

# Control of a humanoid robot by a noninvasive brain–computer interface in humans

Christian J Bell, Pradeep Shenoy, Rawichote Chalodhorn and Rajesh P N Rao

Department of Computer Science and Engineering, University of Washington, Seattle, WA 98195, USA

E-mail: [pshenoy@cs.washington.edu](mailto:pshenoy@cs.washington.edu)

Received 25 March 2008

Accepted for publication 9 April 2008

Published 15 May 2008

Online at [stacks.iop.org/JNE/5/214](http://stacks.iop.org/JNE/5/214)

## Abstract

We describe a brain–computer interface for controlling a humanoid robot directly using brain signals obtained non-invasively from the scalp through electroencephalography (EEG). EEG has previously been used for tasks such as controlling a cursor and spelling a word, but it has been regarded as an unlikely candidate for more complex forms of control owing to its low signal-to-noise ratio. Here we show that by leveraging advances in robotics, an interface based on EEG can be used to command a partially autonomous humanoid robot to perform complex tasks such as walking to specific locations and picking up desired objects. Visual feedback from the robot's cameras allows the user to select arbitrary objects in the environment for pick-up and transport to chosen locations. Results from a study involving nine users indicate that a command for the robot can be selected from four possible choices in 5 s with 95% accuracy. Our results demonstrate that an EEG-based brain–computer interface can be used for sophisticated robotic interaction with the environment, involving not only navigation as in previous applications but also manipulation and transport of objects.

 This article features online multimedia enhancements

(Some figures in this article are in colour only in the electronic version)

Advances in neuroscience and computer technology have made possible a number of recent demonstrations of direct brain control of devices such as a cursor on a computer screen [1–5] and various prosthetic devices [6–8]. Such brain–computer interfaces (BCIs) could potentially lead to sophisticated neural prosthetics and other assistive devices for paralyzed and disabled patients. Some of the more complex demonstrations of control, such as the control of a prosthetic limb, have relied on invasive techniques for recording brain signals [6–8] (although see also [9, 10]). Non-invasive techniques, such as electroencephalography (EEG) recorded from the scalp, have been used in interfaces for cursor control [1–3] and spelling [11, 12] but the low bandwidth offered by such non-invasive signals (20–30 bits min<sup>-1</sup>) makes their use in more complex systems difficult. An approach to overcome this is to incorporate increasing autonomy in the agent that

executes BCI commands, for example a wheelchair with obstacle avoidance [13–15]. Extending this line of work, we demonstrate here that we can leverage advances in robotics and machine learning and, in particular, make use of a sophisticated humanoid robot which only requires high-level commands from the user. The robot autonomously executes those commands without requiring tedious moment-by-moment supervision. By using a dynamic image-based BCI to select between alternatives, our system can seamlessly incorporate newly discovered objects and interaction affordances in the environment. This frees the user from having to exercise control at a very low level while allowing non-invasive signals such as EEG to be used as low-bandwidth control signals. Such an approach is consistent with a cognitive approach to neural prosthetics [16].

In our interface, the subject’s command to the robot is determined based on a visually evoked EEG response known as P3 (or P300) [11, 17, 18] which is produced when an object that the user is attending to suddenly changes (e.g., flashes). Similar techniques have been used previously in speller paradigms [18, 19], and in control of a robotic arm [9] or a wheelchair [14, 15]. In our case, the P3 is used to discern which object the robot should pick up and which location the robot should bring the object to. The robot transmits images of objects discovered by its cameras back to the user for selection. The user is instructed to attend to the image of the object of their choice, while the border around each image is flashed in a random order. Machine learning techniques are used to classify the subject’s response to a particular image flash as containing a P3 or not, thereby allowing us to infer the user’s choice of object. A similar procedure is used to select a destination location. We present results from a study involving nine human subjects that explores the effects of varying the amount of time needed for decoding brain signals, the number of choices available to the user and the customization of the interface to a particular user. Our results show that a command for the robot can be selected from four possible choices in 5 s with 95% accuracy.

## 1. Materials and methods

### 1.1. Human subjects

Nine subjects (eight male and one female) were recruited from the student population at the University of Washington. The subjects had no known neurological conditions and had not participated in prior BCI experiments. The study was approved by the University of Washington Human Subjects Division and each user gave informed consent. Subjects received a small monetary compensation for their participation in the experiment.

### 1.2. User study protocol

The accuracy of the P3-based interface in decoding the user’s intent from brain signals was tested in a user study without the robot in the loop. These experiments consisted of four sessions: (1) two sessions where there were four (randomly selected) images on the screen in a  $2 \times 2$  layout, (2) one session with six images in a  $2 \times 3$  layout and (3) one session with six images in a  $3 \times 2$  layout. The ordering of these sessions was the same for all users. Since our goal was to design an interface that could be dynamically resized and reordered, we used these sessions to explore whether changes in the number of images or their layout in a grid would affect the performance of the interface. Each session lasted up to 12 min and consisted of a number of trials where one image was pre-designated as the target and the subject was instructed to focus on that image. The borders around the images were then flashed one at a time in a random order at intervals of 0.25 s, with ten flashes per image. For the four-image sessions, each image position was the target for ten trials, and for the six-image sessions, each image position was used in five trials as the target. Data from 32 EEG electrodes were recorded at a sampling rate

of 2048 Hz, and stored to disk along with stimulus timing information. The collected data were used for training and testing the classifier under different conditions as described below.

### 1.3. Classification methods

We recorded signals at 2 kHz from a 32-channel layout according to the 10–20 layout [20], using a Biosemi ActiveTwo system (Biosemi, Amsterdam, The Netherlands). The EEG signals were bandpass filtered (0.5–30 Hz) and downsampled to 100 Hz. We used a 0.5 s long data window following each flash, labeled as *target/nontarget*, to train a binary classifier that would predict whether the response to a given flash contained a P300 or not.

While testing in a multiple-image scenario, each image flash was classified using this P300 detector, and the classifier output was added to a running tally of scores for the images. The image with the highest classifier output from the binary classifier was designated as the final result of the selection process.

In addition, we used a recently proposed spatial projection algorithm [21] that is designed for event-related EEG responses. Briefly, the algorithm considers a multichannel time-series response to an event and projects all the channels to form a single time series that is *maximally discriminative*. The criterion used for optimization is as follows. Let  $E_i$  represent an event-related response to event  $i$ , in the form of a  $C \times T$  matrix ( $C$  is the number of channels and  $T$  is the number of time points). Let  $f$  be a projection filter and let  $x_i = f^T E_i$  be the time series of ‘features’ formed by linearly weighting and combining all channels. We can compute the *within-class* and *between-class* scatter matrices ( $S_w$  and  $S_b$  respectively) over the set of all  $x_i$  and maximize the Jacobian  $J$  [21]:

$$J = \frac{\text{tr}(S_b)}{\text{tr}(S_w)}.$$

It can be shown that this is equivalent to maximizing the following quantity:

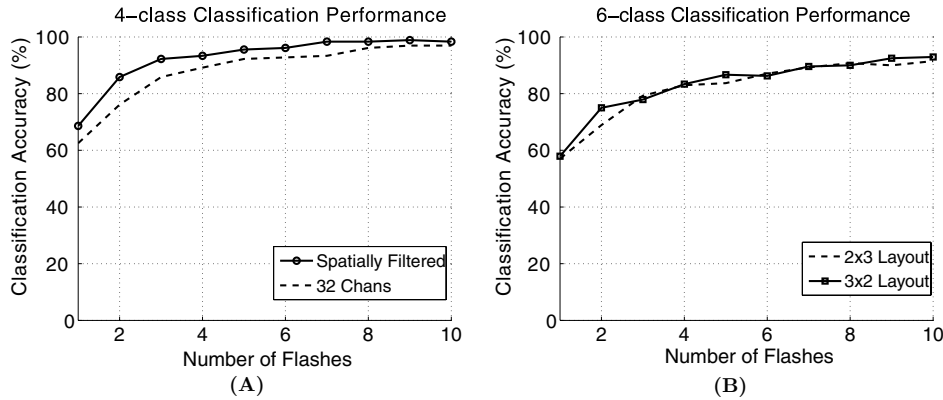
$$J(f) = \frac{f^T S'_b f}{f^T S'_w f},$$

where  $S'_b, S'_w$  are calculated directly from the set of  $E_i$ s [21]. This is a generalized eigenvalue problem and its solution is an orthonormal set of vectors  $f$  ordered numerically by their eigenvalue. We choose the first three filters  $f$  to represent our data. The choice was made empirically from pilot experiments, where we observed that the first 2–3 filters captured most of the discriminative information and that additional filters did not improve classification performance.

We used the LIBSVM classification package [22] to classify the spatially projected data. A linear soft-margin classifier was used, and the free parameter was estimated using cross-validation on the training data.

### 1.4. Brain–robot interface design

We designed an image-based BCI that was dynamic and could accommodate a variable number of images by suitably resizing



**Figure 1.** Performance of the brain–computer interface. (A) Speed–accuracy tradeoffs in the interface. The figure shows the average accuracy across all subjects as a function of the number of flashes for each choice. Note that even with only five flashes, the accuracy rises to roughly 95%, suggesting that commands could be issued at a faster rate if need be with only a small loss in accuracy. The figure also shows that the spatial filter method is on average better than using the raw data from 32 channels. (B) Performance on a six-choice task. The plot shows the average classification accuracy across subjects as a function of the number of flashes per stimulus. The two curves represent two different layouts of the six images on the screen:  $2 \times 3$  and  $3 \times 2$ .

and arranging them in a grid. This allows the interface to incorporate dynamically generated images discovered by the robot and present them as choices to the user.

During a selection procedure, the interface flashes a thin border around each image in a random sequence. The subject focuses attention on the image of interest (for instance, by counting the number of flashes of that border). The flashes occur once every 0.25 s and consist of a red border being visible for the first 0.125 s.

The interface implements the classification scheme described above. The parameters for the data processing, i.e. the spatial projection filter and the linear classifier, are obtained by a training procedure before using the interface. The training session consists of 10 min of data collection in a protocol similar to the user study above. In our online interface, we perform the spatial projection and downsampling first, in order to achieve substantial data reduction. The resulting three projected channels are then bandpass filtered and classified for each flash. A running total of classifier output is maintained for each image, so as to give a continuous estimate of the target image. At any step, the image with the current maximum total is the most likely selection. This running total can be used, for example, to cut short the process after fewer flashes, to focus on fewer images that the classifier is having difficulty distinguishing between, or to increase the number of flashes in the case of uncertainty. Our image-based interface was implemented using the general-purpose BCI2000 software [23].

The robot used was a Fujitsu HOAP-2 humanoid robot with 25 degrees-of-freedom, including a pan-tilt camera head for visual feedback to the user and two hands for grasping objects. The robot was programmed to be able to autonomously walk, navigate based on visual information, and pick up and place objects. The robot used its vision to locate objects of interest, segment them and send them to the BCI as choices for the user. We also used an overhead camera to locate tables that contained objects and that could serve as destinations for the objects that have been picked up.

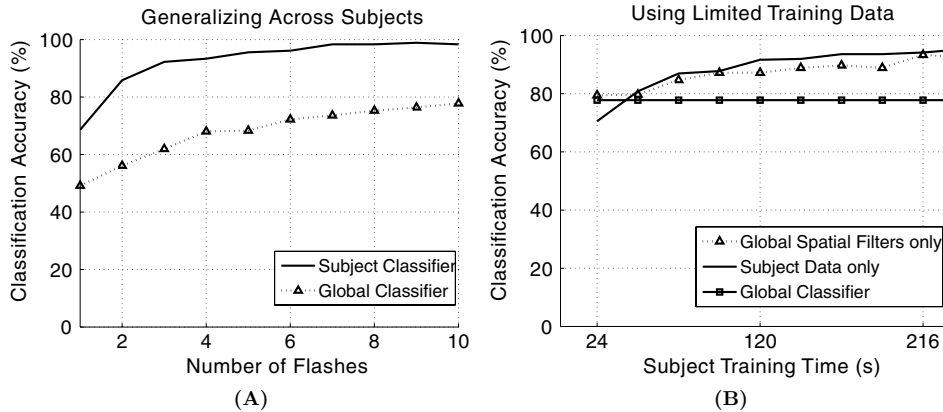
## 2. Results

In the first set of experiments, we collected data from nine users and used the data to test the performance of the P3 classifier used in the brain–computer interface. Users participated in two sessions, each consisting of a number of trials involving four randomly selected images displayed on a computer screen. We used the data from the first session to compute the first three maximally discriminative spatial filters for these data and trained a binary classifier to detect the user’s P3 responses (see section 1 for details). We tested the performance of this classifier on data from the second session. The resulting classification accuracy across subjects was 98.4% with a standard deviation of 2%, i.e. there were almost no misclassified points. Classification performance without the spatial filters, i.e. using all 32 raw EEG channels, was 97%.

In order to explore whether we could reduce the number of flashes used per stimulus and thus study speed–accuracy tradeoffs, we measured classification accuracy as a function of the number of flashes used. Figure 1(A) shows this accuracy measure, averaged across all subjects, with and without the use of the spatial filtering algorithm. Note that the chance level accuracy for this four-class problem is 25%. As shown in the figure, even with as few as five flashes, the accuracy is 95% across subjects. Thus, for a four-choice selection problem, a command can be issued every 5 s with an accuracy of 95%, resulting in a bit rate of  $24 \text{ bits min}^{-1}$ .

### 2.1. Varying the number of choices

Our interface is designed to dynamically accommodate a variable number of choices that are detected and communicated by the robot. A natural concern is whether changing the layout or number of images affects performance. To test this, we used both four-image sessions as training data, and tested this classifier on tasks involving choosing from six



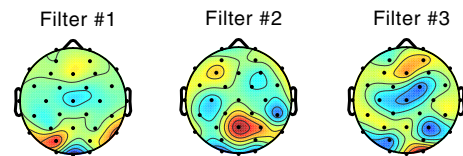
**Figure 2.** Effect of training data on classifier performance. (A) Generalization across subjects. The figure compares the performance of a pre-trained classifier on a novel subject (global classifier) with the performance of a classifier trained on the subject’s own training data set. The average classification accuracy across all subjects is shown as a function of the number of flashes per image. The subject-specific classifier is better as expected, but the pre-trained ‘generic’ classifier (‘global classifier’) performs significantly better than chance (78% accuracy versus 25% chance). (B) Effect of training time. The figure shows classification accuracy across subjects as the amount of training time is increased. Average accuracy for three classifiers are shown: (1) the baseline when a pre-trained across-subject classifier is used (global classifier), (2) a subject-specific classifier built on a generic spatial projection (global spatial filters only) and (3) the subject-specific classifier with subject-specific spatial filters (subject data only). We see that with 3–4 min of training data, the subject-specific classifier shows accuracy above 95%.

images in  $2 \times 3$  and  $3 \times 2$  layouts<sup>1</sup>. As seen in figure 1(B), the classification accuracy on this six-class problem remains high at 93% (standard deviation 3.1% across subjects), as compared to a 16.6% chance level. These results demonstrate that increasing the number of options on the screen from four to six and changing their layout do not significantly impact the performance of the interface.

## 2.2. Generalizing across subjects

We also examined the question of whether it was possible to design a classifier that would generalize across subjects; this would allow one, for instance, to forgo the initial data gathering step for training a subject-specific classifier. We trained a classifier on the data from all but one subject and tested performance on the data from that subject. Figure 2(A) shows the average classification accuracy of this scheme, averaged across all subjects as a function of the number of flashes. The classification accuracy when using the subject’s own data for training is also shown for comparison. It is clear that although a classifier trained on the subject’s own data performs better, the performance of the generic classifier is significantly better than chance (accuracy of about 78% for the ten-flashes case compared to a chance level of 25%). This supports the hypothesis that the visually evoked P3 response to flashed images is robust and shares similarities across subjects. Another recent study [24] using a larger number of options in a speller paradigm found that there was significant individual variation between subjects and that a generalized classifier performed worse than a subject-specific classifier.

<sup>1</sup> For subject 2, the  $3 \times 2$  session was dropped since a malfunction of the EEG device corrupted the recorded signals.

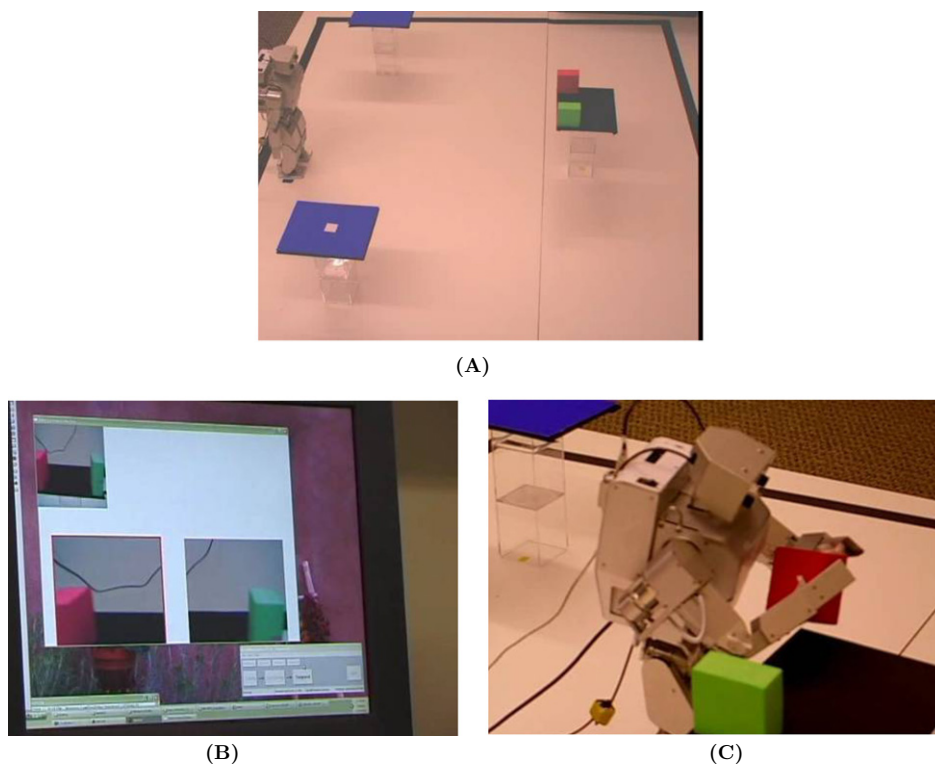


**Figure 3.** Characteristic features of the evoked response. Shown are the first three spatial filters learned from the data across all subjects (red denotes high positive values; blue denotes high negative values). The spatial filters are smoothly varying and are focused on the occipital and parietal regions, regions that are appropriate for a visuo-spatial attention task such as the one utilized by the interface.

## 2.3. Training time required

To investigate how the amount of time devoted to collecting training data affects the performance of the interface, we computed classification accuracy on test data using varying amounts of training data. The data used for training/testing were average responses to each image across the ten flashes in each trial. Figure 2(B) shows the average classification accuracy on test data. For comparison, the accuracy from the across-subjects classifier (figure 2(A), accuracy with all ten flashes) is shown as the baseline case where no training data are used. As seen in the figure, the average accuracy rises steeply with more input training data as expected, but even with only 3–4 min of training data, the accuracy is already up to about 95%. These results support the use of a short training session for each subject for learning a personalized classifier rather than relying on a generic P3 detector.

Finally, to gain a better understanding of the properties of the EEG responses used in this study, we examined the three spatial filters learned from the combined data from all subjects. These filters can be viewed as capturing the distinctive features of the EEG signals across the scalp that characterize the evoked response to a target flash. As seen in figure 3, the parietal and



**Figure 4.** Example of a humanoid robot control using the interface. (A) The robot in an arena consisting of a target table with two objects and two destination tables at nearby locations. (B) The interface showing the robot's eye view and the two discovered objects presented as selection options. (C) The robot picking up the object selected by the user. A video file demonstrating the system in action has been included as supplementary data, available online at [stacks.iop.org/JNE/5/214](https://stacks.iop.org/JNE/5/214) with this paper.

occipital electrodes appear to be significantly involved in a discriminative activity. This is consistent with other studies involving P3 spellers (e.g., [25]), where parietal electrodes have been found to contribute significantly to performance in addition to occipital electrodes classically used for detecting the P3 response.

#### 2.4. Control of the robot using the interface

We used the brain–computer interface to allow users to command a Fujitsu HOAP-2 humanoid robot to pick up a desired object and bring it to a selected location. Figure 4(A) shows an example scenario with two colored objects on a table available for pick-up and two other tables as destination locations. The objects were captured with the robot's cameras, segmented and presented as choices to the user via the brain–computer interface. The destinations were discovered and segmented from an image taken by an overhead camera.

For control of the robot by a given user, we collected data from the user in a 10 min training session using the protocol described above (see section 2). These data were used to learn the spatial filters and the classifier. The user used this trained interface to make selections between various image choices depicting objects or destination locations on the computer screen (figure 4(B)). The user's choice, as inferred by the interface, is conveyed to the robot, which autonomously orients itself with respect to the selected object using visual feedback and executes a sequence of motor commands to pick

up the object (figure 4(C)). Similarly, given the user's choice of a destination location, the robot walks to the selected location using visual feedback and prior knowledge of the environment, and delivers the object to that location. A video of the brain–robot interface in action is included as supplementary data, available online at [stacks.iop.org/JNE/5/214](https://stacks.iop.org/JNE/5/214). A second supporting video depicts a typical P3 response elicited from a subject during the operation of the interface.

### 3. Discussion

Our results demonstrate that non-invasive brain signals can be used to command a sophisticated mobile device such as a humanoid robot to perform a useful task such as navigating to a desired location and fetching a desired object. We used computer vision for discovering interactions afforded by the environment, a dynamic interface to communicate these discovered choices to the user and a semiautonomous robot to implement the user's choices. In principle, the robot could execute a wider range of actions, and arbitrary commands could be presented to the user as choices using images describing available actions.

A BCI system such as the one explored in this paper could potentially be used, for instance, in medical helper robots for paralyzed and disabled patients. Such robots would possess the ability to move about in a home, perform different manipulative actions on objects and provide visual feedback to

the patient for monitoring progress and for subsequent action selection. Such robots could potentially act as surrogate bodies for the paralyzed that are remotely commanded at the cognitive level by a BCI. Our P3-based system can be viewed as a first step toward developing sophisticated robotics-enhanced BCIs, with robotics playing a more complex role than just navigation alone as in previous BCI applications.

Other types of EEG responses, such as steady-state visual-evoked potentials (SSVEPs) [26], EEG during mental tasks [27] and spectral power in particular frequency bands (e.g.,  $\mu$  (8–12 Hz) or  $\beta$  (18–26 Hz) bands) during motor imagery [1], could also be used for command selection as an alternative to or possibly in conjunction with P300. Some of these approaches, such as motor imagery, may require more training data and subject training time, and show wider variation across subjects, but may allow finer grained control. We chose a P300-based approach because it requires very little subject training, given that the P300 is a robust-evoked response across subjects. In addition, it is easier to select between larger numbers of options using the P300 response, as seen in, for example, speller paradigms. However, it does suffer from the limitation that control is coupled to stimulus presentation. We intend to explore the use of some of the other types of EEG responses mentioned above for self-paced asynchronous control of the robot in the near future.

The current implementation of our interface provides a bit rate of up to 24 bits  $\text{min}^{-1}$  with an accuracy of 95% (four choices). Our results show that such a bit rate, which is comparable to the bit rates of other EEG-based BCIs, is sufficient for high-level coarse-grained control of semi-autonomous robots such as the humanoid robot used in this study. Invasive BCIs such as cortical implants would allow control at finer temporal granularity, albeit with the risks associated with invasive devices. Our results suggest that the advent of adaptive, autonomous robotic devices could help pave the way for a new generation of non-invasive brain-machine interfaces that allow direct closed-loop interaction with the physical world.

## Acknowledgments

This research was supported by the National Science Foundation, an ONR Young Investigator award, and a David and Lucile Packard Fellowship to RPNR.

## References

- [1] Wolpaw J R and McFarland D J 2004 Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans *Proc. Natl Acad. Sci.* **101** 17849–54
- [2] Pfurtscheller G *et al* 2000 Current trends in Graz brain computer interface (BCI) research *IEEE Trans. Rehabil. Eng.* **8** 216–9
- [3] Kubler A *et al* 1999 The thought translation device: a neurophysiological approach to communication in total motor paralysis *Exp. Brain Res.* **124** 223–32
- [4] Serruya M D, Hatsopoulos N G, Paminski L, Fellows M R and Donoghue J P 2002 Brain-machine interface: instant neural control of a movement signal *Nature* **416** 141–2
- [5] Musallam S, Corneil B D, Greger B, Scherberger H and Andersen R A 2004 Cognitive control signals for neural prosthetics *Science* **305** 258–62
- [6] Wessberg J *et al* 2000 Real-time prediction of hand trajectory by ensembles of cortical neurons in primates *Nature* **408** 361–5
- [7] Taylor D M, Tillery S I and Schwartz A B 2002 Direct cortical control of 3D neuroprosthetic devices *Science* **296** 1829–32
- [8] Hochberg L R, Serruya M D, Friebs G M, Mukand J A, Saleh M, Caplan A H, Branner A, Chen D, Penn R D and Donoghue J P 2006 Neuronal ensemble control of prosthetic devices by a human with tetraplegia *Nature* **442** 164–71
- [9] Vora J Y, Allison B Z and Moore M M 2004 A P3 brain computer interface for robot arm control *Society for Neuroscience Abstract 30, Program No 421.19*
- [10] Pfurtscheller G, Muller G R, Pfurtscheller J, Gerner H J and Rupp R 2003 ‘Thought’-control of functional electrical stimulation to restore hand grasp in a patient with tetraplegia *Neurosci. Lett.* **351** 33–6
- [11] Farwell L A and Donchin E 1988 Talking off the top of your head: toward a mental prosthesis utilizing event-related brain potentials *Electroencephalogr. Clin. Neurophysiol.* **70** 510–23
- [12] Birbaumer N *et al* 1999 A spelling device for the paralyzed *Nature* **398** 297–8
- [13] del Millan J, Renkens F, Mourino J and Gerstner W 2004 Noninvasive brain-actuated control of a mobile robot by human EEG *IEEE Trans. Biomed. Eng.* **51** 1026–33
- [14] Rebsamen B, Burdet E, Teo C L, Zeng Q, Guan C, Ang M and Laugier C 2006 A brain control wheelchair with a P300 based BCI and a path following controller *The 1<sup>st</sup> IEEE/RAS-EMBS Int. Conf. on Biomedical Robotics and Biomechatronics (BioRob 2006), Pisa, Italy, 20–22 February 2006* pp 1101–6
- [15] Luth T, Ojdanic D, Friman O, Prenzel O and Graeser A 2007 Low level control in a semi-autonomous rehabilitation robotic system via a brain-computer interface *IEEE 10<sup>th</sup> Int. Conf. on Rehabilitation Robotics (ICORR 2007) June 13–15, Noordwijk, The Netherlands, 2007* pp 721–8
- [16] Keirn Z A and Aunon J I 1990 A new mode of communication between man and his surroundings *IEEE Trans. Biomed. Eng.* **37** 1209–14
- [17] Bayliss J D and Ballard D H 2000 Recognizing evoked potentials in a virtual environment in S A Solla, T K Leen and K-R Miller, eds *Advances in Neural Information Processing Systems* vol 12 pp 3–9 (Cambridge, MA: MIT Press)
- [18] Krusienski D J *et al* 2006 A comparison of classification techniques for the P300 speller *J. Neural Eng.* **3** 299–305
- [19] Thulasidas M, Guan C and Wu J 2006 Robust classification of EEG signal for brain computer interface *IEEE Trans. Neural Syst. Rehabil. Eng.* **14** 24–9
- [20] Jasper H H 1958 The ten-twenty electrode system of the international federation *Electroencephalogr. Clin. Neurophysiol.* **10** 371–5
- [21] Hoffmann U, Vesin J-M and Ebrahimi T 2006 Spatial filters for the classification of event-related potentials *Proc. ESANN* pp 47–52
- [22] Chang C-C and Lin C-J 2001 LIBSVM: a library for support vector machines (Software available at url <http://www.csie.ntu.edu.tw/~cjlin/libsvm>)
- [23] Schalk G, McFarland D J, Hinterberger T, Birbaumer N and Wolpaw J R 2004 BCI2000: a general-purpose brain-computer interface (BCI) system *IEEE Trans. Biomed. Eng.* **51** 1034–43

- [24] Sellers E W, Krusienski D J, McFarland D J, Vaughan T M and Wolpaw J R 2006 A p300 event-related potential brain-computer interface (BCI): the effects of matrix size and inter stimulus interval on performance *Biol. Psychol.* **73** 242–52
- [25] Krusienski D J, Sellers E W, McFarland D J, Vaughan T M and Wolpaw J R 2008 Toward enhanced p300 speller performance *J. Neurosci. Methods* **167** 15–21
- [26] Muller-Putz G, Scherer R, Brauneis C and Pfurtscheller G 2005 Steady-state visual evoked potential (ssvep)-based communication: impact of harmonic frequency components *J. Neural Eng.* **2** 123–30
- [27] Millán J del R, Renkens F, Mouriño J and Gerstner W 2004 Non-invasive brain-actuated control of a mobile robot by human EEG *IEEE Trans. Biomed. Eng.* **51** 1026–33