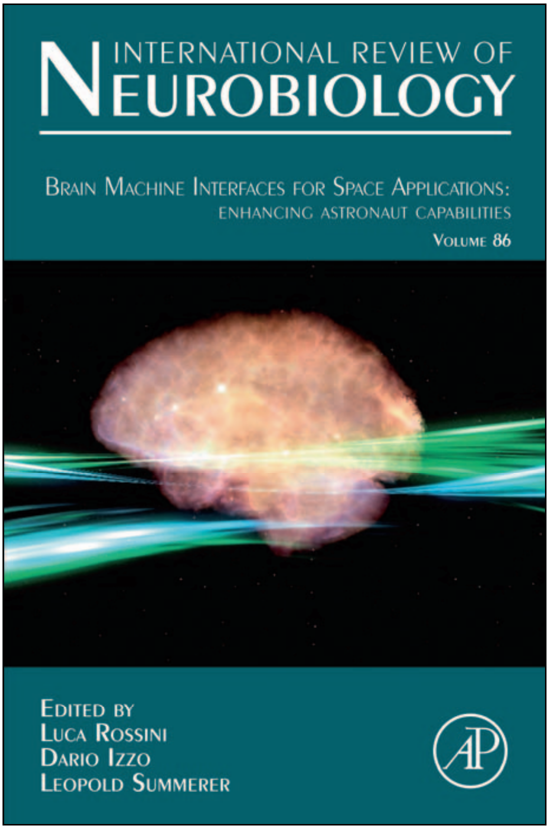

From: José del R. Millàn, Pierre W. Ferrez and Tobias Seidl, Validation of Brain–Machine Interfaces during Parabolic Flight.
In Luca Rossini, Dario Izzo and Leopold Summerer, editors: International Review of Neurobiology, Vol. 86, Burlington: Academic Press, 2009, pp. 189-197.
ISBN: 978-0-12-374821-8

VALIDATION OF BRAIN–MACHINE INTERFACES DURING PARABOLIC FLIGHT

José del R. Millàn,*,† Pierre W. Ferrez,* and Tobias Seidl‡

*Idiap Research Institute, Rue Marconi 19, 1920 Martigny, Switzerland
†Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
‡Advanced Concepts Team, European Space Agency, Keplerlaan 1, 2201 AZ Noordwijk, The Netherlands

Here we report on a validation study on brain–machine interfaces (BMIs) performed during the December 2007 ESA parabolic flight campaign. We investigated the feasibility of using BMIs for space applications by performing tests in microgravity. Brain signals were recorded with noninvasive electroencephalography before (calibration sessions) and during the parabolic flights on two subjects with prior BMI experience. The results of our experiments show that an experienced BMI user can achieve stable performance in all gravity conditions examined and, hence, demonstrate the feasibility of operating noninvasive BMIs in space.

I. Introduction

Triggered by the promising review of three Ariadna[1] studies (Carpi and De Rossi, 2006; Millán *et al*., 2006; Tonet *et al*., 2006) initiated by ESA's Advanced Concepts Team, we experimentally evaluated the functionality of BMIs in different gravity conditions, including microgravity, onboard a parabolic flight

[1]Ariadna is the name of a framework for cooperative research between the ESA Advanced Concepts Team and universities (http://www.esa.int/gsp/ACT/ariadna/index.htm).

DOI: 10.1016/S0074-7742(09)86014-5

0074-7742/09 $35.00

(47th ESA PFC campaign).[2] Brain signals were recorded with noninvasive electroencephalogram (EEG), currently the most promising BMI for space applications (see other chapters for the nature of EEG and the possibilities that BMI open up to astronauts). In this chapter we report the performance of two healthy volunteer subjects with some previous BMI experience during various experimental conditions, including the calibration session run on ground prior to the parabolic flights that is used as a baseline to compare flight performance. The analysis focuses on two different aspects of BMI, the mental commands sent by the user to drive the BMI and the error potentials (ErrP) generated by a feedback that does not match the subject's intent. These ErrP can be used as a verification procedure: if an ErrP follows the feedback associated to the BMI response, the system can cancel the command and therefore filter errors made by the BMI. Ferrez and Millán (2008a,b) describe ErrP for BMI and demonstrate their benefits.

II. Methods

Figure 1 shows the task subjects have to perform. It consists of mentally moving a virtual blue balloon on a standard computer display from a start position at the top of a pyramid to pseudo-randomly selected targets either on the left or on the right bottom of the pyramid. Every 2 s, the balloon goes down one step, either to the left or to the right depending on the BMI's interpretation of the user's mental command. The BMI continuously analyzes the subject's EEG signals to recognize his intent and makes a decision every 2 s. This classification process continues until the balloon reaches the bottom row. In parallel, after each single step of the balloon, the BMI analyzes a small time window to check the presence of an ErrP which would indicate an erroneous feedback (i.e., wrong response of the BMI).

During the parabola of the flight, subjects reached two targets per gravity phase (intertrial interval of around 2 s, for a total of around 18 s). As shown in Fig. 2, a parabola consists of five phases of 20 s each: normal gravity (1*g*), hypergravity (1.8*g*), microgravity (0*g*), hypergravity (1.8*g*), and normal gravity (1*g*). Each subject executed 15 parabolas. In addition, subjects performed 10 calibration sessions on ground a few days before the flight, each consisting of 10 targets equally distributed.

Data from the calibration sessions were used to build a classifier (see Section III for details). Then, during the parabolic flight, EEG preprocessing and classification was done online. However, the feedback delivered to the subjects was not the

[2]ESA Parabolic flight campaign: http://www.spaceflight.esa.int/users/index.cfm?act=default.page&level=11&page=paraf

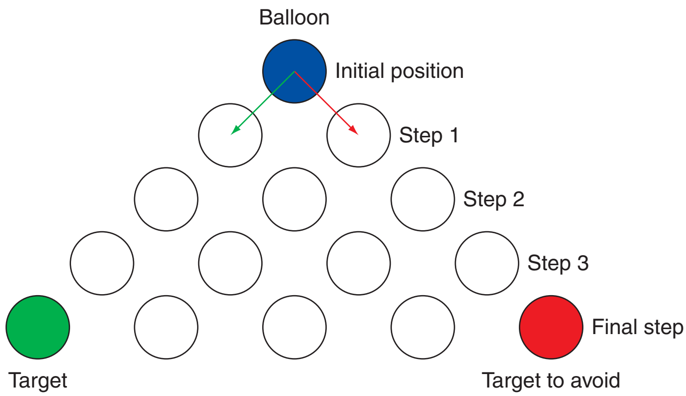

FIG. 1. Experimental task. The balloon (blue) appears at the top of the pyramid. The goal is to bring it to the green target (left in this example) that is chosen randomly. The subject executes the corresponding mental task (imagination of a left arm movement) until the balloon reached the bottom of the pyramid. The balloon makes a step down every 2 s, either to the left or to the right.

actual response of the BMI. Instead, the balloon moved with a 30% error rate—that is, at each step there was a 0.3 probability that the balloon moved to the wrong direction, thus replicating the performance of the online BMI (see Section III. A). It is our experience that this approach facilitates initial user training (either in early stages or in complex novel conditions) and yields EEG data of higher quality, even if the subjects are aware of the nature of the feedback. The reason is that it helps users to maintain their concentration and avoid frustration or confusion because of a poor performance of the BMI, which in our case can be due to dramatic changes in the EEG induced by hyper- or microgravity (Pletser and Quadens, 2003). Ultimately, this approach eliminates a potential showstopper during the first assessment of BMI for space applications.

In order to deliver mental commands, subjects were instructed to execute two mental tasks in a self-paced way—that is, at their own pace without needing any external stimulation. The two mental tasks were imagination of left hand movements, which is associated to the command "left," and words association, for the command "right." The words association task consists in searching for words starting with the same letter chosen randomly at the beginning of the trial. EEG signals were processed following the protocol described by Ferrez and Millán (2008a) and Millán *et al.* (2008). As a reminder, for recognition of mental tasks, we analyze EEG in the frequency domain and compute 112 EEG samples during the 8 s that lasts a trial; for ErrP detection, analysis is performed in the time domain and there are four EEG samples per trial, one per step of the balloon.

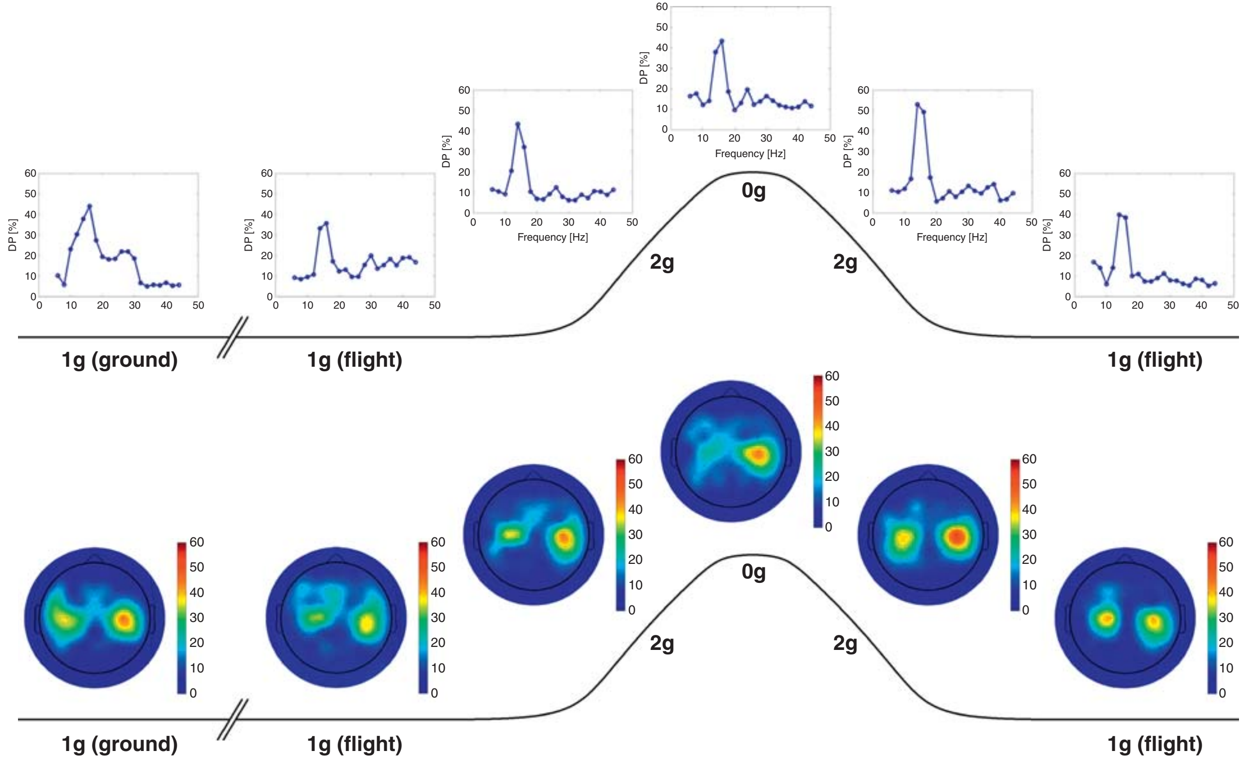

FIG. 2. Stability of features for subject 1 over the different gravity conditions. *Top:* Degree of relevance of frequencies. *Bottom:* Degree of relevance of EEG electrodes. Data for subject 2 are similar and omitted due to space restrictions.

We use machine learning techniques at two levels, namely feature selection and training the two classifiers embedded in the BMI. The approach aims at discovering subject-specific patterns embedded in the continuous EEG signal. At the first level, we select those features that are more relevant for recognizing either the mental tasks or ErrP. Thus, we select spatio-frequency features for mental tasks (relevant electrodes and frequency components) and relevant electrodes for ErrP. Ferrez (2007) and Millán *et al.* (2008) provide details of the different feature selection methods we use.

The vector of relevant features is extracted from each EEG sample and fed to a statistical Gaussian classifier. Its output is an estimation of the posterior class probability distribution for a single EEG sample; that is, the probability that the sample belongs to one of the two classes (left or right for mental commands, and error or correct for ErrP). In this statistical classifier, every Gaussian unit represents a prototype of one of the classes to be recognized, and we use several prototypes per class. During learning, the centers of the Gaussian units are pulled toward the samples of the class they represent and pushed away from the samples of the other class (see Millán *et al.*, 2004). For the classification of mental commands, the BMI combines the outputs of the Gaussian classifier over 2 s; while for ErrP recognition, the BMI simply takes the output of the classifier to each single sample.

No artifact rejection algorithm was applied and all EEG samples were kept for analysis. It is worth noting, however, that after a visual a posteriori check of the samples we found no evidence of eye/muscular artifacts that could have contaminated one condition differently from the other.

III. Experimental Results

For each of the two subjects, data from the calibration sessions performed on ground were split in two groups of five consecutive sessions. The first one, training set, was used to select the features and build a classifier. The performance of this classifier was tested on the second group, testing set, to have a baseline against which to compare the subjects' performance during the parabolic flights. Regarding the data from the parabolic flight, we split it in three groups of five consecutive parabolas. Then, we built a classifier for each group and type of gravity condition, which was tested on the next group. Final performance for each gravity condition is the average for the three groups, which yields a more robust estimation of the BMI performance since we are always testing it on new data recorded on later parabolas than those used for building the classifiers. This procedure is the same for both aspects of the BMI, namely the mental commands and the ErrP.

Relevant features, selected on the training set of the calibration sessions, are kept fixed for the parabolic flight sessions. For the recognition of mental commands, the relevant features are electrodes {C1, C3, C5, CP3, CP5, C2, C4, C6, CP4, CP6} and frequencies {14, 16} Hz for subject 1, and electrodes {FC3, C1, C3, C4} and frequencies {12, 14} Hz for subject 2. These features are in agreement with previous studies where sensorimotor rhythms over the two hemispheres have allowed operating a BMI (Pfurtscheller and Neuper, 2001). Interestingly, the relevant features for ErrP detection are similar for both subjects, namely electrodes FCz and Cz, in accordance with our previous experiments. This is also in agreement with all neurophysiological evidence that ErrP has a centro-frontal focus along the midline (Falkenstein *et al.*, 2000).

A. Classification of Mental Commands

Although the overall task for the subjects was to reach the target at the bottom of the pyramid, here we analyze the classification accuracy at the level of each single EEG sample. This is a much harder task, but yields a better picture of the short-time performance and stability of subjects during parabolic flights. Task-level performance is, in general, better than single-sample performance (provided the latter is above chance level), as each step taken by the balloon is a combination of the outputs of the classifier to several consecutive samples. Also, achieving the task only requires getting closer to the target than to the opposite corner. Thus, correct performance at the task level can accommodate errors at the sample level.

Performance is above chance level for all gravity conditions (or phases) for both subjects, with a global accuracy in between 72 and 79% (Table I). Despite the stress, noise, and novelty of parabolic flight, performance during the flight does not degrade much with respect to ground (our baseline) for subject 1 and is

TABLE I
Percentages (Mean and Standard Deviation) of Correctly Classified Single Samples for the Two Experimental Subjects

	Left arm (%)		Words association (%)		Accuracy (%)	
	Subject 1	Subject 2	Subject 1	Subject 2	Subject 1	Subject 2
1*g* (ground)	84.5	73.5	73.7	73.6	79.1 ± 7.6	73.6 ± 0.1
1*g* (flight)	74.8 ± 12.5	76.2 ± 4.2	69.5 ± 1.5	77.6 ± 2.9	72.2 ± 3.7	76.9 ± 1.0
2*g*	77.6 ± 13.0	77.4 ± 8.8	75.8 ± 6.9	80.0 ± 2.8	76.7 ± 1.3	78.7 ± 1.8
0*g*	81.4 ± 3.6	74.0 ± 1.4	62.8 ± 1.9	74.1 ± 0.7	72.1 ± 13.2	74.1 ± 0.1
2*g*	89.7 ± 0.5	78.2 ± 7.4	68.1 ± 13.9	79.3 ± 1.0	78.9 ± 15.3	78.8 ± 0.8
1*g* (flight)	88.4 ± 1.3	76.0 ± 2.4	57.4 ± 3.1	81.3 ± 12.9	72.9 ± 21.9	78.7 ± 3.7
Average	82.7 ± 5.9	75.9 ± 1.8	67.9 ± 6.9	77.7 ± 3.2	75.3 ± 3.3	76.8 ± 1.2

even better for subject 2. While subject 2 achieves a well-balanced accuracy among both mental tasks, subject 1 has a bias toward "left."

The stability of the EEG patterns during the different gravity conditions of the flight (and with respect to ground) is a key requirement for a successful and reliable BMI in space applications. To check it, we have run the feature selection algorithms to identify the relevant features characterizing each gravity condition. Remarkably, the relevant features, frequencies and electrodes, are very similar for all conditions (Fig. 2 for subject 1). Indeed, the most relevant frequencies are 14 and 16 Hz, whereas the most relevant electrodes are located around C3 and C4. Subject 2 also exhibits a high stability of relevant features for all conditions.

B. Recognition of Error-Related Potentials

ErrP are similar for both subjects and, on average, above 80% for both error and correct steps (Tables II and III for subjects 1 and 2, respectively). These recognition rates are similar to the performances of all subjects we have worked with until now (Ferrez, 2007). The benefit of integrating ErrP detection into a BMI becomes obvious since it always improves its bit-rate—that is, how many correct bits it can communicated per step—for any gravity condition (Tables II and III). On average, ErrP detection doubles the bit-rate of the BMI for both subjects (see Ferrez, 2007 for bit-rate computation of a BMI).

TABLE II
Percentages (Mean and Standard Deviation) of Correctly Classified Error Samples and Correct Samples, Global Accuracy of the BMI (from Table I), Bit-rate of the BMI, and Increase in Performance Introduced by ErrP Detection

				Bit-rate		
	Error (%)	Correct (%)	BMI (%)	No ErrP	ErrP	Increase (%)
1g (ground)	82.8	78.9	79.1	0.260	0.459	76
1g (flight)	85.0 ± 7.1	91.7 ± 7.1	72.2	0.147	0.475	223
2g	90.6 ± 13.3	77.1 ± 3.0	76.7	0.217	0.477	120
0g	90.6 ± 2.4	82.7 ± 1.7	72.1	0.146	0.466	219
2g	70.3 ± 9.4	89.6 ± 9.5	78.9	0.256	0.456	78
1g (flight)	77.5 ± 3.5	83.4 ± 7.4	72.9	0.157	0.374	138
Average	82.8 ± 7.9	83.9 ± 5.8	75.3	0.193	0.445	130

Performances for subject 1 over all gravity conditions.

TABLE III
PERCENTAGES (MEAN AND STANDARD DEVIATION) OF CORRECTLY CLASSIFIED ERROR SAMPLES AND CORRECT SAMPLES, GLOBAL ACCURACY OF THE BMI (FROM TABLE I), BIT-RATE OF THE BMI, AND INCREASE IN PERFORMANCE INTRODUCED BY ERRP DETECTION

				Bit-rate		
	Error (%)	Correct (%)	BMI (%)	no ErrP	ErrP	Increase (%)
1*g* (ground)	84.6	84.4	73.6	0.167	0.441	163
1*g* (flight)	85.0 ± 7.1	87.1 ± 1.3	76.9	0.220	0.505	129
2*g*	89.9 ± 1.4	88.4 ± 1.8	78.7	0.253	0.578	128
0*g*	86.4 ± 6.4	77.6 ± 2.4	74.1	0.175	0.416	138
2*g*	81.2 ± 1.7	68.8 ± 1.1	78.8	0.255	0.372	46
1*g* (flight)	80.8 ± 11.5	86.6 ± 5.3	78.7	0.253	0.496	96
Average	84.7 ± 3.4	82.2 ± 7.6	76.8	0.218	0.467	113

Performances for subject 2 over all gravity conditions.

IV. Discussion

The results of the December 2007 ESA campaign show that it is possible for a subject with some prior BCI experience to achieve stable performances in normal gravity as well as in microgravity and hypergravity, and hence demonstrate the feasibility of operating noninvasive BMI in space. Both subjects show encouraging performance despite their little experience in microgravity. On average, both of them reached 75% of global accuracy for the recognition of two mental commands and more than 80% of correct classification for ErrP. Although the BMI performance does not achieve the results of experiments run on ground, they are still satisfactory considering the various sensorial stress experienced during parabolic flights. As previous BMI research shows, these performances can be improved with further training and experience of the subjects in the use of BMI during parabolic flights. These results, and hypothesis, need to be confirmed with further experiments in future parabolic flight campaigns that should involve more subjects sufficiently trained previously on ground.

Acknowledgments

The experiments were performed onboard the A 300 Zero-G of Novespace during the 47th ESA parabolic flight campaign (December 2007, Bordeux). We are grateful to the organizers of this campaign for acceptance on the flight and great support throughout the entire campaign.

References

Carpi, F., and De Rossi, D. (2006). Non invasive brain-machine interfaces. European Space Agency, Advanced Concepts Team, Ariadna Final Report (05-6402).

Falkenstein, M., Hoormann, J., Christ, S., and Hohnsbein, J. (2000). ERP components on reaction errors and their functional significance: A tutorial. *Biol. Psychol.* **51,** 87–107.

Ferrez, P. W. (2007). Error-related EEG potentials in brain-computer interfaces. Ph.D. Thesis, École Polytechnique Fédérale de Lausanne.

Ferrez, P. W., and Millán, J. d. R. (2008a). Error-related EEG potentials generated during simulated brain-computer interaction. *IEEE Trans. Biomed. Eng.* **55,** 923–929.

Ferrez, P. W., and Millán, J. d. R. (2008b). Simultaneous real-time detection of motor imagery and error-related potentials for improved BCI accuracy. *In* "4th International Brain-Computer Interface Workshop and Training Course," Graz, Austria, September 18–21.

Millán, J. d. 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,** 1026–1033.

Millán, J. d. R., Ferrez, P. W., and Buttfield, A. (2006). Non invasive brain-machine interfaces European Space Agency, Advanced Concepts Team, Ariadna Final Report (05-6402).

Millán, J. d. R., Ferrez, P. W., Galán, F., Lew, E., and Chavarriaga, R. (2008). Non-invasive brain-machine interaction. *Intern. J. Pattern Recognit. Artif. Intell.* **22,** 959–972.

Pfurtscheller, G., and Neuper, C. (2001). Motor imagery and direct brain-computer communication. *Proc. IEEE* **89,** 1123–1134.

Pletser, V., and Quadens, O. (2003). Degraded EEG response of the human brain in function of gravity levels by the method of chaotic attractor. *Acta Astronaut.* **52,** 581–589.

Tonet, O., Tecchio, F., Sepulveda, F., Citi, L., Rossini, P. M., Marinelli, M., Tombini, M., Laschi, C., and Dario, P. (2006). Non invasive brain-machine interfaces. European Space Agency, Advanced Concepts Team, Ariadna Final Report (05-6402).