Proceedings of the
7th Graz Brain-Computer Interface Conference 2017 DOI: 10.3217/978-3-85125-533-1-17

REAL-TIME FMRI CONTROL OF A HUMANOID ROBOT USING TWO BRAIN NETWORKS SIMULTANEOUSLY: A PILOT STUDY

Ori Cohen[1,2], François Keith[3], Moshe Koppel[2], Abderrahmane Kheddar[4,5], Rafael Malach[6], and Doron Friedman[1]

[1]The Interdisciplinary Center, Herzliya 46150, Israel
[2]Bar-Ilan University, Ramat-Gan 52900, Israel
[3]CEA-Saclay, France
[4]CNRS-UM LIRMM, 161 rue Ada 34095 Montpellier Cedex 5, France
[5]CNRS-AIST Joint Robotics Laboratory, UMI3218, 305-8568 Tsukuba, Japan
[6]Weizmann Institute of Science, Department of Neurobiology, Rehovot 76100, Israel

E-mail: doronf@idc.ac.il, orioric@gmail.com

ABSTRACT: We have previously shown that real-time fMRI, despite the low temporal resolution of the BOLD signal, can be used for BCI navigation, using motor-imagery and -execution. Here we leverage the superior spatial resolution of fMRI to implement a BCI paradigm going beyond a single brain network for control, while retaining an intuitive mapping between brain activity and BCI functionality. The experiments simulate non-trivial navigation and item selection tasks by a subject teleoperating an HRP-4 humanoid-robot. Motor actions are mapped into simple navigation commands inside a room and visual attention is mapped to direct the robot's arm toward one of three objects placed on a table. When the correct item has been selected, the subject navigates the robot toward the experimenter in order to simulate the delivery of the object. We describe a method based on two parallel classifiers, with four and three classes (independent of the first four), offline and real-time classification results from a single-subject pilot, performing several times.

INTRODUCTION

This research is part of a thread of studies aimed at dissolving the boundary between the human body and a surrogate robotic representation in a physical reality. The subject is expected to act as if the robotic body is his own body, and our aim was to provide the subject with an intuitive thought-based control of this surrogate representation. The subject was located in Israel and the robot was located in France; this geographic split was only made due to the availability of the facilities.

Electroencephalogram (EEG)-based brain-computer interface (BCI) for device control, despite much recent progress, is still mostly based on three paradigms (with some variants): motor imagery, P300, and steady state visually evoked potential (SSVEP). Our overarching goal in this research is to leverage the superior spatial resolution of blood-oxygen-dependent-signal (BOLD) in order to explore novel BCI control paradigms based on multiple brain systems simultaneously, such that we map different types of mental patterns to relevant functional goals, approximating a realistic task. Specifically, in this study we allow the subject to navigate using three motor classes, to select one of three objects using the visual system, and a null class.

There have been several studies including EEG-based BCI control of avatars [1, 2, 3] and teleoperation of a humanoid robot [4, 5], including studies with spinal cord injured people [6]. We have demonstrated teleoperating a humanoid robot using motor imagery and execution with real-time functional magnetic resonance imaging (fMRI) [7, 8], and others have demonstrated navigating a robot using covert visuospatial attention [9]. In this pilot study we aim going beyond these studies, using two different brain systems simultaneously.

MATERIALS

fMRI scans were performed on a 3T Trio Magnetom Siemens scanner as described in [7, 10], with a repetition time (TR) of 2000ms. Our system includes a tool for whole brain classification of raw data in real-time as described in [11]. Visual feedback is provided by a mirror, placed 11cm from the eyes of the subject and 97.5cm from a screen, which results in a total distance of 108.5cm from the screen to the eyes of the subject.

We used the vision system framework (VSF)[1] to acquire, transcode and transmit the video stream between the scanner (in Israel) and the robot (in France) with minimum latency.

METHODS

We created a complete software suite for running a wide range of real-time fMRI studies, which is able to process

[1] https://github.com/LIRMM-Beziers/visionsystem

 DOI: 10.3217/978-3-85125-533-1-17

both brain data arriving in real time from the fMRI scanner and pre-recorded fMRI data [11]. It supports various experimental protocols, includes several analysis methods, is integrated with the Unity3D game engine for virtual environment feedback, and can interface with other external devices. Our tool is efficient in terms of processing, can be configured for a wide range of experimental protocols and was previously tested in several types of real-time fMRI BCI experiments. It is based on statistical machine learning classification of subjects' brain state in real time, based on whole brain activity.

Figure 1: Sample stimuli for testing the visual category task. From left to right the categories are: faces, tools and houses.

Training and applying classifiers in real-time requires that learning be executed faster than is generally done in the application of machine learning to fMRI. Our system is optimized for memory usage, processing speed, and classification speed using feature reduction, feature selection, and redundant data reduction. The system uses pre-recorded raw brain data for the purpose of learning a classifier using Platt's sequential minimal optimization (SMO) version of the support vector machine (SVM) learning algorithm [12]. The system culls empty voxels and the subject's eyes and corrects non-linear non-homogeneous drifts. For classification and feature selection we use Weka, which is a collection of machine learning algorithms [13]. For feature selection we use the information gain (IG) measure to select the most relevant voxels [14]. The filtered dataset is passed into Weka's [15] implementation of multiclass [16] SVM [12], using default parameters. The result of the training phase is an SVM classifier model that can classify previously unseen vectors. The system automatically verifies that the model classifies the training data with perfect accuracy ("test on train" for sanity check) and displays the selected voxels.

In the real-time classification stage, the subjects perform a task and the system classifies their intentions in real time. The system classifies a brain scan every time resolution unit (TR), which in this case is 2 seconds. It uses the filtering and normalization methods as in the training stage and select the same voxels based on the IG filtering performed at model training. The data is then passed into the trained SMO model, and the classification result is then transmitted to the external application using a user datagram protocol (UDP). The classification process takes approximately 50 milliseconds. Before moving to a free choice task the subject undergoes a cue-based part of the study, the task is similar to training but feedback is provided based on real time classification.

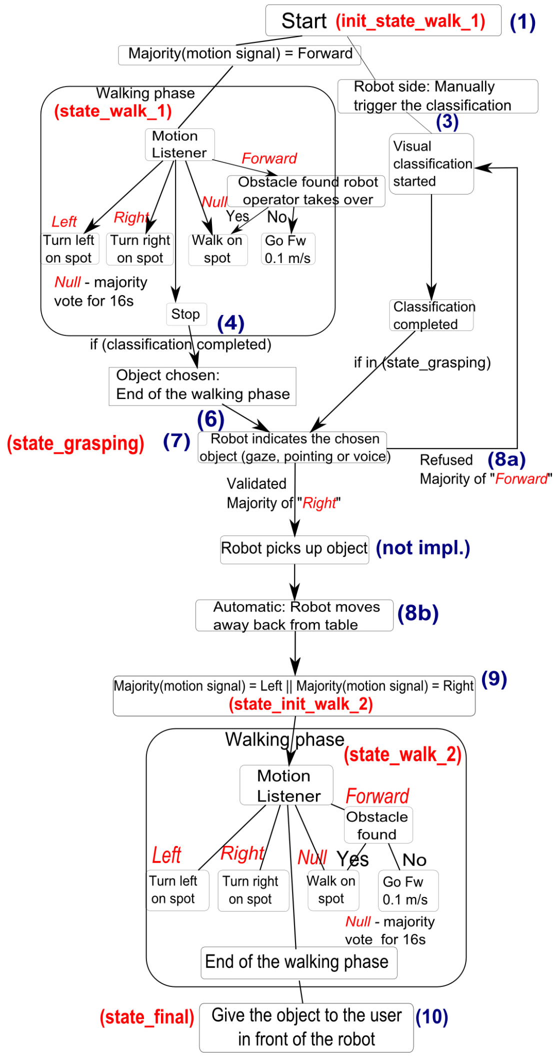

Figure 2: The HRP-4 humanoid robot's state-machine protocol for the visual-motor task.

The subject (female, 31) underwent several training sessions for each classification task: i) motor execution (4 sessions) – moving left fingers, right fingers, toes, and a null (rest) class; ii) motor imagery (4 sessions) – imagining left hand, right hand, feet, and a null class; and iii) visual categories (5 sessions) – viewing images of houses, faces and tools. The motor execution classifier was trained 3 months prior to this experiment, the motor imagery classifier was trained 2.5 months prior and the visual categories classifier was trained 24 days prior.

Each motor training session included 40 events from each category, i.e., for four training sessions there are 160 labelled samples; all details are as in [10]. For visual classification training is done using a block design – a sequence of images from the same category is flashed, one per second, for 12 seconds, followed by a duration of six seconds during which the category images were painted in white to allow the signal to return to baseline. The subject

 DOI: 10.3217/978-3-85125-533-1-17

was trained with 36 image sequences for each category, i.e., for five training sessions there were 180 samples. In test sessions each stimulus includes three images from the three categories simultaneously, and the target category is indicated by a fixation dot (Fig. 1). The subject carried out three such test runs, with 30 stimuli in each session (10 from each category).

The motivation behind the visual paradigm is to allow the subject to select one of several objects by visual attention, even if the multiple objects are seen together. Our processing pipeline assures that information from the areas surrounding the eyes is pruned prior to being fed into the machine learning system [11]. Thus, we expect that the machine learning system can classify this task by decoding, in real time, the activity of the well-known visual areas in the cortex, corresponding to the visual categories of faces, tools, and houses. Such decoding in real time and as part of a BCI task has never been attempted to our knowledge, and the task is especially challenging given that the training and testing of the algorithm are not done in the same conditions – the training is based on a single image display and the testing is based on three images shown simultaneously. The goal in this paradigm is to move towards new naturalistic BCI paradigms, such that training is done in controlled conditions, and the model can be applied in real time, in naturalistic conditions, which may be different than those available for training. In our case, we expect the classifier to be trained separately over different object categories, whereas the free choice task includes multiple object categories at the same time.

The experiment and the robot's control were based on a predefined state machine (Fig. 2). The task included several stages, including navigation and object selection (see Fig. 3 and companion video[2]). The robot was placed in a fixed orientation. First, the subject steered the robot towards a table, by passing an obstacle chair and by utilizing all four motor commands in order to reach the table (Fig. 4, 3). The subject was instructed to guide the robot around the chair and then turn towards the printed sign as seen in 4(b) (on the right). The subject had to read, through the eyes of the robot, the instructions from the sign with the target object to select. After seeing the instruction, the subject was expected to navigate towards a table, eventually stopping in front of it. On top of the table we placed three objects: a (toy) doll's head, a (toy) house, and a tool (either a hammer or a tea cup). Prior to the experiment, the subject was instructed that in order to select an object she had to rotate towards the sign to learn about the target object, and from that moment she had to focus her attention to the target physical object on the table at all times and study it, including while walking to table, until it is selected. For example, if the experimenter revealed the word "face" then the subject had to focus attention to the head's eyes, nose and chin, i.e., pay attention to the features of the object. If the word "tool" is revealed, then the subject had to imagine herself using that tool.

[2]http://y2u.be/eYSb9Q5PcP8

After walking the path, the subject was expected to stop within grasping distance from the objects. Once the subject reached the table the steering was deactivated, the robot stood still and the robot's left hand was pointed toward one of the items on the table. The item was selected by a majority vote from the classification at times 0-16 seconds following the instruction from the experiment (8 TRs). The 16 seconds delay was based on the optimal classification time as determined in offline evaluation. If a majority vote did not take place (i.e., a draw) the classification continued until there was a decision.

Following the classification, the subject had to indicate whether she agrees with the selection or not using motor categories. The subject used either motor execution or motor imagery (in different runs), as follows: feet – try again, right hand – activate a grasp motion. If the motor action was classified as null, then the subject received feedback indicating her to repeat the motor action. The subject repeated this step until she was satisfied with the category that was selected. Immediately after the subject activated the grasp motion by selecting the category, the robot takes a few steps backward away from the table. Only then the system is re-activated the steering and the subject was allowed to control the robot and navigate it towards the experimenter.

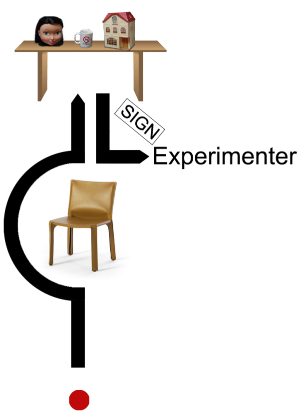

Figure 3: A schematic drawing of the intended walking path inside the room. The red dot indicates the robot's fixed position. The path goes around the chair rotating towards the sign, and then towards the table that has the three category items placed on top. Following the visual task, the path continues toward the experimenter.

During the task our system ran two classifiers in parallel: motor (motor execution or motor imagery) and visual categories, the former with four categories and the latter

 DOI: 10.3217/978-3-85125-533-1-17

with three. The subject teleoperated the robot using all seven classes: left, right, forward, null class, house, face and tool. The longest run was 12 minutes. For each run the subject was assigned with a different target category (face, house, or tool). During the navigation part of the experiment, the flow of high level commands (forward, left, right, null, face, tool, house) was sent to the robot through a user UDP connection with a latency of 100-150 milliseconds using the VSF.

RESULTS

Throughout the experiment we used three classifier models: motor execution, motor imagery, and visual categories. A combination of motor execution and visual categories was used on the first day of the experiment, and a combination of motor imagery and visual categories on the second day. Estimating accuracy for the free choice is difficult, so we provide offline evaluation of the classification models.

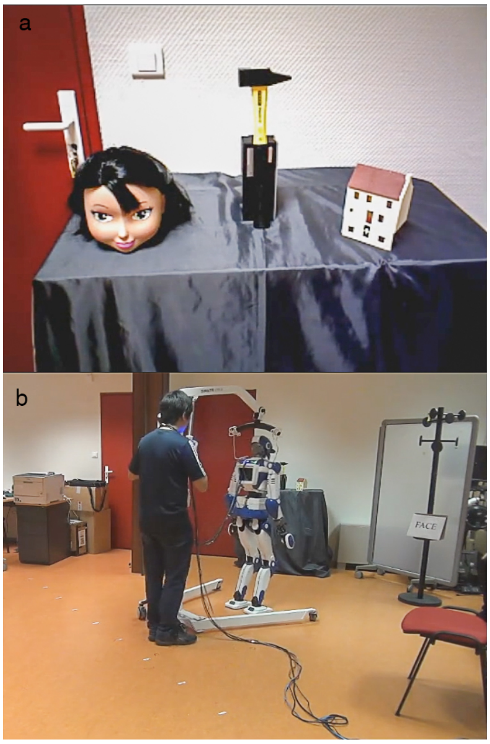

Figure 4: The HRP-4 humanoid robot during a visual-motor task, standing in front of three objects. Top: the target objects as seen from the robot camera by the subject, bottom: the robot performing the task, escorted by an experimenter for robot safety.

Offline analysis of the motor classifier is based on a single training session (Fig. 5). As expected, a cue-based session with real time feedback using the same classifier yields similar results. More real-time data is required to assess the difference in accuracy in TR3. The motor execution classifier was trained and tested three months prior to this real-time experiment and test. Similarly, single-run motor imagery classification accuracy results can be seen in Fig. 6; we note that the pre-recorded and real-time tests were done 3 & 2.5 months apart, respectively.

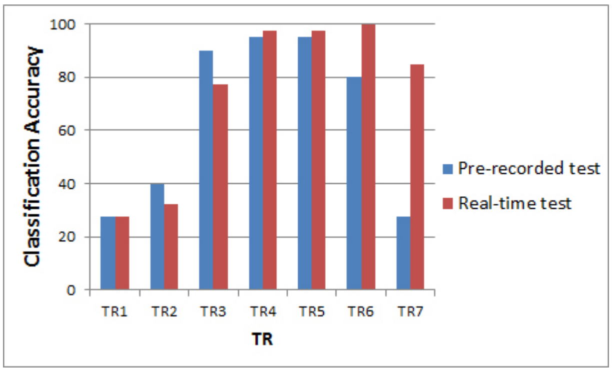

Figure 5: Motor execution cue-base classification accuracy comparison between a single pre-recorded and a single real-time run.

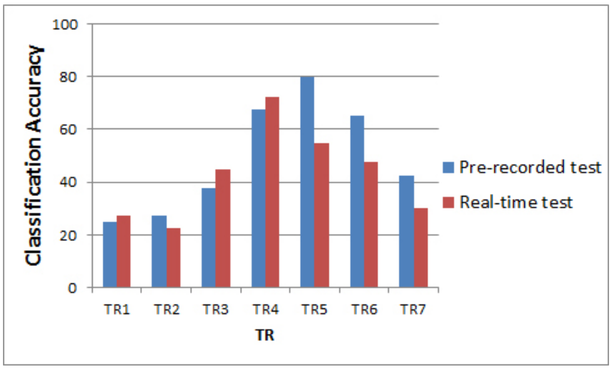

Figure 6: Motor imagery cue-base classification accuracy comparison between a single pre-recorded and a single real-time run.

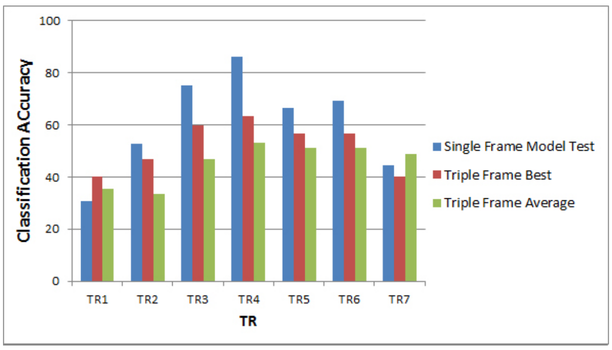

Figure 7: Cue-base classification accuracy, comparing single and triple frame (best run and average).

Offline analysis of the visual classifier was performed as follows. A single model was trained on five sessions – overall 180 stimuli, each including one category. The model was then tested in two conditions: i) another run of 36 stimuli with one image displayed on screen, and ii) a

DOI: 10.3217/978-3-85125-533-1-17

run of 36 stimuli, each comprising of all three categories displayed on the screen simultaneously. Fig. 7 presents the results, indicating that while both methods perform significantly better than chance (33%), testing on a single image is superior to testing on three parallel images.

A qualitative assessment of the subject's performance can be provided for the free choice task. In the motor execution experiment the subject performed the navigation part successfully in all four experimental sessions. In the motor imagery experiments the subject failed to stop the robot near the table and was only successful in the third session.

Thus, our subject had five attempts at the visual task: four in the motor execution conditions and one in the motor imagery condition. The subject succeeded in all cases, but only in the second attempt (in all of the motor execution sessions) or in the fourth attempt (in the motor imagery session).

The classification seemed to be skewed towards the face category. Fig. 8(A) shows several red dots that correspond to a successful classification of the "face" command in each time point during 8 TRs. However, when the subject was instructed to focus on one of the two other categories (house or tool) there was category rivalry between the classes. Fig. 8(B) is a rivalry example that show a fluctuation between "face" and "house". When category rivalry occurs, it prolongs the classification stage and it is harder to get a majority vote. In other words, without rivalry there are less classification attempts and a majority vote occurs quickly.

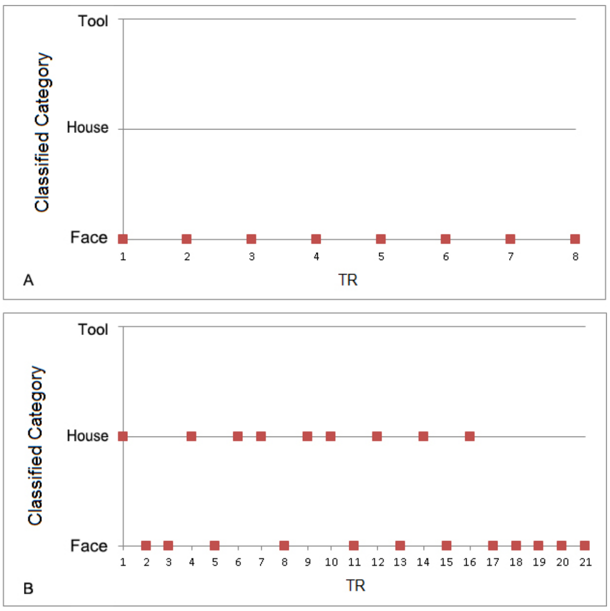

Figure 8: Visual categories real-time classification examples. The subject was instructed to focus on a) face, b) house.

DISCUSSION

We have developed a novel paradigm, based on simultaneous classification of both motor and visual brain networks, and have evaluated it in the context of a complex navigation and object-selection task, involving teleoperating of a humanoid robot. The pilot study with one subject serves to demonstrate that our system is fully operable, and provides a preliminary evaluation of the paradigm. Our results indicate that the task can be performed, although motor imagery and visual classification are challenging. Specifically, further work is required to refine the visual paradigm. Our offline evaluation results suggest that training subjects on simultaneous images (in the same fashion as the actual task) may be more appropriate.

ACKNOWLEDGEMENTS

This research was supported by the European Union FP7 Integrated Project VERE (No 657295), `www.vereproject.eu`. We would like to thank the subject for her participation, and the Weizmann Institute fMRI scanner staff Edna Furman-Haran, Nachum Stern and Fanny Attar for their help in this experiment.

REFERENCES

[1] C Kapeller, C Hintermüller, and C Guger. Augmented control of an avatar using an SSVEP based BCI. In *Proceedings of the 3rd Augmented Human International Conference*, page 27. ACM, 2012.

[2] J Faller, G Müller-Putz, D Schmalstieg, and G Pfurtscheller. An application framework for controlling an avatar in a desktop-based virtual environment via a software SSVEP brain–computer interface. *Presence: teleoperators and virtual environments*, 19(1):25–34, 2010.

[3] D Friedman, R Leeb, L Dikovsky, M Reiner, G Pfurtscheller, and M Slater. Controlling a virtual body by thought in a highly-immersive virtual environment. *GRAPP 2007*, pages 83–90, 2007.

[4] P Gergondet, S Druon, A Kheddar, C Hintermüller, C Guger, and M Slater. Using brain-computer interface to steer a humanoid robot. In *Robotics and Biomimetics (ROBIO), 2011 IEEE International Conference on*, pages 192–197. IEEE, 2011.

[5] E Tidoni, P Gergondet, A Kheddar, and S M Aglioti. Audio-visual feedback improves the BCI performance in the navigational control of a humanoid robot. *Frontiers in neurorobotics*, 8:20, 2014.

[6] E Tidoni, P Gergondet, G Fusco, A Kheddar, and S M Aglioti. The role of audio-visual feedback in a thought-based control of a humanoid robot: A BCI study in healthy and spinal cord injured people. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2016.

[7] O Cohen, S Druon, S Lengagne, A Mendelsohn, R Malach, A Kheddar, and D Friedman. fMRI robotic embodiment: A pilot study. In *Biomedical Robotics and*

 DOI: 10.3217/978-3-85125-533-1-17

Biomechatronics (BioRob), 2012 4th IEEE RAS EMBS International Conference on, pages 314–319, June 2012.

[8] O Cohen, S Druon, S Lengagne, A Mendelsohn, R Malach, A Kheddar, and D Friedman. fMRI-based robotic embodiment: Controlling a humanoid robot by thought using real-time fMRI. *PRESENCE: Teleoperators and Virtual Environments*, 23(3):229–241, 2014.

[9] P Andersson, J PW Pluim, M A Viergever, and N F Ramsey. Navigation of a telepresence robot via covert visuospatial attention and real-time fMRI. *Brain topography*, 26(1):177–185, 2013.

[10] O Cohen, M Koppel, R Malach, and D Friedman. Controlling an avatar by thought using real-time fMRI. *Journal of Neural Engineering*, 11(3):035006, 2014.

[11] O Cohen, M Ramot, R Malach, M Koppel, and D Friedman. A generic machine-learning tool for online whole brain classification from fMRI. In *The 6th International Brain-Computer Interface Conference*, 2014.

[12] J Platt. Sequential minimal optimization: A fast algorithm for training support vector machines. 1998.

[13] M Hall, E Frank, G Holmes, B Pfahringer, P Reutemann, and I H Witten. The weka data mining software: An update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

[14] J R Quinlan. Induction of decision trees. *Machine Learning*, 1(1):81–106, 1986.

[15] I H Witten, E Frank, and A M Hall. *Data Mining: Practical machine learning tools and techniques, 3rd Edition*. Morgan Kaufmann, 2011.

[16] T Hastie and R Tibshirani. Classification by pairwise coupling. In I.M. Jordan, J.M. Kearns, and A.S. Solla, editors, *Advances in Neural Information Processing Systems*, volume 10. MIT Press, 1998.