiol* 89, 989-1002.
- Hastie, T., Tibshirani, R., Friedman, J.H., 2001. *The elements of statistical learning : data mining, inference, and prediction : with 200 full-color illustrations*. Springer, New York.

- Haynes, J.D., Rees, G., 2006. Decoding mental states from brain activity in humans. *Nat Rev Neurosci* 7, 523-534.
- Haynes, J.D., Sakai, K., Rees, G., Gilbert, S., Frith, C., Passingham, R.E., 2007. Reading hidden intentions in the human brain. *Curr Biol* 17, 323-328.
- Hoshi, E., Tanji, J., 2004. Functional specialization in dorsal and ventral premotor areas. *Prog Brain Res* 143, 507-511.
- Hoshi, E., Tanji, J., 2007. Distinctions between dorsal and ventral premotor areas: anatomical connectivity and functional properties. *Curr Opin Neurobiol* 17, 234-242.
- Jankelowitz, S.K., Colebatch, J.G., 2002. Movement-related potentials associated with self-paced, cued and imagined arm movements. *Exp Brain Res* 147, 98-107.
- Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. *Nat Neurosci* 8, 679-685.
- Kriegeskorte, N., Goebel, R., Bandettini, P., 2006. Information-based functional brain mapping. *Proc Natl Acad Sci U S A* 103, 3863-3868.
- Lau, H.C., Rogers, R.D., Haggard, P., Passingham, R.E., 2004. Attention to intention. *Science* 303, 1208-1210.
- Lee, S., Halder, S., Kübler, A., Birbaumer, N., Sitaram, R., in press. Effective functional mapping of fMRI data with support vector machines. *Human Brain Mapping*.
- Logothetis, N.K., 2008. What we can do and what we cannot do with fMRI. *Nature* 453, 869-878.
- Lotze, M., Montoya, P., Erb, M., Hulsmann, E., Flor, H., Klose, U., Birbaumer, N., Grodd, W., 1999. Activation of cortical and cerebellar motor areas during executed and imagined hand movements: an fMRI study. *J Cogn Neurosci* 11, 491-501.
- Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P., 1997. Multimodality image registration by maximization of mutual information. *IEEE Trans Med Imaging* 16, 187-198.
- Malouin, F., Richards, C.L., Desrosiers, J., Doyon, J., 2004. Bilateral slowing of mentally simulated actions after stroke. *Neuroreport* 15, 1349-1353.
- Mazzoni, P., Bracewell, R.M., Barash, S., Andersen, R.A., 1996. Motor intention activity in the macaque's lateral intraparietal area. I. Dissociation of motor plan from sensory memory. *J Neurophysiol* 76, 1439-1456.

Michelon, P., Vettel, J.M., Zacks, J.M., 2006. Lateral somatotopic organization during imagined and prepared movements. *J Neurophysiol* 95, 811-822.

Mulliken, G.H., Musallam, S., Andersen, R.A., 2008. Decoding trajectories from posterior parietal cortex ensembles. *J Neurosci* 28, 12913-12926.

Munzert, J., Lorey, B., Zentgraf, K., 2009. Cognitive motor processes: the role of motor imagery in the study of motor representations. *Brain Res Rev* 60, 306-326.

Musallam, S., Corneil, B.D., Greger, B., Scherberger, H., Andersen, R.A., 2004. Cognitive control signals for neural prosthetics. *Science* 305, 258-262.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97-113.

Pessoa, L., Padmala, S., 2005. Quantitative prediction of perceptual decisions during near-threshold fear detection. *Proc Natl Acad Sci U S A* 102, 5612-5617.

Polyn, S.M., Natu, V.S., Cohen, J.D., Norman, K.A., 2005. Category-specific cortical activity precedes retrieval during memory search. *Science* 310, 1963-1966.

Quiñero, R., Snyder, L.H., Batista, A.P., Cui, H., Andersen, R.A., 2006. Movement intention is better predicted than attention in the posterior parietal cortex. *J Neurosci* 26, 3615-3620.

Sabate, M., Gonzalez, B., Rodriguez, M., 2007. Adapting movement planning to motor impairments: the motor-scanning system. *Neuropsychologia* 45, 378-386.

Sakata, H., Taira, M., Murata, A., Mine, S., 1995. Neural mechanisms of visual guidance of hand action in the parietal cortex of the monkey. *Cereb Cortex* 5, 429-438.

Simon, S.R., Meunier, M., Pieltre, L., Berardi, A.M., Segebarth, C.M., Boussaoud, D., 2002. Spatial attention and memory versus motor preparation: premotor cortex involvement as revealed by fMRI. *J Neurophysiol* 88, 2047-2057.

Sirigu, A., Daprati, E., Ciancia, S., Giraux, P., Nighoghossian, N., Posada, A., Haggard, P., 2004. Altered awareness of voluntary action after damage to the parietal cortex. *Nat Neurosci* 7, 80-84.

Sitaram, R., Zhang, H., Guan, C., Thulasidas, M., Hoshi, Y., Ishikawa, A., Shimizu, K., Birbaumer, N., 2007. Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing a brain-computer interface. *Neuroimage* 34, 1416-1427.

- Snyder, L.H., Batista, A.P., Andersen, R.A., 1997. Coding of intention in the posterior parietal cortex. *Nature* 386, 167-170.
- Snyder, L.H., Batista, A.P., Andersen, R.A., 2000. Intention-related activity in the posterior parietal cortex: a review. *Vision Res* 40, 1433-1441.
- Soon, C.S., Brass, M., Heinze, H.J., Haynes, J.D., 2008. Unconscious determinants of free decisions in the human brain. *Nat Neurosci* 11, 543-545.
- Thyriion, C., Roll, J.P., 2009. Perceptual integration of illusory and imagined kinesthetic images. *J Neurosci* 29, 8483-8492.
- Wolbers, T., Weiller, C., Buchel, C., 2003. Contralateral coding of imagined body parts in the superior parietal lobe. *Cereb Cortex* 13, 392-399.
- Wolpaw, J.R., Birbaumer, N., McFarland, D.J., Pfurtscheller, G., Vaughan, T.M., 2002. Brain-computer interfaces for communication and control. *Clin Neurophysiol* 113, 767-791.
- Worsley, K.J., Friston, K.J., 1995. Analysis of fMRI time-series revisited--again. *Neuroimage* 2, 173-181.

Figure Legends

Fig. 1. Stimulus presentation paradigm. The paradigm consisted of a fixation block varying in length between 1 and 5 TR (pseudo-randomized) followed by randomized right or left hand/arm movement intention block and kinaesthetic motor imagery. During the onset of the movement intention block, the participants received an indication of the chosen side for the forthcoming imagery. Movement intention and imagery durations were fixed with 1.5 and 4.5 seconds respectively.

Fig. 2. Movement intention lateralization. Areas more activated for the contrast left > right movement intention included activations in right SMA (MNI coordinates: 15,-23, 66; $t = 11.47$), right inferior parietal lobule (MNI coordinates: 34,-45,46; $t = 9.01$), and right superior parietal lobule (MNI coordinates: 26,-60,55; $t = 8.67$). For the contrast right > left movement intention the highest activations included left primary motor cortex (MNI coordinates: -26,-29,61; $t = 9.75$), left rolandic operculum (MNI coordinates: -44,-22,11; $t = 7.84$), and left superior parietal lobule (MNI coordinates: -28,-58,52; $t = 7.81$). Results are depicted on a standard brain (sagittal view; left column depicting the right hemisphere; right column depicting the left hemisphere). Subtraction analysis was performed on the group level using data acquired via second-level random effects analysis, $P < 0.001$, uncorrected. Colour bar represents t -values.

Fig. 3. Decoding accuracy of fixation vs. movement intention in 4 brain areas (PPC, SMA, PMC, M1) and the corresponding E(ffect)-maps. (A) Classification accuracy over 10 participants (% mean accuracy \pm % standard deviation). The classification accuracy was computed from 40 classification accuracies (i.e., 4-fold CV X 10 participants). The figure shows two decoding accuracies: without consideration of hemodynamic delay (i.e., at the onset of stimulus) and with consideration of hemodynamic response (i.e., 3 scan delay (4.5s)). (B) E-maps corresponding to the decoding accuracies with hemodynamic delay of 0 and 3 scans. In the computation of the E-maps, design labels of the intention and fixation for a SVM classifier are given as 1 and -1, respectively. E-maps are drawn by selecting 10% of voxels having the highest EV. For the purpose of display, positive and negative EVs are coded into red and blue color.

Fig. 4. Decoding accuracy of condition left vs. right hand/arm in the specified brain area across all conditions (intention and imagery) and the corresponding E-maps. (A) Classification accuracy over 10 participants across conditions (% mean accuracy \pm % standard deviation). The mean accuracy and standard deviation were computed from 40 classification accuracies (i.e., 4-fold CV X 10 participants). At the point T0, classification accuracy is calculated from data obtained from left and right fixation. T1 and T2 indicate the onset of intention and imagery conditions. T4 (red broken line) and T5 (blue broken line) represent the onsets of intention and imagery conditions, respectively, in the presence of hemodynamic delay. (B) E-map in T5. This map shows 6 horizontal slices at spatial interval of 6 mm obtained from the 4 areas (i.e., PPC, SMA, PMC, and M1) in MNI space. In computation of the E-maps, design labels of the left and right hand/arm for a SVM classifier are assigned as 1 and -1, respectively. After selecting top 20% of voxels having the highest EV, the E-maps are drawn for the purpose of display. Positive and negative EVs are shown as into red and blue colored pixels.

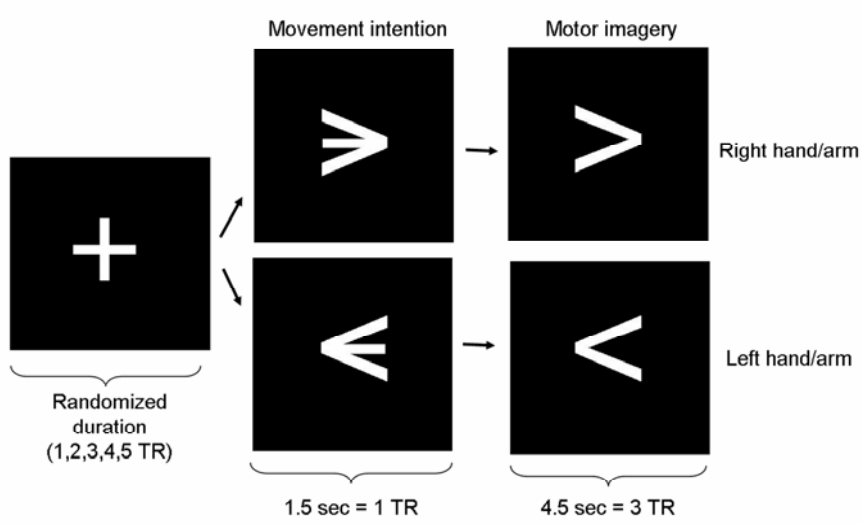


Fig. 1

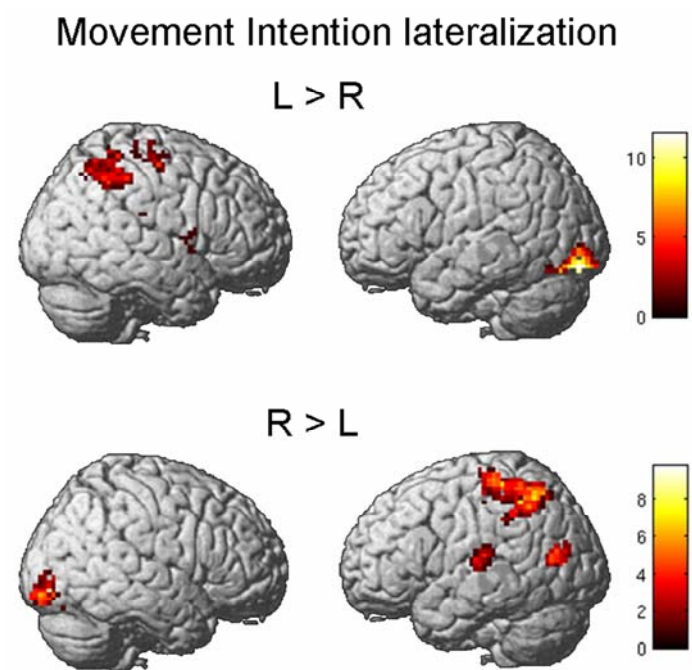


Fig. 2

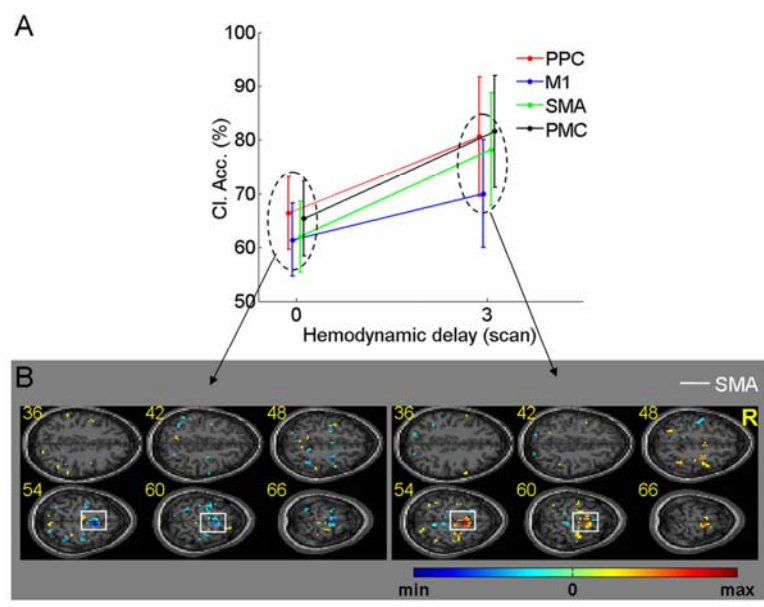


Fig. 3

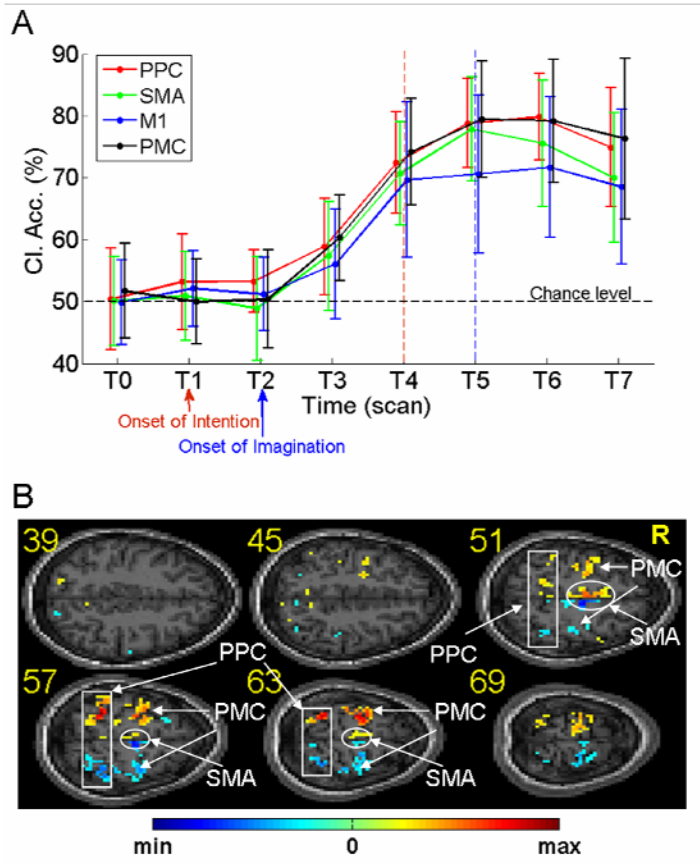


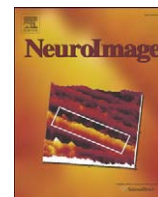
Fig. 4

4. 5. Multivariate prediction of movement intention in the human fronto-parietal cortex



Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Real-time support vector classification and feedback of multiple emotional brain states

Ranganatha Sitaram^{a,*}, Sangkyun Lee^{a,b,*}, Sergio Ruiz^{a,b,d}, Mohit Rana^{a,b}, Ralf Veit^a, Niels Birbaumer^{a,c}

^a Institute of Medical Psychology and Behavioral Neurobiology, University of Tübingen, Gartenstr. 29, 72074 Tübingen, Germany

^b Graduate School of Neural & Behavioural Sciences, International Max Planck Research School, 72074 Tübingen, Germany

^c Ospedale San Camillo, Istituto di Ricovero e Cura a Carattere Scientifico, Venezia, Lido, Italy

^d Department of Psychiatry, Faculty of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile

ARTICLE INFO

Article history:

Received 25 February 2010

Revised 31 July 2010

Accepted 3 August 2010

Available online xxxx

ABSTRACT

An important question that confronts current research in affective neuroscience as well as in the treatment of emotional disorders is whether it is possible to determine the emotional state of a person based on the measurement of brain activity alone. Here, we first show that an online support vector machine (SVM) can be built to recognize two discrete emotional states, such as happiness and disgust from fMRI signals, in healthy individuals instructed to recall emotionally salient episodes from their lives. We report the first application of real-time head motion correction, spatial smoothing and feature selection based on a new method called Effect mapping. The classifier also showed robust prediction rates in decoding three discrete emotional states (happiness, disgust and sadness) in an extended group of participants. Subjective reports ascertained that participants performed emotion imagery and that the online classifier decoded emotions and not arbitrary states of the brain. Offline whole brain classification as well as region-of-interest classification in 24 brain areas previously implicated in emotion processing revealed that the frontal cortex was critically involved in emotion induction by imagery. We also demonstrate an fMRI-BCI based on real-time classification of BOLD signals from multiple brain regions, for each repetition time (TR) of scanning, providing visual feedback of emotional states to the participant for potential applications in the clinical treatment of dysfunctional affect.

© 2010 Elsevier Inc. All rights reserved.

Introduction

Prediction of emotional states from brain activity constitutes a major scope of affective neuroscience and would solve several pressing clinical problems such as the assessment of affect in verbally incompetent people with dementia, minimally conscious state and locked-in-syndrome, and the detection of deception. Recent advances in multivariate pattern classification of functional magnetic resonance imaging (fMRI) signals are especially important due to the high spatial resolution, whole brain coverage and non-invasiveness of fMRI. It has been pointed out (Haynes and Rees, 2006) that multivariate approaches are more sensitive in decoding brain states as they integrate spatial and temporal information from different regions of the brain, in contrast to univariate statistical parametric mapping (SPM) which analyses each brain location in isolation.

Studies on pattern classification of fMRI signals can be grouped into three major themes. Firstly, a number of methodological studies aimed to incorporate and adapt the existing methods in the field of machine learning to the classification of fMRI signals (LaConte et al., 2003, 2005; Martinez-Ramon et al., 2006; Shaw et al., 2003; Strother et al., 2004). LaConte and colleagues examined the classification of block-design fMRI data using linear discriminant analysis (LDA) (LaConte et al., 2003) and support vector machines (SVM) (LaConte et al., 2005) in contrast to canonical variates analysis (CVA). Mourao-Miranda and colleagues (2005) compared SVM and the Fisher linear discriminant (FLD) classifier and demonstrated that SVM outperforms FLD in prediction accuracy as well as in robustness of the spatial maps obtained. Shaw and colleagues (2003) showed that preprocessing strategies such as spatial and temporal smoothing and classification parameters could be derived in a subject-specific manner to result in optimal prediction. Martinez-Ramon and colleagues (2006) later developed an approach for multi-class classification by segmenting partially normalized activation maps into functional areas using a neuroanatomical atlas, classified each area separately with local classifiers, and finally performed a weighted aggregation of the multiclass outputs.

A second topic of research pertains to the application of pattern classification for obtaining greater insight into spatial and temporal

* Corresponding authors. Institute of Medical Psychology and Behavioural Neurobiology, Eberhard-Karls-University of Tübingen, Gartenstr. 29, D-72074 Tübingen, Germany. Fax: +49 7071 295956.

E-mail addresses: sitaram.ranganatha@uni-tuebingen.de (R. Sitaram),

lee.sangkyun@gmail.com (S. Lee).

¹ Authors contributed equally.

patterns of brain activity during cognitive, affective and perceptual states, represented by the following recent studies: neural antecedents of voluntary movement (Soon et al., 2008), visual processing (Kamitani and Tong, 2005), memory recall (Polyn et al., 2005), detection of deception (Davatzikos et al., 2005) and emotion perception (Pessoa and Padmala, 2005).

A third class of studies in pattern classification of brain signals, to which the present work belongs, is related to the rapidly advancing field of brain–computer interfaces (BCIs) and neurorehabilitation. A series of studies (Caria et al., 2006; deCharms et al., 2004, 2005; Posse et al., 2003; Rota et al., 2006; Sitaram et al., 2005; Veit et al., 2006; Weiskopf et al., 2003, 2004; Yoo et al., 2004; Yoo and Jolesz, 2002) with fMRI-BCIs demonstrated that healthy individuals as well as patients can learn voluntary regulation of localized brain regions and presented evidence for behavioural modifications that accompany self-regulation training. If the neurobiological basis of a disorder is known in terms of abnormal activity in certain regions of the brain, fMRI-BCI can be potentially targeted to modify activity in those regions with high specificity for treatment. However, a disadvantage of the existing fMRI-BCIs is in the restriction to one single region of interest (ROI) for extracting signals and providing feedback of activation to the participant. A major argument for moving away from deriving feedback signals from single ROIs is that perceptual, cognitive or emotional states generally recruit a distributed network of brain regions rather than single locations. Training subjects to merely increase or decrease BOLD signals in single ROIs may not completely model the network dynamics of the target brain state. Applying multivariate pattern classification presents itself as a potential solution to this problem as it does not make prior assumptions about functional localization and cognitive strategy. However, due to the limited computational time and resources available during online decoding, much of the offline classification methods developed so far might not be directly applicable to real-time fMRI. LaConte and colleagues (2007) reported the first real-time fMRI system with multivariate classification. Using SVM, the authors showed the feasibility of online decoding and feedback from single (repetition time, TR) fMRI data during block-design left- and right-hand motor imagery and further demonstrated the classifier's ability to decode other forms of cognitive and emotional states. The present study is in line with the above work and extends by demonstrating robust online classification and feedback of multiple emotional states.

fMRI-BCI is a promising tool for affective neuroscience and has shown potential for neurorehabilitation to alleviate emotional disorders such as anxiety, sociopathy, chronic pain and schizophrenia and brain communication in the locked-in syndrome (Caria et al., 2007; deCharms et al., 2005; Rota et al., 2008; Ruiz et al., 2008). However, Phan et al. (2002) in a meta-analysis indicated that emotions such as happiness, sadness, fear, anger and disgust activate networks of several brain regions such as insula, amygdala, hippocampus, thalamus, medial and lateral prefrontal cortex, orbitofrontal cortex and anterior and posterior cingulate. In consideration of the above evidence and the limitations of single ROI feedback, we intended to implement a high performance real-time SVM classifier that could decode as well as provide feedback of emotional states of the brain from fMRI signals. It would be beneficial if the classifier could learn to generalize across intra-subject variations of emotion regulation.

The purpose of the present study is to demonstrate that multiple discrete emotional states (e.g., happy, disgust and sad) could be classified and the information could be fed back to the participant in real time. We first demonstrate two-class classification between happy and disgust emotions as these two emotions are quite distinct from each other in emotional valence. Further, these emotions are interesting from the point of view of clinical applications—namely, rehabilitation in patients with dysfunctional affect. A novelty of the present work is the demonstration of the generalization of classification by training the classifier on one type of imagery or scenario and testing it on another type of imagery. In other words, we show that the classifier is able to

distinguish between two different imageries, e.g., winning a lottery and playing with a pet, as the same emotion, i.e., happy.

However, performing two-class (happy vs. disgust) classification alone may raise the question whether the classification is between positive and negative emotions in general, or whether specific discrete emotions could indeed be classified. To answer this question, we further aimed to demonstrate multiclass classification of 3 discrete emotional states, namely, happy, disgust and sad. We chose sad as the third emotion to show that the classifier can distinguish between two emotions that are similar in valence (e.g., negative), namely, disgust and sad.

We first demonstrate in 12 healthy volunteers, a high performance, real-time, two-class prediction of positive (happy) and negative (disgust) emotions and visual feedback of the predicted brain states for each TR. We show that robust classification performance could be obtained even with limited computational time by the application of preprocessing (realignment, detrending and spatial smoothing) and a method of feature selection called Effect mapping. Our previous work (Lee et al., 2010) has already reported results of rigorous comparisons between the Effect mapping method and some of the existing methods, showing clearly that the Effect mapping method fares better in classification performance.

Further to the two-class classification, we illustrate for the first time a multiclass (happiness, disgust and sadness) prediction and neurofeedback on 4 more healthy volunteers. To test whether classifier-based feedback training can help participants to improve emotion regulation, we trained two additional healthy participants for three sessions of feedback training. Finally, to obtain greater insight into the brain regions activated in emotion imagery and regulation, we carried out the following offline analyses: (1) SVM classification and Effect mapping of whole brain fMRI signals and (2) ROI classification of fMRI signals extracted from 24 brain regions previously implicated in emotion imagery and regulation.

Methods

Our aim was to develop a system that could train a classification model based on an initial set of fMRI data and brain state labels determined by the experimental protocol (training data set), thereafter to use the classification model to predict brain states from every volume of brain images (at the end of each repetition time, TR) that are acquired from the scanner, and to subsequently update the visual feedback based on the classifier's prediction.

Real-time Implementation

The real-time brain state classification system (Fig. 1) comprises of the following subsystems: (1) image acquisition subsystem written in the C programming language, (2) fMRI-BCI that performs image preprocessing, brain state classification and visual feedback, coded in a combination of C programming language and Matlab scripting language (Mathworks, Natwick, MA). The image acquisition system is centered around a modified echo-planar imaging (EPI) sequence developed in-house on the Siemens 3T TIM whole body scanner (Siemens Magnetom Trio Tim, Siemens, Erlangen, Germany), based on the image acquisition software, Syngo version VB13. We used the same real-time EPI sequence that was employed in our earlier studies on real-time fMRI (Rota et al., 2006, 2008; Weiskopf et al., 2003). The original implementation was described in Weiskopf et al. (2003). In brief, a standard echo-planar imaging sequence provided by the firm Siemens (Erlangen, Germany) was modified in collaboration with the manufacturer. The modifications enabled storage of functional image files in the ANALYZE format (*.hdr, *.img) to a specified directory of the host computer of the MRI scanner. The EPI sequence, otherwise, had facilities for changing the MR parameters, such as repetition time (TR), echo time (TE), number of slices and so on, similar to a standard

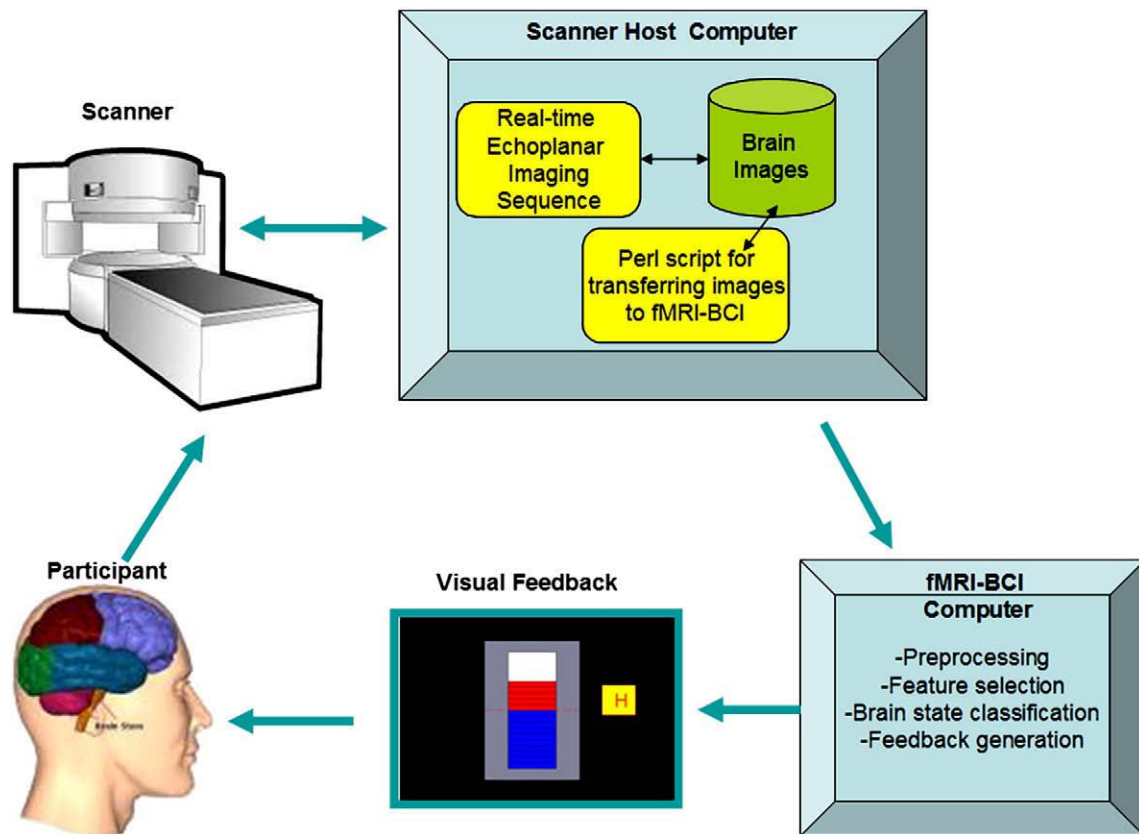


Fig. 1. The real-time fMRI brain state classification system comprises of the following subsystems: (1) image acquisition subsystem, which is a modified version of the standard echo-planar imaging (EPI) sequence written in C and executed on the scanner host computer, and (2) fMRI-BCI subsystem, which performs image preprocessing, brain state classification and visual feedback, implemented in C and Matlab scripts (Mathworks, Natwick, MA) and executed on a 64-bit Windows desktop. A Perl-script on the scanner host transfers the acquired images after every scan (at an interval of 1.5 s) to the fMRI-BCI computer.

EPI sequences available from the manufacturer. In other words, no parameters needed be hard-coded into the sequence.

Functional images were acquired with a standard 12-channels head coil by the real-time EPI sequence and were then stored in a user-specified directory on the scanner host computer. Sixteen images pertaining to 16 slices (voxel size = $3.3 \times 3.3 \times 5.0 \text{ mm}^3$, slice gap = 1 mm) were accessible for preprocessing and further analysis at the end of each repetition time. The real-time sequence incorporated the following image acquisition parameters: TR = 1.5 s, matrix size = 64×64 , echo time TE = 30 ms, flip angle $\alpha = 70^\circ$, bandwidth = 1.954 kHz/pixel. For superposition of functional maps upon brain anatomy, during offline analysis, a high-resolution T1-weighted structural scan of the whole brain was collected from each subject (MPRAGE, matrix size = 256×256 , 160 partitions, 1 mm^3 isotropic voxels, TR = 2300 ms, TE = 3.93 ms, TI = 1100 ms, $\alpha = 8^\circ$). The Perl-script was a simple program that copied the functional image files in the ANALYZE format from a folder in the host computer to a Windows workstation that hosted the fMRI-BCI system.

The fMRI-BCI reads the images as they arrive, and performs online preprocessing, brain state classification and generation of visual feedback to the participant. Unlike LaConte and colleagues (2007), we did not alter the image reconstruction (IR) system of the scanner in implementing the fMRI-BCI. As the present study incorporates additional preprocessing steps such as online realignment, spatial smoothing and feature extraction, we anticipated that these procedures may introduce unpredictable delays that may affect the online reconstruction of the functional images in the image reconstruction system (IRS). To avoid this potential problem, we implemented online preprocessing, feature extraction and classification as a separate software module and executed it on the fMRI-BCI workstation. We implemented this subsystem in a

64-bit version of Matlab (version 7.1, Mathworks, Natwick, MA) by building around the C-language implementation of the core engine from SVMlight (Joachims, 1999). This approach offered us flexibility in repeated modify-and-test software life cycle of several intermediate implementations of preprocessing, brain masking, feature/voxel selection and feedback algorithms.

Real-time brain state classification on each participant is performed in several steps, as depicted in the Fig. 2. First, signal preprocessing is performed on each volume of brain images as they arrive from the scanner. In this step, real-time realignment for head motion correction and spatial smoothing are performed. In our implementation, the preprocessing was performed online during SVM training as well as testing (see Fig. 2). In the next step, a first-pass feature selection is done, whereby brain voxels are selected by choosing voxels above a user-specified intensity threshold. Subsequently, a second-pass thresholding is done, by which informative voxels are chosen from the first-pass brain mask by training an SVM classifier on the training data set. In the penultimate step, the classifier is retrained on the same data set but now with the features selected in step 3. Finally, the classifier is tested on new data collected from the same participant.

Real-time preprocessing

To correct for head motion artifacts, all brain scans were realigned to the first brain scan of the first run. The real-time implementation was adapted from the realignment algorithm (Ashburner et al., 2003; see Supplementary information and <http://www.fil.ion.ucl.ac.uk/spm/doc/books/hbf2/>) of SPM2. After realignment, the images were spatially smoothed in real-time with an 8-mm fixed width half maximum (FWHM) window.

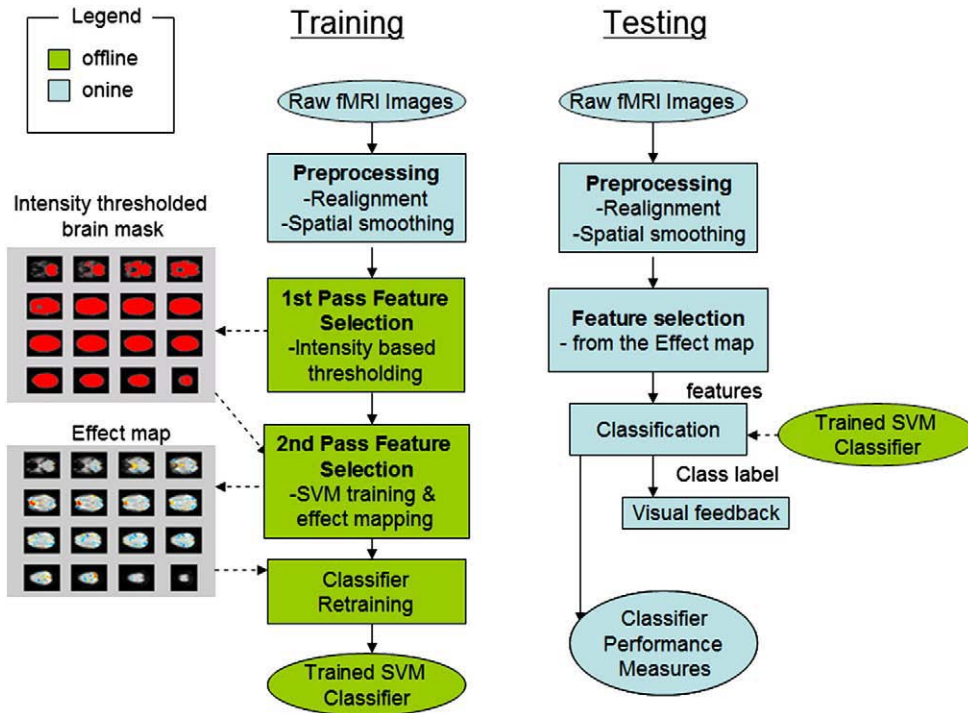


Fig. 2. Flow chart for fMRI signal preprocessing and classification. Brain state classification is performed in the following steps: (1) signal preprocessing for online realignment and spatial smoothing, (2) first-pass feature selection for selecting brain voxels by applying an intensity threshold resulting in a brain mask, (3) second-pass thresholding for selecting informative voxels from the first-pass brain mask by the method of Effect mapping resulting in the final brain mask, (3) classifier retraining based on the brain mask obtained in step 3, and (4) real-time classifier testing on new data using the second-pass brain mask.

Support vector classification

To classify an individual scan of fMRI data, all brain voxels or selected brain voxels from each repetition time (TR) can be composed of an input vector \mathbf{x}^i . SVM determines a scalar class label Y^i ($Y^i = \text{sgn}(y^i) = \text{sgn}(\mathbf{w}^T \mathbf{x}^i + b) = \pm 1, i = 1, \dots, N$, where N is the number of input vectors, T is the transpose of a vector, b is a constant value, $\text{sgn}(\cdot)$ is a signum function, $\text{sgn}(x) = +1, 0, -1$ if $x > 0, x = 0, x < 0$, respectively) from the input vector.

When the input vectors \mathbf{x}^i and the designed labels Y^i_L are taken from the training data set, the weight vector \mathbf{w} of SVM is obtained by minimizing objective function L of Eq. (5) with constraints Eqs. (6) and (7),

$$L = \frac{1}{2} \mathbf{w}^T \mathbf{w} + C \sum_{i=1}^N \xi^i, \quad (5)$$

$$\text{with } Y^i_L (\mathbf{w}^T \mathbf{x}^i + b) \geq 1 - \xi^i, \quad (6)$$

$$\text{and } \xi^i \geq 0 \quad (7)$$

where slack variable ξ^i is introduced to describe a non-separable case, i.e., data that cannot be separated without classification error, C denotes the weighting on the slack variable, i.e., the extent to which misclassification is allowed, and a value of $C=1$ was used in our implementation because of the following reasons. Firstly, model selection to determine the C value is hard to perform in the context of real-time classification due to limitations of time available for SVM training. What is important is that real-time, online classification should work robustly in the majority of participants and sessions. In many previous fMRI classification studies (Haynes et al., 2007; LaConte et al., 2005, 2007; Mourao-Miranda et al., 2005, 2007, 2006; Soon et al., 2008), $C=1$ was successfully used. Furthermore,

LaConte et al. (2005) showed that prediction accuracy does not vary much with the selection of C .

Feature selection and SVM training

The performance of a pattern classifier could be improved by a procedure called feature selection, where informative features from the data are extracted for input to the classifier. We performed feature selection in 2 steps: (1) a first-pass intensity thresholding and (2) a second-pass voxel selection using Effect mapping (Lee et al., 2010). For the first-pass intensity thresholding, an image of the brain, called intensity thresholded mask, was created by removing voxels below a certain BOLD intensity value in a mean image of all the brain scans. For the second-pass feature selection, training data from 2 runs were divided into 5 sequential subsets. That is, each subset consisted of 4.5 blocks (for example, in the case of binary classification: 2.5 blocks of condition happy, 2 blocks of condition disgust or 2.5 blocks of condition disgust, 2 blocks of condition happy) from a total number of 24 blocks. Z-normalization (z-value: $(x - \text{mean}(x)) / \text{standard deviation}(x)$, x : samples) was applied across all the time-course signals at each voxel to account for the variability of BOLD signals across runs. Z-normalization was performed to correct for the variability of the BOLD signal and to convert it to a signal of zero mean and unit variance. BOLD values from brain voxels of the subsets were used to create input vectors for training SVMs and subsequently to generate five Effect maps (please see Section 3.5 for a description of the Effect maps). Informative voxels were selected by applying a threshold on the mean of five Effect maps (Fig. 2). The online SVM classification software written in Matlab allows for specifying user-specified threshold. The threshold that was incorporated in this study was 0.5 for all participants and sessions. In our experience, this threshold provided a good balance between reduction of the size of the data input while maintaining good prediction accuracy. Finally, an SVM was trained from the BOLD values of the feature-selected voxels after Z-normalization across all the training samples.

Ideally, SVM output y^i should be centered about zero, so that when the output is greater than zero the brain state is assigned to one emotion (e.g., happy), and when it is less than zero it is assigned to the other emotion (e.g., disgust). However, due to participants' head movements and other systemic changes, a gradual drift in the classifier output can be expected (LaConte et al., 2007). To remove this bias during online classification, we subtracted the mean of the SVM outputs during the rest condition from each SVM output during the active condition (i.e., happy or disgust).

Effect mapping

Based on the parameters of the trained SVM model, we analyzed the fMRI data with a new measure called *effect value* (Lee et al., 2010) which estimates the effect of the BOLD activation at each voxel on the classifier output by first computing the *mutual information* between the voxel and the output, and then multiplying it with the weight value of the voxel as estimated by SVM. Mutual information (MI) is derived from relative entropy or the Kullback–Leibler divergence, which is defined as the amount of information that one random variable contains about another random variable (Cover and Thomas, 1991). Hence, the effect value (EV) E_k of a voxel k is defined as:

$$E_k = w_k I(x_k; y), \quad k = 1, \dots, M \quad (8)$$

where $I(x_k; y)$ is the mutual information between the voxel and the output, y is the SVM output after excluding the sign function, w_k and x_k are weight value and activation at voxel k , respectively. To reduce the variability of EVs from Eq. (8) dependent on data set, normalization of E_k from Eq. (9) was applied as follows:

$$|E_k = \text{sgn}(E_k) \log(1 + \text{std}(|E|)), k = 1, \dots, M \quad (9)$$

where $\text{sgn}(\cdot)$ is a signum function, and $\text{std}(|E|)$ is standard deviation across all the brain voxels. For ease of voxel selection based on the EVs, we have used Eq. (9) for generating the Effect maps at the contrast: happy vs. disgust in binary classification, and happy vs. disgusted, happy vs. sad, and disgusted vs. sad, in our multiclass classification, respectively. Such maps can be thresholded at a suitable level by visual inspection by the experimenter and the resulting map can be used as a brain mask. This brain mask is subsequently applied to functional images arriving at each time point (TR), to reduce the data dimension and to choose the most important voxels for classification.

Generation of visual feedback

Visual feedback was provided in the form of a graphical thermometer whose bars increase or decrease in number from the baseline value (represented by the middle red dashed line of the thermometer) proportional to the classifier output after each TR. The thermometer could be operated in two modes by selection in our Matlab toolbox: (1) feedback mode and (2) non-feedback mode. In the non-feedback mode, the thermometer was displayed without any bars for the whole duration of the experiment. In the feedback mode, the graphical thermometer is shown in an animated fashion by increasing or decreasing the number of bars. The classifier output after correction of the classifier drift was used to generate the visual feedback. Depending on the classification result, i.e., $+1$ or -1 , the thermometer bars were updated with respect to the previous TR in the positive or negative direction, respectively. The thermometer was designed to indicate not only the correctness of classification (in terms of the bar changing in the positive or negative direction) at the end of each TR, but also the cumulative performance of the past trials in terms of the height of the bars above or below the baseline. This way the participant receives feedback about his instantaneous

(defined by the length of the TR) brain state as well as his past performance.

Multiclass classification

The objective of the multiclass classification experiment was to test whether the online classifier is able to classify three discrete emotions, namely, happy, sad and disgust. The multiclass problem was formulated as a problem of finding the best classification among three trained binary classifiers (i.e., classifier 1: happy vs. disgust, classifier 2: happy vs. sad, and classifier 3: disgust vs. sad) based on the selected voxels from an Effect map of each combination as described above.

Multiclass classification was performed based on the framework of error correcting output code (ECOC) from 3 binary classifications (Mourao-Miranda et al., 2006). The ECOC method assigns a unique binary string of length n , called the “code word”. Subsequently, n binary classifiers are trained, one for each bit position of the binary string. During the classifier testing, new data are classified by evaluating each of the binary classes to arrive at an n -bit string, which is then compared to all the different code words. The test data is finally assigned to the code word that is the closest based on a distance measure. In our approach to the ECOC, each class has its own code \mathbf{m}_i as follows: $\mathbf{m}_1 = [1 \ 1 \ 0]$ for classifier 1 (happy vs. disgust), $\mathbf{m}_2 = [-1 \ 1 \ 0]$ for classifier 2 (happy vs. sad), $\mathbf{m}_3 = [0 \ -1 \ -1]$ for classifier 3 (disgust vs. sad). Let \mathbf{p} be a vector of predictions from the 3 classifiers. Then, the final decision was made by selecting the closest code from the prediction vector $r = \min_{\arg i} i = 1^3 \text{dist}(\mathbf{m}_i, \mathbf{p})$. We used the Euclidian distance measure in our implementation.

Experimental paradigm

Our experiment (Fig. 3) was divided into three parts: experiment 1 for investigating real-time binary classification, experiment 2 for further feasibility testing of multiclass prediction, and experiment 3 for assessing the effect of extended feedback training with a real-time classifier. Twelve healthy students from the Department of Medicine, University of Tuebingen, Germany, participated in experiment 1 and four more students participated in experiment 2. The participants were aged in the range of 22–26 years, with a mean age of 25 years. All participants signed a written informed consent, and the study was approved by the local institutional review board.

Experiment 1 (see Fig. 3) was envisioned in 4 succeeding stages, each stage prepared as a block-design protocol with alternating blocks of happy and disgust imagery of 24-s duration, interleaved with rest blocks of 4.5-s duration. In each run, there were 6 blocks of happy imagery, 6 blocks of disgust imagery and 12 blocks of rest. In total, 233 scans were collected in a run including the 5 initial scans that were not used in the analysis, due to magnetic equilibration effects that could potentially distort the data. The rest blocks were incorporated for the sole purpose of avoiding cognitive and emotional overload due to sudden changes of imagery. During classifier training (stage 1) containing 2 runs of the above protocol, there were 12 blocks of happy imagery, 12 blocks of disgust imagery and 24 blocks of rest. During each type of the emotional imagery block, an empty thermometer (without feedback) was shown at the centre of the screen with a letter beside it indicating the participant to perform happy ('H') or disgust ('D') mental imagery. The purpose of the empty thermometer during the SVM training stage was to maintain consistency in visual stimulation with the testing stage during which the thermometer could be updated for visual feedback of the brain state. Two runs of the stage 1 protocol were performed to collect sufficient amount of data for training the SVM classifier (stage 2).

Participants were instructed well in advance of the experiment to identify one or more emotional episodes from their personal lives for each type of emotion (e.g., happy, sad or disgust) that they were

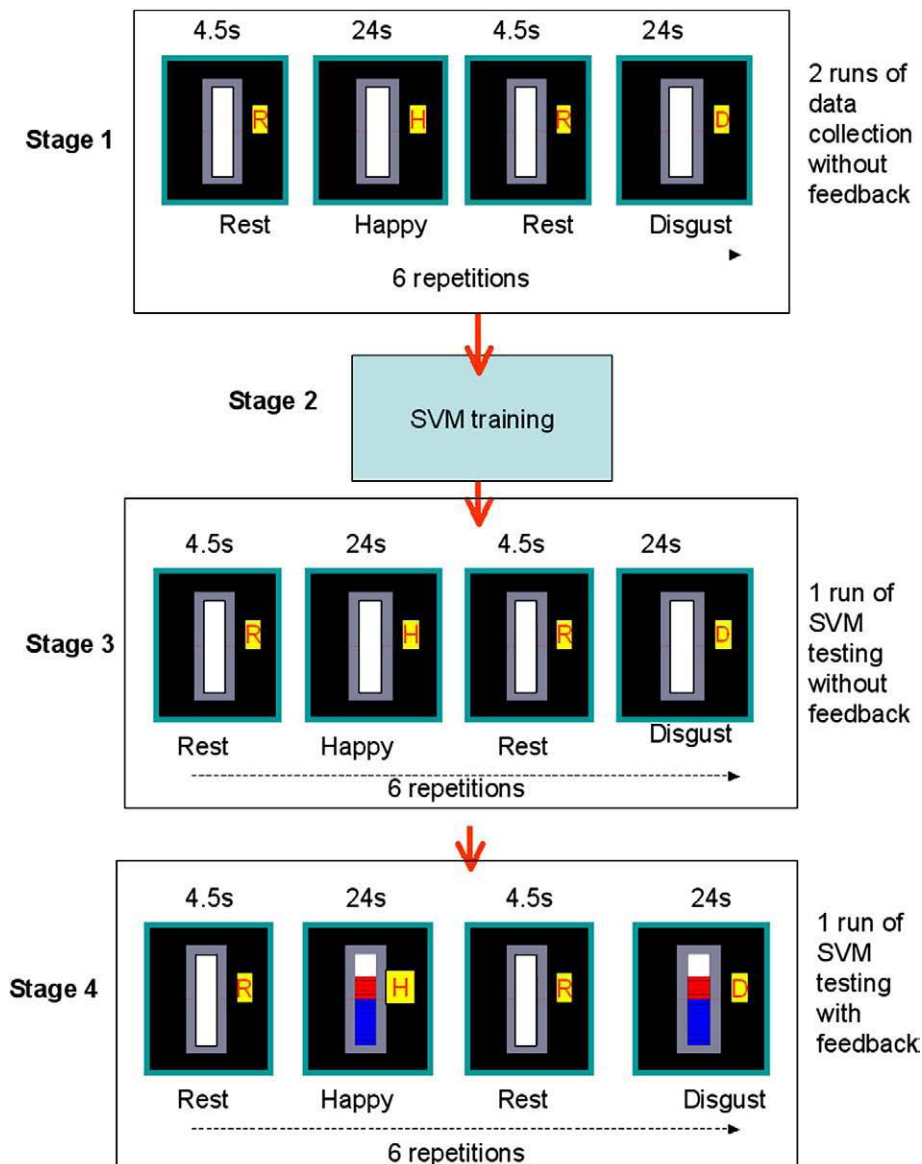


Fig. 3. Experimental protocol. Both binary (happy and disgust) and multiclass (happy, sad and disgust) classification experiments were performed in 4 succeeding stages, each stage prepared as a block-design protocol with alternating blocks of emotion imagery each of 24-s duration, interleaved with rest blocks of 4.5-s duration. During SVM training, an empty thermometer was shown at the centre of the screen with a letter beside it indicating the participant to perform happy ('H'), disgust ('D') or sad ('S') mental imagery. Two runs of the stage 1 protocol were performed to collect data for training the SVM classifier (stage 2). Stages 3 and 4 were performed to test the classifier without and with feedback, in that order.

required to recall during the experimental blocks. To maintain consistency between classifier training and testing conditions we showed participants pre-selected pictures as a reminder of the specific type of emotion imagery that they need to use in each session. Towards this end, participants were asked to identify one picture from the International Affective Picture System (IAPS) (Lang et al., 1997) that best epitomised each type of emotional episode. At the beginning of a training or testing run, participants were shown each selected picture for a duration of 30 s to remind and strengthen their emotional recall strategies. We could have alternately displayed text to describe and remind them of the specific imagery to use, but we found that pictures captured more succinctly the complexity of an emotional recall scenario. However, to avoid a potential problem that subjects might just recall the images rather than the emotion, we specifically instructed subjects to use the IAPS images as mere references to their imagery and to elaborate on the images to invoke the required discrete emotion (happy, disgust or sad emotion). We later confirmed the type and level of emotion imagery they performed with an interview and subjective rating (see Table 1 and Supplementary Table 2).

Stages 3 and 4 were meant to test the classifier without and with feedback, respectively. These two tests were performed to compare real-time classification performance in the absence and presence of feedback to evaluate the eventual application of real-time brain state decoding for BCI and clinical rehabilitation. At the end of each run, participants were asked, over the scanner audio system, to complete a self-assessment scale (1–9, 1 indicating poor performance and 9 indicating best performance) of their level of emotion regulation for each type of emotion. The purpose of this self-report was to perform an offline analysis of the correlation between subjective report and online classifier performance.

Four out of the 12 participants in experiment 1 underwent an additional run of the stage 4 protocol to test whether the classifier could robustly predict emotion states even when participants do not use the same emotion inducing strategies that were used during SVM training runs. To assess whether feedback training helps participants to improve emotion recall, we recruited two more participants and trained them on two-class (happy vs. disgust) classification for 3 sessions (in addition to 2 sessions for collection of training samples).

Table 1
PANAS scores, subjective ratings and strategies for binary and multiclass emotion imagery.

Subject	Mean positive affect sub-score of PANAS	Mean negative affect sub-score of PANAS	Subjective ratings of emotion imagery (scales 1–9, 1 = worst, 9 = best)*				Strategies for happy (H), disgust (D) and sad (S) emotions
			Training run 1	Training run 2	Test run (no feedback)	Test run (with feedback)	
<i>Binary classification with same strategies for training and testing</i>							
S1	30	24	7.5	7.5	6	6.5	H: romantic love D: toilet/feces
S2	41	21	6	6	6	5	H: family reunion D: eating insects
S3	26	12	8	7.5	7.5	7.5	H: good marks in exams D: blood in an open wound
S4	26	13	6	6.5	6	7	H: graduation day D: a dirty toilet
S5	27	26	6.5	5	4.5	4.5	H: riding a bike D: infected skin wound
S6	20	23	2.5	3	4	3.5	H: romantic love D: feces
S7	28	30	6	6	5.5	6.5	H: romantic love D: a dirty toilet
S8	28	18	9	6	6	7	H: baby D: a dirty toilet
S9	29	22	6.5	6	5	6	H: friends gathering D: a dirty toilet
S10	34	10	7	7	7	9	H: cats D: an used toilet
S11	23	16	7	8	9		H: friends gathering D: dirty and sick people
S12			4	8	9	9	H: falling in love D: spiders
<i>Binary classification with different strategies for training and testing</i>							
S9	28	13	7	6.5	6.5	5.5	Training runs: H: romantic love D: baby with feces Testing runs: H: family reunion D: dirty bathroom
S10	14	17	6	7	6	7.5	Training runs: H: romantic love D: spiders Testing runs: H: party with friends D: dirty bathroom
S11	39	10	8.5	8.5	8	9	Training runs: H: romantic love D: toilet Testing runs: H: playing the guitar D: rotten food
S12	43	11	8	8.5	8	6.5	H: getting the visa D: a dirty laboratory Testing runs: H: romantic love D: toilet
<i>Multiclass classification with same strategies for training and testing</i>							
S13	31	10	7.7	8.4	8	7.7	H: pets D: toilet S: braking up with boyfriend
S14	24	14	7.7	7.8	7.7	7.8	H: graduation day D: toilet S: death of a friend
S15	27	11	6.2	8	8	9	H: a boat trip D: bloody car accident S: braking up with boyfriend
S16	14	16	6.8	7.3	6.3	7.8	H: romantic love D: spiders S: view of blind beggars

*Values correspond to the mean between happy and disgust score in binary classification, or between happy, disgust and sad scores in the multiclass classification.

To adapt the classifier to the changes in the participant's brain activation with thermometer feedback and learned regulation, we retrained the classifier after each feedback session on a combined data set of the training data set (collected in the absence of feedback) and the feedback data set (collected in the presence of feedback).

Experiment 2 was designed to test real-time multiclass classification, both with and without feedback. Four more student participants (mean age = 25) were recruited for this experiment. Conceptually, experiment 2 was designed similar to experiment 1, except that additional blocks pertaining to a third emotion (sadness) was added

to train and test a three-class classifier. That is, each run consists of 4 blocks of each emotion with 12 blocks of rest condition. The block length is exactly same as the binary case. The total duration of each run of experiment 2 was also 233 scans (5.825 min).

Experiment 3 was conducted to answer the question: How can a pattern classifier provide contingent feedback in a self-regulation application where participants are normally expected to improve their ability with training to induce themselves into desired brain states? One way to approach this would be by adapting the classifier to participant's learning by incrementally retraining the classifier with new samples at the end of each session of feedback training. We have adapted this method in an extended experiment in two participants to incrementally enhance the classifier's ability to recognize changes in brain activity with self-regulation learning. We trained participants on 2-class (happy vs. disgust) classification, with 3 sessions of feedback training (in addition to 2 sessions for collection of training samples) in a day. We retrained the classifier after each feedback session by combining the previous training set with the new test data. Thermometer feedback was provided as described in section 3.6. Participants were instructed to perform emotion recall in order to increase the bars of the thermometer. They were told that thermometer bars going up meant that they were doing the task right and bars going down meant that they were probably not doing it right.

Offline whole brain and ROI analysis with classification

To assess the level of involvement of different brain regions, two additional types of SVM classifications were performed offline: (1) single-subject whole brain analysis with the E-map (Lee et al., 2010) (binary classification on 12 subjects and multiclass classification on 1 subject), and (2) ROI analysis (binary classification on 12 subjects). We expected that BOLD signals from brain regions known to be involved in emotion processing would show higher decoding accuracy for the discrete emotions under consideration (i.e., happy and disgust) than in other regions. Our ROI analysis is based on the expectation that classification accuracy will be higher in ROIs that have been previously implicated in emotion regulation and induction (Amodio and Frith, 2006; Ochsner et al., 2005, 2004; Phan et al., 2002). Towards this end, fMRI images for all participants were preprocessed with the following procedures: realignment, normalization to MNI space, and spatial smoothing with 8 mm Gaussian FWHM. First, single-subject whole brain classification was performed wherein BOLD signals from all brain voxels for each TR were collected into an input vector. For the ROI classification, 24 masks (see Supplementary Fig. 2) from brain regions previously reported to be involved in emotion imagery, recall and regulation (Amodio and Frith, 2006; Ochsner et al., 2005, 2004; Phan et al., 2002) were applied to the preprocessed data to extract BOLD signals as input to separate classifiers. Z-normalization was performed on the BOLD values at each voxel across the whole time course to correct for the variance of BOLD signals of different runs and different participants. The normalized BOLD values from each ROI of each TR were then collected into an input vector. Data from 2 training sessions and 1 testing session were used in both the single-subject whole brain and ROI classifications. Linear SVM, with the regularization parameter $C=1$, based on freely available SVM library SVMlight (Joachims, 1999) was used to perform both single-subject and ROI classifications. The single-subject whole brain analysis was performed based on the E-maps obtained after training a separate SVM model from each participant's data. The ROI classification performance was evaluated through 12-fold cross validation (CV). In each fold, the data of 11 participants were used to train the classifier, and then the data of 1 remaining participant were used to test the classifier.

Affect scores and subjective ratings of emotion recall

Before the beginning of fMRI data collection, we asked each participant to fill out the Positive Affect Negative Affect Schedule (PANAS), which is a

psychometric scale developed to measure the independent constructs of positive and negative affect both as states and traits (Watson et al., 1988). PANAS contains a list of 10 descriptors for positive scale: attentive, interested, alert, excited, enthusiastic, inspired, proud, determined, strong and active; and 10 descriptors for negative scale: distressed, upset-distressed; hostile, irritable-angry; scared, afraid-fearful; ashamed, guilty; nervous, and jittery. The PANAS has been found to exhibit the following characteristics: (1) a significant level of stability in every time frame; (2) no consistent sex differences; (3) inter-correlations and internal consistency reliabilities are all acceptably high (ranging from 0.86 to 0.90 for PA and 0.84–0.87 for NA); (4) the reliability of the scales is unaffected by the time instructions used; and (5) has high scale validity and high item validity.

At the end of each SVM training or testing run, participants were instructed to rate their degree of success in being able to recall each type of emotion (i.e., happiness, disgust or sadness) for each block of the run on a scale of 1–9, where 1 represented the worst and 9 represented the best regulation (see Table 1).

Results

PANAS scores and subjective reports of emotion imagery

Participants used several emotion recall scenarios and imagery strategies for inducing themselves into happy, disgust and sad (see Table 1).

A negative correlation was found between the mean of the negative subscale of the PANAS and participants' ratings of their performance in emotion imagery tasks of session 3 ($r_s = -.74, p < .01$) and 4 ($r_s = -.72, p < .01$). A similar trend was also found for session 2 ($r_s = -.56, p = .07$, two-tailed). No significant correlations were found between positive subscale of the PANAS and the participants' ratings of their performance across the sessions of emotion imagery task. Also, no significant correlations were found between the prediction accuracy either with the PANAS scores, or with participants' ratings of their performance in emotion imagery.

In addition, we collected subjective reports of the intensity of the emotion attained, attention to the imagery task and the consistency of the mental strategy within each session, in a 9-point scale (1–worst, 9–best). Due to lack of scanning time, these data could be collected only in a subset of participants: 4 participants in the binary classification and 4 participants in the multiclass classification (see Supplementary Table 2).

Real-time binary and multiclass classification

Fig. 4 shows the prediction accuracies of the online classifier for all participants for SVM testing runs, with and without feedback, for binary classification (see Figs. 4a and b) and for multiclass classification (see Fig. 4c). Fig. 4a shows the performance of the binary classifier when participants used the same type of emotion imagery both during SVM training and testing runs (e.g., a participant used imageries of family reunion for happy emotion and eating insects for disgust emotion, respectively). Fig. 4b shows the performance of the binary classifier when participants used different types of emotion imagery during SVM training and testing runs (e.g. a participant used imageries of romantic love for happy emotion and baby with feces for disgust emotion for SVM training; and family reunion for happy emotion and dirty bathroom for disgust emotion for SVM testing). The online classifier showed the following classification accuracies across participants, indicated by mean and standard deviation of correct classification: (1) for binary classification (chance accuracy = 50%) testing with same emotion imagery, without feedback ($92\% \pm 6\%$) and with feedback ($80\% \pm 13\%$) (see Fig. 4a); (2) for binary classification testing with different emotion imagery, without feedback ($80\% \pm 10\%$) and with feedback ($65\% \pm 18\%$) (see Fig. 4b); and (3) for multiclass classification (chance accuracy = 33%) testing without

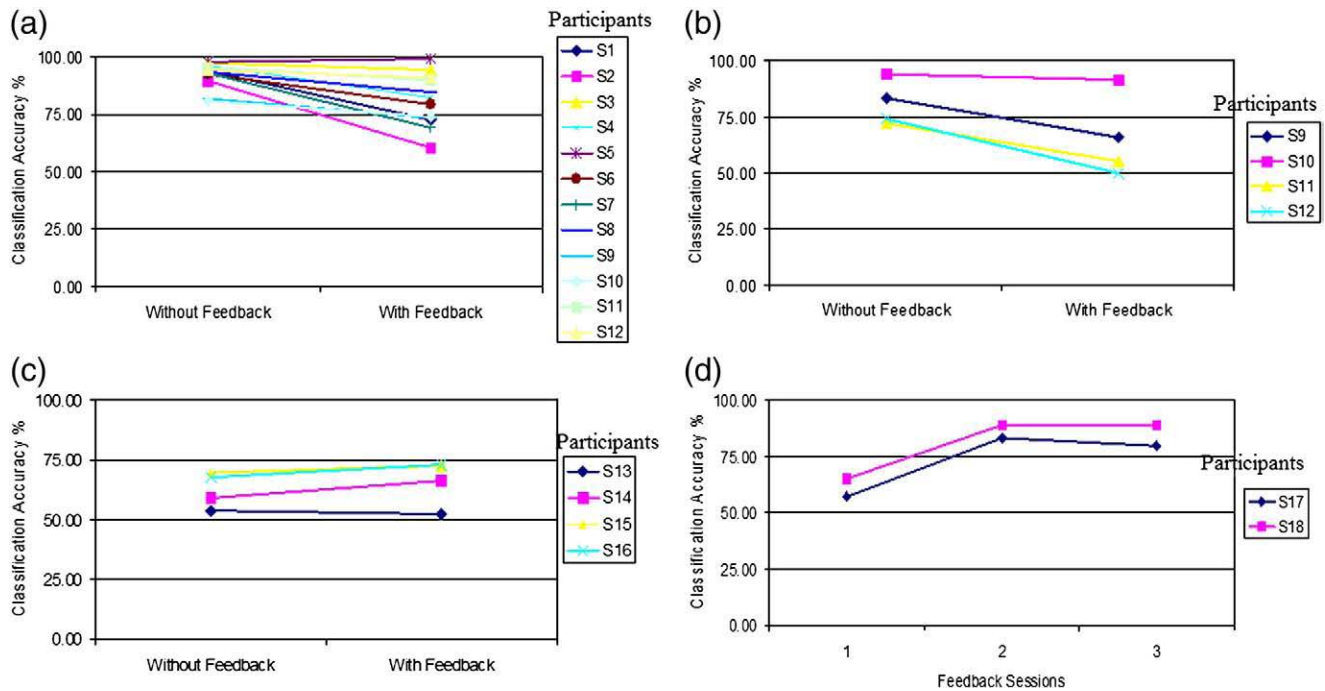


Fig. 4. Online SVM classification performance. (a) Binary classification (happy vs. disgust) accuracies (chance level = 50%) in 12 participants who used the same emotion imagery during both SVM training and testing sessions. (b) Binary classification accuracies in 4 participants who used different emotion imagery between SVM training and testing sessions. (c) Multiclass classification (happy vs. disgust vs. sad) accuracies (chance level = 33%) in 4 participants who used same emotion imagery in SVM training and testing sessions. (d) Classification accuracies (chance level = 50%) of extended feedback training on 2 participants on two-class (happy vs. disgust) classification.

feedback ($62\% \pm 14\%$) and with feedback ($60\% \pm 16\%$). Classification performance indicates the beneficial effects of online realignment, feature selection and optimized experimental procedures. Robust prediction accuracies were seen even when participants were intentionally instructed to use different strategies between SVM training and testing runs, indicating the ability of the classifier to generalize across varied emotion imagery strategies and memory recall scenarios.

Significantly higher accuracies of the SVM classifier were found in the sessions without feedback (median = 88.5) compared to the sessions with feedback (median = 76), $z = -2.02$, $p < .05$, $r = -.37$ (Wilcoxon signed-rank test). To assess the effect of extended feedback training on participants' performance, we trained 2 new participants for 3 sessions of emotion recall in the presence of feedback and incremental retraining of the classifier. Our results show that classification performance improves already in the second feedback session and classification accuracy continues to be maintained in the third session (see Fig. 4d). However, by the end of 5 training sessions, participants were unable to continue due to tiredness and hence feedback training had to be stopped. Nevertheless, our data suggests that feedback training with incrementally retrained classifier could enhance participant's performance.

SVM outputs for binary and multiclass classification

Supplementary Figs. 3a and b show the online SVM output for a participant before and after bias removal. Examination of the SVM outputs across different participants for multiclass classification showed robust classifier performance on a block-by-block basis within each testing session (see Supplementary Fig. 3c).

Effect maps and ROI classification performance

Extensive results of the single-subject whole brain classification and Effect maps generated thereof (in an offline analyses performed

after the real-time experiments to reveal the discriminating brain regions) have been documented in the Supplementary information (Supplementary Sections 2.2 and 2.3, including supplementary Table 3 (a-k), Supplementary Fig. 1 (a-k) and Table 4 (a-c)). The Supplementary Tables also provide, for each brain region, the effect value (EV) and the cluster size, which together indicate the degree of importance of the cluster in discriminating between the emotions. Much inter-subject variability was observed in brain activations, possibly due to the diverse emotion imagery employed by the participants (see Table 1 for list of emotion imagery). The following brain regions were also observed as commonly activated among the participants: orbital medial frontal cortex (oMFC), anterior rostral medial frontal cortex (arMFC) and posterior rostral medial frontal cortex (prMFC) (these anatomical segregation were based on Amodio and Frith (2006)); superior and lateral frontal cortices, anterior cingulate cortex (ACC), insula, superior temporal gyrus (STG), primary and association areas of the visual cortex, posterior cingulate cortex (PCC) and precuneus. In terms of Brodmann areas, the following areas were commonly observed: BA10, 11, 14 and 25 belonging to the oMFC, BA32 to arMFC, BA8, 9 to the prMFC; and BA17, 18 and 19 belonging to the visual cortex. Fig. 5 shows exemplary Effect maps depicting the 3 contrasts of (happy vs. disgust), (happy vs. sad) and (disgust vs. sad) emotions.

To better identify the most important brain regions for emotion imagery and regulation, we performed offline ROI classifications, based on fMRI signals extracted from 24 different brain masks (see Supplementary Fig. 2) previously implicated in emotion processing (Amodio and Frith, 2006; Ochsner et al., 2005, 2004; Phan et al., 2002). This analysis was conducted separately on 12 subjects who had performed happy and disgust imagery (see Fig. 3, Table 1). Our results (see Fig. 5d) reveal that high classification performance ($>75\%$) was obtained in the following brain regions: middle frontal gyrus (MiFG), superior frontal gyrus (SFG), ventrolateral prefrontal cortex (VLPFC), STG and precuneus. Other ROIs that showed high prediction rates included dorsolateral prefrontal cortex (DLPFC), ACC and the insula.

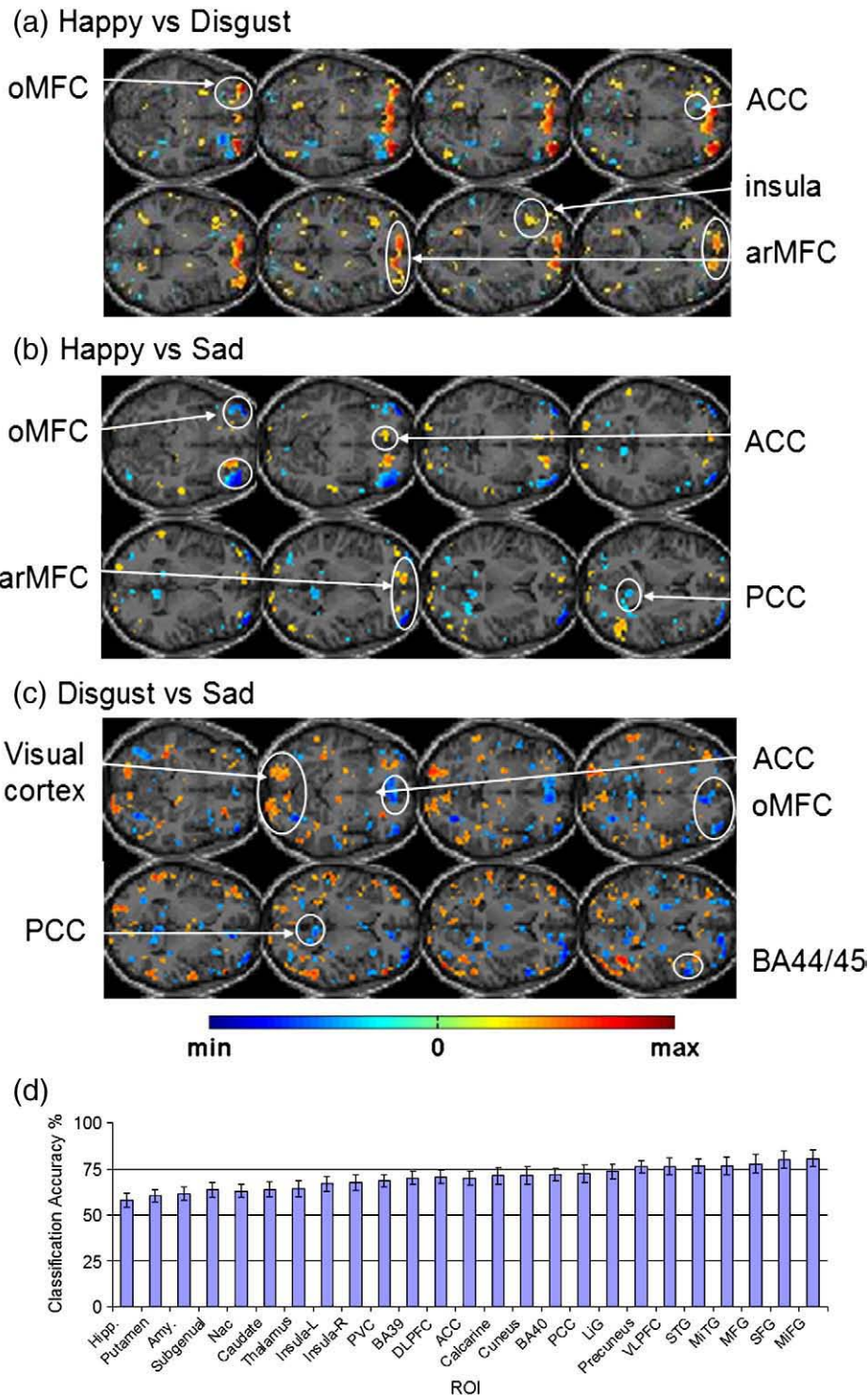


Fig. 5. Effect maps generated from single-subject whole brain SVM classification showing discriminating voxels for: (a) happy vs. disgust classification (for participant S1), (b) happy vs. sad classification and (c) disgust vs. sad classification (for participant S13). Second level threshold applied was 0.5, slices shown are in the range, $Z(\text{MNI coordinates}) = -11$ to $+10$ in steps of 3 units. Brain regions: oMFC—orbital medial frontal cortex, arMFC—anterior rostral MFC (based on Amodio and Frith, 2006); OFC—orbitofrontal cortex, ACC—anterior cingulate cortex, PCC—posterior cingulate cortex. (d) Prediction accuracy of 24 regions of interest (ROIs) in a single-subject ROI classification from 12 participants performing happy and disgust motor imagery. ROIs: Subgenual—subgenual cingulate (BA25), Nac—nucleus accumbens, PVC—primary visual cortex, DLPFC—dorsolateral prefrontal cortex, LiG—lingual gyrus, VLPFC—ventrolateral prefrontal cortex, STG—superior temporal gyrus, MiTG—middle temporal gyrus, MFG—medial frontal gyrus, SFG—superior frontal gyrus, MiFG—middle frontal gyrus.

Discussion

The present implementation of brain state decoding and fMRI-BCI is based on real-time SVM classification of fMRI signals after online head motion correction, spatial smoothing and feature selection based

on our new approach called Effect mapping (EM) (Lee et al., 2010). EM is a method of generating functional activations by multiplying two types of information at each voxel of the brain: (1) SVM trained weight value at the voxel, and (2) the mutual information (MI) between the BOLD value at the voxel and the SVM output. Activation

maps so derived represent a hybrid of univariate and multivariate functional mapping, resulting at once in spatially distributed activations as well as in large clusters centered about brain regions presumably important for discriminating between the brain states. In the present study, EM was used as the basis for feature selection to reduce the multidimensionality of fMRI data, and to choose the most discriminating clusters of voxels for classification. Our results show that the application of feature selection results in consistently good predictions of emotions for binary classification as well as for multiclass classification.

Subjective reports collected from participants ascertained that participants performed emotion imagery and that the online classifier decoded emotions and not arbitrary states of the brain. This study also shows a relationship between the participants' affect scores as measured by PANAS and the subjective ratings of their performance in the emotion imagery task, indicating that participants who report greater negative affect rate themselves relatively lower on the imagery tasks.

One may expect that higher subjective reports of success in recalling emotions should correlate with higher prediction accuracies. However, one should note that pattern classification works well and shows higher accuracy based on how similar the samples in the test set are in comparison to the training set. The greater the similarity, the greater is the accuracy. However, if the test set has different brain signals generated due to higher success in recall than in the training set, then the similarity of patterns is reduced and so is the classification accuracy. This means that improved success in recalling emotions may not result in greater decoding accuracy. This is in contrast to the situation with single ROI real-time feedback experiments where improved success in recall may correlate with higher % BOLD in the target region of interest.

This may then lead one to ask the question: How then can a pattern classifier provide contingent feedback in a self-regulation application where participants, including patients, are normally expected to improve their ability (with more training) to induce themselves into desired brain states? One way to approach this would be by adapting the classifier to participant's learning by incrementally retraining the classifier with new samples at the end of each session of feedback training. Another way could be to carry out feedback training of patients or beginners using a brain state classifier which has been previously trained on data from experts or individuals who are already good at self-regulation. In this approach, the classifier would show improved performance as the trainees achieve brain states similar to the experts. In such a classifier, one can expect improved prediction accuracy with greater success in recall or regulation. In any method, the experimenter has the prerogative to ascertain whether the participant is performing self-regulation as desired and whether the relevant brain regions are activated by scrutinizing the multivariate functional activations.

In our experiments, the performance of the classifier in the presence of feedback was lower than in its absence although a few participants performed better in the presence of feedback than in its absence. This lower accuracy of the classifier when the feedback was introduced could be due to as yet unlearned ability of the participants to perform consistent emotion imagery while simultaneously monitoring the thermometer. These data are in line with self-reports by participants after scanning sessions, about the difficulties they had in performing emotion regulation when feedback was introduced for the first time. It has been known that the introduction of feedback initially creates a task conflict between attention to the imagery task and the visual feedback (Rota et al., 2008), which could be overcome with longer training (Caria et al., 2007; deCharms, 2008; Rota et al., 2008). A second reason for the reduction in classification performance could be the extra activation of the visual cortex due to the graphical animation of the thermometer that was previously absent during the classifier training. This additional BOLD activation could act as noise or confound to the classifier input,

reducing its prediction accuracy. To overcome the above limitations and to assess the effect of extended feedback training on participants' performance, we recruited two more participants and trained them on two-class (happy vs. disgust) classification, with 3 sessions of feedback training by incorporating an incremental method of classifier retraining. Our results show that participants did learn to improve their performance in the presence of feedback. An elaborate study on the effects of feedback training is beyond the scope of this work. Future studies could be dedicated to the systematic evaluation of the effect of feedback training on self-regulation performance, behaviour and changes in relevant neural network.

We performed additional offline pattern classification and multivariate statistical mapping based on the EM (Lee et al., 2010) approach to identify brain regions for discriminating emotion under consideration. Our results are in line with previous findings. Firstly, the most observed involvement of the medial frontal cortex in our clusters of activation and Effect maps is of interest. A theoretical review of the involvement of medial frontal cortex (MFC) by Amodio and Frith (2006) delineates the functional roles of the 3 regions of MFC, namely, arMFC, prMFC and oMFC. The authors associated the activation of the arMFC (the most anterior part of the frontal cortex including BA10, BA32 and rostral ACC) with the monitoring of one's own emotional state, such as rating one's emotions in response to pictures of varying valence. The more posterior region, the rostral MFC (prMFC), including BA8, BA9 and dorsal ACC, is activated by cognitive tasks such as action monitoring and attention. The orbital MFC (oMFC), including BA10, BA14 and BA25, was linked to the monitoring of task outcomes associated with punishments or rewards. Activations of the prefrontal cortex, ACC and insula found in our analyses are in line with an influential meta-analysis by (Phan and colleagues (2002) which states that the MFC has a general role in emotional processing and that induction by emotion imagery recruits the anterior cingulate and insula. Our offline analyses serve to confirm that participants performed emotion imagery and that the SVM results are reliable. Additionally, to our knowledge, our study provides the first objective comparison, through pattern classification, of the degree of involvement of different brain regions (see Fig. 5 and Supplementary Table 4) in emotion imagery and regulation.

In our real-time implementation of the SVM classifier, we programmed the online SVM classifier and the BCI system in a combination of C programs and Matlab scripts on a dedicated workstation. This design decision was beneficial in the following ways: (1) it allowed us greater flexibility in the repeated software cycle of source code modification and testing, (2) it helped us avoid interfering with the normal operation of the scanner and (3) it resulted in faster SVM training and online SVM online classification. This modular approach to software design has an added advantage in that when future versions of the scanner operating software (e.g., Syngo version VB15) are implemented, it is no longer necessary for the user to modify, recompile and relink the SVM classification source code.

Our decision to use was SVM based on recent developments in the pattern classification of neuroimaging data (LaConte et al., 2003, 2005; Martinez-Ramon et al., 2006; Shaw et al., 2003; Strother et al., 2004). SVM has been shown to have certain advantages in the classification of fMRI signals in comparison to other methods such as linear discriminant analysis (LaConte et al., 2005) and multilayer neural network. SVM is less sensitive to preprocessing (LaConte et al., 2005), better capable of handling large data sizes (Haynes and Rees, 2006), techniques developed to perform multiclass classification and most importantly produces unique optimal solutions (Collobert et al., 2006). Although, SVM model training is computationally more intensive, current availability of faster yet cheaper processors more than compensate for this drawback (LaConte et al., 2007).

Online multiclass brain state decoding has potential applications in lie detection (critical lies, non-critical lies and truths), BCIs (detecting

multiple movement states and thoughts) for communication and control, to name a few. The majority of the present implementations of brain state decoding including real-time classification, however, have two main drawbacks. Firstly, the present methods work based on spatial pattern of brain activity alone as input in discriminating different states of the brain and ignore the temporal pattern or time evolution of brain activity. Considering that brain states may differ in their temporal signature, in terms of time evolution of activity in one region and also in terms of its dynamic interaction with other regions of the brain, spatial classification of brain activity alone is largely limiting. For example, emotion regulation involves the dynamic, temporal evolution of BOLD activity connecting different parts of the limbic region that includes insula, amygdala, anterior cingulate cortex and medial prefrontal cortex. Hence, a pattern classification system that uses both spatial and temporal information but does so in a computationally efficient way so that it could be used in real-time would be of great practical importance and should be the focus of future research.

Another limitation of our present work and possible topic of future research is the lack of an implementation for real-time, subject-independent classification of brain states. Existing methods are optimized for each participant because SVM training is carried out on subject-specific data. However, these methods have two major disadvantages: (1) one needs to collect initial data for classifier training that consumes time and participant's attention and involvement, and (2) this approach is not suitable for certain applications such as real-time lie detection for forensics and security on persons from whom data for SVM training might not be available. An adaptive classifier needs to be developed similar to those previously developed in the pattern recognition field for subject-independent speech and character recognition. Such a subject-independent brain state classifier could then be applied to new subject data without prior classifier training, and in addition, could be adapted to the idiosyncrasies of every individual's brain size, shape and activation patterns. An essential technical improvisation to be achieved in this regard is the real-time coregistration and normalization of functional images to a standard brain so that inter-subject variations in brain size, shape and activations could be overcome. We anticipate that a subject-independent classifier could find use in clinical rehabilitation, where patients with brain abnormalities pertaining to motor, cognitive or emotion processing could be retrained to achieve normal level of functioning by providing feedback from a real-time pattern classifier that is trained on healthy subjects. By repeated operant training with contingent reward from the classifier, patients could perhaps learn to mimic brain activation of healthy individuals in order to ameliorate their problem.

Acknowledgment

This work was supported by grants from the SFB 437 "Kriegserfahrungen", DFG BI 195/59-1 and DFG BI 195/56-1 of the Deutsche Forschungsgemeinschaft (DFG) and DFG-grant KU 1453/3-1. Author S. L. is grateful to DAAD (German Academic Exchange Service) for supporting this research. Author S.R. is thankful to the Faculty of Medicine, Pontificia Universidad Católica de Chile, and Conycit for supporting this research.

Appendix A. Supplementary data

Supplementary data to this article can be found online at [doi:10.1016/j.neuroimage.2010.08.007](https://doi.org/10.1016/j.neuroimage.2010.08.007).

References

Amodio, D.M., Frith, C.D., 2006. Meeting of minds: the medial frontal cortex and social cognition. *Nat. Rev. Neurosci.* 7, 268–277.

- Ashburner, J., Friston, K., Penny, W., 2003. *Human Brain Function*, 2nd ed. .
- Caria, A., Sitaram, R., Veit, R., Gaber, T., Kübler, A., Birbaumer, N., 2006. Can we learn to increase our emotional involvement? Real-Time fMRI of Anterior Cingulate Cortex During Emotional Processing. Organization for Human Brain Mapping, Florence, Italy.
- Caria, A., Veit, R., Sitaram, R., Lotze, M., Weiskopf, N., Grodd, W., Birbaumer, N., 2007. Regulation of anterior insular cortex activity using real-time fMRI. *Neuroimage* 35, 1238–1246.
- Collobert, R., Sinz, F., Weston, J., Bottou, L., 2006. Trading convexity for scalability. Proceedings of the 23rd international conference on Machine learning. ACM New York, NY, USA.
- Cover, T.M., Thomas, J.A., 1991. *Elements of information theory*. Wiley, New York.
- Davatzikos, C., Ruparel, K., Fan, Y., Shen, D.G., Acharyya, M., Loughhead, J.W., Gur, R.C., Langleben, D.D., 2005. Classifying spatial patterns of brain activity with machine learning methods: application to lie detection. *Neuroimage* 28, 663–668.
- deCharms, R.C., 2008. Applications of real-time fMRI. *Nat. Rev. Neurosci.* 9, 720–729.
- deCharms, R.C., Christoff, K., Glover, G.H., Pauly, J.M., Whitfield, S., Gabrieli, J.D., 2004. Learned regulation of spatially localized brain activation using real-time fMRI. *Neuroimage* 21, 436–443.
- deCharms, R.C., Maeda, F., Glover, G.H., Ludlow, D., Pauly, J.M., Soneji, D., Gabrieli, J.D., Mackey, S.C., 2005. Control over brain activation and pain learned by using real-time functional MRI. *Proc. Natl. Acad. Sci. U. S. A.* 102, 18626–18631 (Epub 12005 Dec 18613).
- Haynes, J.D., Rees, G., 2006. Decoding mental states from brain activity in humans. *Nat. Rev. Neurosci.* 7, 523–534.
- Haynes, J.D., Sakai, K., Rees, G., Gilbert, S., Frith, C., Passingham, R.E., 2007. Reading hidden intentions in the human brain. *Curr. Biol.* 17, 323–328.
- Joachims, T., 1999. Making large-scale SVM learning practical. In: Schölkopf, B., Burges, C., Smola, A. (Eds.), *Advances in Kernel Methods—Support Vector Learning*. MIT-Press.
- Kamitani, Y., Tong, F., 2005. Decoding the visual and subjective contents of the human brain. *Nat. Neurosci.* 8, 679–685.
- LaConte, S., Anderson, J., Muley, S., Ashe, J., Frutiger, S., Rehm, K., Hansen, L.K., 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., Strother, S., Cherkassky, V., Anderson, J., Hu, X., 2005. Support vector machines for temporal classification of block design fMRI data. *Neuroimage* 26, 317–329.
- LaConte, S.M., Peltier, S.J., Hu, X.P., 2007. Real-time fMRI using brain-state classification. *Hum. Brain Mapp.* 28, 1033–1044.
- Lang, P.J., Bradley, M.M., Cuthbert, B.N., 1997. *International Affective Picture System (IAPS): Technical Manual and Affective Ratings*. The Center for Research in Psychophysiology, University of Florida, Gainesville.
- Lee, S., Halder, S., Kubler, A., Birbaumer, N., Sitaram, R., 2010. Effective functional mapping of fMRI data with support-vector machines. *Hum. Brain Mapp.* (Epub ahead of print).
- Martinez-Ramon, M., Koltchinskii, V., Heileman, G.L., Posse, S., 2006. fMRI pattern classification using neuroanatomically constrained boosting. *Neuroimage* 7, 7.
- Mourao-Miranda, J., Bokde, A.L., 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 (Epub 2005 Nov 2004).
- Mourao-Miranda, J., Friston, K.J., Brammer, M., 2007. Dynamic discrimination analysis: a spatial-temporal SVM. *Neuroimage* 36, 88–99.
- Mourao-Miranda, J., Reynaud, E., McGlone, F., Calvert, G., Brammer, M., 2006. The impact of temporal compression and space selection on SVM analysis of single-subject and multi-subject fMRI data. *Neuroimage* 33, 1055–1065.
- Ochsner, K.N., Beer, J.S., Robertson, E.R., Cooper, J.C., Gabrieli, J.D., Kihlstrom, J.F., D'Esposito, M., 2005. The neural correlates of direct and reflected self-knowledge. *Neuroimage* 28, 797–814.
- Ochsner, K.N., Ray, R.D., Cooper, J.C., Robertson, E.R., Chopra, S., Gabrieli, J.D., Gross, J.J., 2004. For better or for worse: neural systems supporting the cognitive down- and up-regulation of negative emotion. *Neuroimage* 23, 483–499.
- Pessoa, L., Padmala, S., 2005. Quantitative prediction of perceptual decisions during near-threshold fear detection. *Proc. Natl. Acad. Sci. U. S. A.* 102, 5612–5617.
- Phan, K.L., Wager, T., Taylor, S.F., Liberzon, I., 2002. Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. *Neuroimage* 16, 331–348.
- Polyn, S.M., Natu, V.S., Cohen, J.D., Norman, K.A., 2005. Category-specific cortical activity precedes retrieval during memory search. *Science* 310, 1963–1966.
- Posse, S., Fitzgerald, D., Gao, K., Habel, U., Rosenberg, D., Moore, G.J., Schneider, F., 2003. Real-time fMRI of temporolimbic regions detects amygdala activation during single-trial self-induced sadness. *Neuroimage* 18, 760–768.
- Rota, G., Sitaram, R., Veit, R., Erb, M., Weiskopf, N., Dogil, G., Birbaumer, N., 2008. Self-regulation of regional cortical activity using real-time fMRI: the right inferior frontal gyrus and linguistic processing. *Hum. Brain Mapp.* 30 (5), 1605–1614.
- Rota, G., Sitaram, R., Veit, R., Weiskopf, N., Birbaumer, N., Dogil, G., 2006. fMRI-neurofeedback for operant conditioning and neural plasticity investigation: a study on the physiological self-induced regulation of the BA 45. Proceedings of the Cognitive Neuroscience conference, San Francisco.
- Ruiz, S., Sitaram, R., Lee, S., Caria, A., Soekadar, S., Veit, R., Birbaumer, N., 2008. Learned control of insular activity using fMRI brain computer interface in schizophrenia. 1st Schizophrenia International Research Society Conference, Venice, Italy.
- Shaw, M.E., Strother, S.C., Gavrilescu, M., Podzbenko, 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.
- Sitaram, R., Caria, A., Veit, R., Gaber, T., Kuebler, A., Birbaumer, N., 2005. Real-time fMRI based brain-computer Interface enhanced by interactive virtual worlds. 45th Annual Meeting Society for Psychophysiological Research, Lisbon, Portugal.
- Soon, C.S., Brass, M., Heinze, H.J., Haynes, J.D., 2008. Unconscious determinants of free decisions in the human brain. *Nat. Neurosci.* 11, 543–545.
- Strother, S., La Conte, S., Kai Hansen, L., Anderson, J., Zhang, J., Pulpura, S., Rottenberg, D., 2004. Optimizing the fMRI data-processing pipeline using prediction and reproducibility performance metrics: I. A preliminary group analysis. *Neuroimage* 23 (Suppl 1), S196–207.
- Veit, R., Lotze, M., Caria, A., Gaber, T., Sitaram, R., Birbaumer, N., 2006. Real-time fMRI of human anterior insula during emotional processing. *Human Brain Mapping* (Florence, Italy).
- Watson, D., Clark, L.A., Tellegen, A., 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Pers. Soc. Psychol.* 54, 1063–1070.
- Weiskopf, N., Scharnowski, F., Veit, R., Goebel, R., Birbaumer, N., Mathiak, K., 2004. Self-regulation of local brain activity using real-time functional magnetic resonance imaging (fMRI). *J. Physiol. Paris* 98, 357–373.
- 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.
- Yoo, S.S., Fairney, T., Chen, N.K., Choo, S.E., Panych, L.P., Park, H., Lee, S.Y., Jolesz, F.A., 2004. Brain-computer interface using fMRI: spatial navigation by thoughts. *Neuroreport* 15, 1591–1595.
- Yoo, S.S., Jolesz, F.A., 2002. Functional MRI for neurofeedback: feasibility study on a hand motor task. *Neuroreport* 13, 1377–1381.

6. Acknowledgements

First, I would like to appreciate Prof. Dr. Niels Birbaumer and Dr. Ranganatha Sitaram for providing a good opportunity to carry out my doctoral research and for supporting me through my doctoral work. I would like to thank Prof. Dr. Christoph Braun, Prof. Dr. Boris Kotchoubey, Dr. Ute Strehl, and Prof. Dr. Matthias Bethge for reviewing my dissertation and good comments. I would also like to thank Sergio Ruiz, Dr. Ralf Veit, Dr. Andrea Caria, and Sebastian Halder who collaborated with me on different projects and helped me. Many thanks to Prof. Dr. Horst Herbert and Dr. Katja Deiss from Graduate school of Neural & Behavioral Sciences, International Max Planck Research School for guiding all the administration procedures, and also DAAD (German Academic Exchange Service) which provided all the financial supports for my doctoral study of 3 years. Also I really appreciate many members of Niels's lab (Markus, Emanuele, Guilia, Woosang, Daniele, Ander, Balint, Adrian, Tamara, Linda, Ernesto, Mohit, Angela, Lydia, and all other members) for providing many nice personal experiences and good friendships. Finally, I would like to mention my girl friend (Boram Kim; currently my wife), my parents, and my sister. Without their supports and encouragement, I could not have successfully completed my doctoral studies. I am really grateful to them.