Single-Trial Classification of Feedback Potentials within Neurofeedback Training with an EEG Brain-Computer Interface

Eduardo López-Larraz, Iñaki Iturrate, Carlos Escolano, Isabel García, Luis Montesano, Javier Minguez

Abstract— **Neurofeedback therapies are an emerging technique used to treat neuropsychological disorders and to enhance cognitive performance. The feedback stimuli presented during the therapy are a key factor, serving as guidance throughout the entire learning process of the brain rhythms. Online decoding of these stimuli could be of great value to measure the compliance and adherence of the subject to the training. This paper describes the modeling and classification of performance feedback potentials with a Brain-Computer Interface (BCI), under a real neurofeedback training with five subjects. LDA and SVM classification techniques are compared and are both able to provide an average performance of approximately 80%.**

I. INTRODUCTION

Neurofeedback is a biofeedback modality administered with the objective of providing the users with operant control of specific brain rhythms. The basic principle of neurofeedback consists of measuring the brain activity, decoding or identifying the brain patterns of interest, and then providing positive or negative feedback stimuli to the user depending on the desired working levels of these patterns. The stimuli can be any of the sensory modalities, such as visual, auditive or somatosensory; and presentation can be continuous (during the execution of the training) or discrete at specific times (usually as a global training evaluation). The EEG is the most accepted brain recording technique for neurofeedback, as it is relatively cheap, portable, easy to use and has a low set-up cost. EEG Neurofeedback has been successfully applied in two main areas: (i) cognitive enhancement (e.g., in the improvement of the attention and/or working memory [1] or cognitive tasks [2]); and (ii) neurotherapy (e.g., treatment of neurological and psychological disorders, such as attention deficit/hyperactivity disorder [3], and epilepsy [4], among others).

Improvement in the control of specific brain patterns is achieved through a learning process via feedback modulation. As the human processing of feedback inputs is a paramount element in the learning process [5], the characterization of the brain potentials involved in this process is of great value to understand neurofeedback. Additionally, the online detection of these brain patterns could provide a metric of adherence and compliance of the subject to the process. When the feedback potentials are adequately elicited, it indicates that the subject is engaged in the task and correctly perceives the task and his/her own performance.

The brain potentials behind the presentation of feedback in a task were studied in [6] and have been recently detected online in well-controlled time estimation tasks [7], [8]. This paper goes a step further by analyzing feedback potentials in a real neurofeedback training protocol, and by demonstrating several strategies to detect feedback potentials with a brain-computer interface. In opposition to studies that design specific protocols to study brain potentials (e.g., [7] for feedback potentials), the data used herein was obtained from a 10-day neurofeedback training for cognitive enhancement. This entails important differences and difficulties. Firstly, there are non-stationarities between the different sessions [9]. Secondly, the feedback potentials are fully dependent on the user performance within the exercise. As a result, the number of available examples can be scarce and it is not possible to balance the presentation of positive and negative feedbacks, which could result in an unpredictable and unbalanced number of examples of each type of potential.

II. EXPERIMENTATION PROTOCOL

A. Neurofeedback Training Procedure

The objective of the work was to study feedback brain potentials obtained during a real neurofeedback training. The training focused on the enhancement of sensory-motor rhythms (SMR) while simultaneously maintaining Theta and low Beta bands at original levels in central sensory-motor areas. This type of neurofeedback training has been reported to produce an impact in memory and attention of healthy subjects [1].

Five participants took part in the neurofeedback training, two females and three males in the range 25.8 ± 2.3 years of age. Each participant completed ten sessions in two consecutive weeks (one session per day, from Monday to Friday). Each session lasted approximately 30 minutes, divided into three phases of ten minutes with several relaxing minutes between phases. Each phase consisted of a one-minute calibration period where a baseline was obtained with the mean values of the three bands, and of four feedback trials. Each trial followed the structure of Figure 1 with two types of feedback: a visual progressive feedback, and a visual plus auditory evaluation feedback. The progressive feedback was computed as a function of the progressive power levels in the correspondent bands. The result was displayed on a computer screen with a green or red color bar, depending on whether the bands were at the correct levels. The bar size was dependent on the minimum power deviation from the
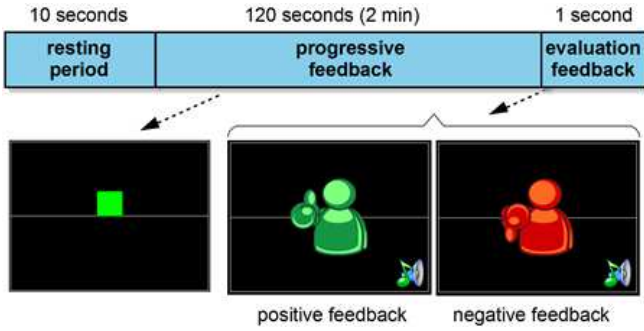
Fig. 1. Trial structure of the neurofeedback protocol. After a resting period, progressive feedback (a moving bar) was continuously shown. The evaluation feedback was provided at the end of the trial. The evaluation feedback consisted of a correct/incorrect audio-visual stimulus. The ERPs should appear in the time window [0 - 1000] ms after this cue.

baseline. At the end of each trial, an audio-visual evaluation feedback indicating the successfulness of the trial block was presented for 1 second. It was positive if the time with the bands at correct levels was more than 50% of the trial duration, and negative otherwise. The evaluation feedback showed a positive or negative image (Figure 1) and played an harmonious jingle (positive feedback) or low tone (negative feedback) for 300 ms. The task of the participants in the experiment was to establish an individual mental strategy to maintain the feedback bar at the correct levels.

B. EEG Recordings

EEG signals were measured from 16 active electrodes, placed at FP1, FP2, F3, Fz, F4, FCz, C3, C1, Cz, C2, C4, CPz, P3, Pz, P4 and Oz (according to the international 10/10 system). The ground and reference electrodes were placed on FPz and on the left earlobe, respectively. The signals were amplified using a commercial gTec system. The EEG was digitized at a sampling frequency of 256Hz, power-line notch-filtered at 50Hz, and bandpass-filtered between 0.5 and 60Hz. EEG acquisition, signal processing, and feedback presentation were developed using a self-made software written in C++ and running on a Windows machine. The EEG signal power of Theta (4-7Hz), SMR (12-15Hz) and low Beta (18-22Hz) bands was calculated through a sliding FFT algorithm, including a Hanning window (2048 points) with 30 ms overlapping.

At the end of each trial, the audio-visual evaluation feedback was presented. It was supposed to elicit the event-related potentials (i.e., feedback potentials) analyzed in this paper. In summary, each session contained 12 evaluation feedback responses (3 phases x 4 trials/phase) and a total of 120 feedbacks were accumulated per participant. In order to analyze the brain potentials, the acquired EEG was bandpass-filtered between 0.5 and 10 Hz and a Common Average Reference (CAR) filter was applied to remove any background activity present on the EEG.

III. RESULTS

This section presents a characterization of the feedback potentials obtained during the training, the feature extraction process, and the classification results.

A. Characterization of Feedback Potentials

The grand averages of the potentials were computed for both conditions (positive and negative feedback) for all participants and all neurofeedback sessions. Figure 2a displays the average potentials and the difference between negative and positive responses for the FCz electrode. The difference shows three main well-defined peaks, and a fourth peak with a sparse ending. These results agree with those obtained in a previous study of visual feedback potentials during a time-estimation task [7] (with the three first peaks roughly at the same time instants but with a notably higher amplitude). This higher amplitude could be explained by the fact that in the present protocol, the saliency of the stimuli was higher as the subjects did not know when the feedback stimuli were going to appear. In the aforementioned protocol [7], the feedback stimuli were presented more frequently and presentation time was easily anticipated [10].

The Dynamo source localization technique [11] was used to analyze the sources of feedback potential activity. Although source localization techniques are not accurate with a low number of electrodes [12], the results suggested that Brodmann Areas 24, 32 (ACC) and 23, 31 (PCC) were systematically activated. These areas are usually associated with error processing [5], [10].

B. Feature Analysis and Extraction

The r^2 analysis was performed to identify the temporal and spatial areas with most statistical differences between positive and negative potentials [13]. Figure 2b shows the r^2 coefficient for each channel during one second after feedback presentation, averaged for the five participants. The results showed that the most significant information belonged to fronto-central channels at the time window [200-600] ms. Thus, this information was used to select the features used throughout the paper: channels FCz, C1, C2 and Cz during time window [200-600] ms were selected and subsampled to 64Hz. The data of each channel was concatenated in a single feature vector of dimension 104 for each response.

One important aspect of this neurofeedback setting was that non-stationarities were present in the EEG, due to the fact that it was carried out in several sessions. The feature vectors were used as a reference to analyze this issue. Then the feedback potentials were grouped according to the acquisition day, totaling ten groups. For each participant, a r^2 analysis was carried out for the feature vectors belonging to each training session, in order to study how this measurement changed throughout days. If one session did not present examples of one class, the r^2 coefficients were assigned a null value. Figure 2c depicts the evolution of r^2 values in the selected features throughout the 10 sessions for participant 2. Note that some features remained stable throughout sessions, but others varied. This indicates that the representative features differentiating classes in one session might not be the same in other sessions. Also note that r^2 values in Fig. 2b are an average of subjects and sessions, while Fig. 2c shows coefficients for each session of a single subject. As the most representative features can vary between subjects
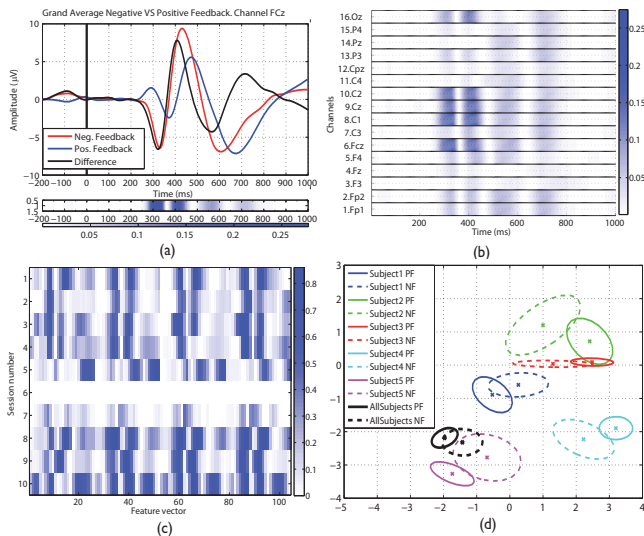
Fig. 2. (a) Time-locked grand averages for channel FCz, averaged for all participants. (b) r^2 Coefficients computed for all channels, averaged for all participants. (c) Evolution of r^2 coefficients computed for selected feature vectors throughout sessions, corresponding to Participant 2. (d) 2D distributions of the two most representative features to discriminate between positive feedback (PF) and negative feedback (NF), computed for each participant separately and for the pooled data.

and sessions, their average value appears to be notably lower than for a single subject and session.

Another important aspect was the variability of features with the subjects. Principal Component Analysis (PCA) was applied to the feature vectors of each participant and to the pooled dataset of all participants to decorrelate the features. Then, a second r^2 analysis was performed to identify the two decorrelated features with most discriminant information between classes. Finally, the distributions of both features were displayed for positive and negative feedback responses, separately for each subject as well as for the five subjects grouped (Figure 2d). The figure suggested that the distributions of both types of potentials for each subject were separated. Additionally, the results showed that the distributions were different for each subject, and that the subjects seemed to be divided into two clusters. Features of subjects 2, 3 and 4 presented positive responses rightwards, while subjects 1 and 5 and the pooled dataset presented responses leftwards. Thus, using information of certain subjects to model the responses of other subjects could be unproductive, as the relative position of distributions can be interchanged.

C. Calibration and Classification

The last step was to analyze the classification possibilities of the potentials. Note that there was a great difficulty imposed by the fact that the potentials were acquired during a neurofeedback training, which was not explicitly designed as a standard calibration session in BCI systems. This resulted in a low and unbalanced number of examples in addition to the inherent non-stationarities among sessions.

The feature vectors defined in the previous section were normalized in the range [0-1] and used for classification. Two different classifiers were compared: Linear Discriminant

Analysis (LDA) and Support Vector Machine (SVM). Both are representative examples of linear and non-linear classifiers (commonly used for EEG classification [14]). Moreover, the classifiers were compared in two different contexts: (i) an offline analysis of the classification accuracy that could be obtained for each subject by cross-validation; and (ii) a more realistic setup of supervised online classifiers, which took into account previous information of the subject or other subjects.

TABLE I
CLASSIFICATION RATES

	LDA			SVM		
	Pos. F.	Neg. F.	Avg. Responses	Pos. F.	Neg. F.	Avg. Responses
P1	72.46	60.78	67.50	91.30	62.75	79.17
P2	74.60	68.42	71.67	92.06	80.70	86.67
P3	61.76	65.12	64.17	41.18	89.53	75.83
P4	60.98	65.79	62.50	98.78	60.53	86.67
P5	61.76	51.92	57.50	97.06	61.54	81.67
Average	66.31	62.41	**64.67**	84.08	71.01	**82.00**

The offline performance was calculated using a leave-one-out cross-validation strategy, providing a measure of the maximum performance that could be obtained for each subject individually. Table I presents the mean accuracy for positive and negative responses as well as the average for both types of responses. Rows $P1$ to $P5$ show the results obtained for each of the five participants for each condition, while the $Average$ row corresponds to the mean of them. The column $Avg.Responses$ represents the global performance for each subject, i.e., the percentage of correctly detected feedback potentials. Note that, as the number of examples for each class was not balanced, this value did not represent the mean between positive and negative accuracies. Results for SVM show that all subjects presented an average classification accuracy higher than 75%, and the average result for the five participants reached 82%. However in four subjects, one of the classes provided a poor classification performance, probably due to the low number of examples available. LDA results were notably lower in practically all cases, indicating that the non-linear classifier SVM worked better to model this classification problem with all the available data.

The complementary analysis was the supervised online classification. Sequential supervised classifiers were trained to study the behavior of LDA and SVM when dealing with a low number of examples and when training with several subjects to classify a different one. Three different strategies are compared in this subsection:

a) Incremental Strategy: The data obtained from sessions $[1 \ldots i-1]$ was used to train the classifier, whereas the test set was formed by the i-th session ($i \in [2 \ldots 9]$). This strategy was performed for each subject separately. The results for the SVM classifier are shown in Fig. 3a. Several observations can be made. Firstly, the first session did not have previous data to train the classifier. Secondly, offline SVM results provided higher accuracies than incremental strategy in almost all sessions. Thirdly, the LDA classifier required almost all sessions to estimate the empirical covariance matrix and consequently is not reported here[1].

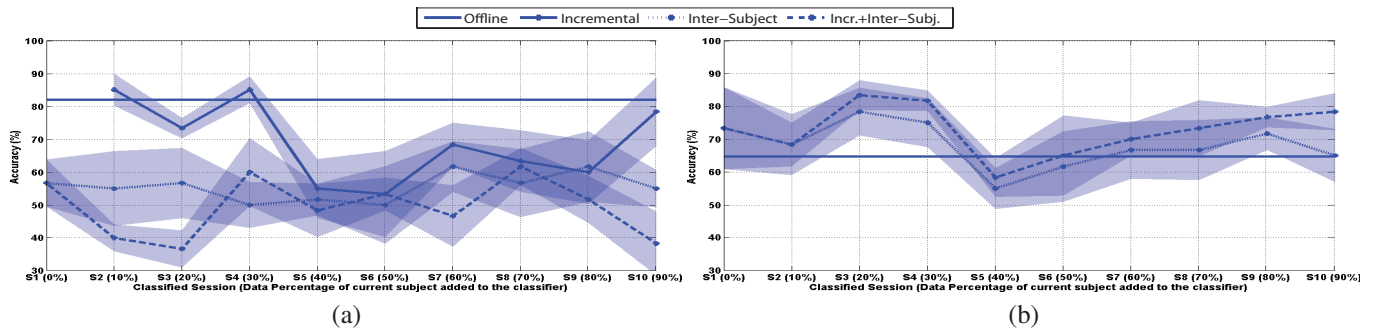[1]It would be possible, however, to use regularization.

Fig. 3. Mean performance for the three classification methods averaged for all the participants using SVM (a) and LDA (b). The shadowing represents the standard error. The x-axis indicates the session classified in each point, and the percentage of the total sessions added to the incremental classifiers is indicated in brackets. The y-axis shows the accuracy obtained when classifying each session. Horizontal lines in both plots represent the maximum accuracy obtained previously through leave-one-out cross-validation. The solid line in (a) shows the average performance of the per-subject incremental classifiers. Dotted lines show the performance of the classifiers built with data from the remaining four subjects. Dashed lines show the performance of the classifiers built with data from four subjects, retrained each session with data of the fifth subject.

b) Inter-Subject Strategy: In order to address the short-comings of the previous strategy, the classifier was initially trained with information of four subjects. Then, the sessions of the fifth subject were tested sequentially using that classifier. The five possible combinations of subjects were tested. The average performance for all subjects is plotted in Figure 3a-b (dotted lines). The results show better than random accuracies since first session. Thus, it is possible to use this classifier since the beginning of the neurofeedback training.

SVM provided lower accuracies than the incremental strategy for the remaining sessions. LDA showed higher accuracies than SVM. Moreover, it exceeded the offline maximum accuracy obtained previously in almost all sessions.

c) Inter-Subject plus Incremental Strategy: Here the two previous approaches were combined. The classifier was initially trained using the examples from four subjects and then retrained after each session with the fifth subject (dashed lines in Fig. 3a-b). This strategy implemented using SVM did not produce any improvements with respect to the previous strategy. In opposition, LDA improved previous strategy in all sessions from third one forwards. Furthermore, it showed a continuous growing trend that began on fifth session.

In summary, the best results for SVM corresponded to the incremental strategy, while the performance degraded almost to randomness for the second and third strategies. LDA worked better than SVM in the second and third strategies, in opposition to the results obtained for the offline classification. Moreover, sequentially-retrained LDA classified almost all sessions with a better performance than the maximum offline accuracy.

IV. Conclusions and future work

This paper described the single-trial classification of feedback potentials within an EEG-based neurofeedback training protocol. In opposition to standard classification results with BCI, these feedback potentials were acquired during a real neurofeedback training (not in a calibration session) and during different sessions. This resulted in an unbalanced low number of examples that changed between sessions.

Different classification methods were proposed. The offline cross-validation achieved a considerably stable behavior and a performance near 80%. Additionally, this performance was also achieved with a supervised LDA-based incremental

classifier. Both results suggested that there is a possibility of using an EEG brain-computer interface to monitor these potentials during neurofeedback training.

References

[1] D. Vernon, "Can Neurofeedback Training Enhance Performance? An Evaluation of the Evidence with Implications for Future Research," *App. Psychophysiology and Biofeedback*, vol. 30, pp. 347–364, 2005.

[2] B. Zoefel, R. J. Huster, and C.S. Herrmann, "Neurofeedback training of the upper alpha frequency band in EEG improves cognitive performance," *NeuroImage*, vol. 54, no. 2, pp. 1427–1431, Jan. 2011.

[3] J.F. Lubar, M.O. Swartwood, J.N. Swartwood, and P.H. O'Donnell, "Evaluation of the effectiveness of EEG neurofeedback training for ADHD in a clinical setting as measured by changes in T.O.V.A. scores, behavioral ratings, and WISC-R performance," *Applied Psychophysiology and Biofeedback*, vol. 20, pp. 83–99, 1995.

[4] M.B. Sterman, "Basic concepts and clinical findings in the treatment of seizure disorders with EEG operant conditioning," *Clinical Electroencephalography*, vol. 31, no. 1, pp. 45 – 55, 2000.

[5] S. Nieuwenhuis, C.B. Holroyd, N. Mol, and M.G.H. Coles, "Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance," *Neuroscience & Biobehavioral Reviews*, vol. 28, no. 4, pp. 441 – 448, 2004.

[6] W.H.R. Miltner, C.H. Braun, and M.G.H. Coles, "Event-Related Brain Potentials Following Incorrect Feedback in a Time-Estimation Task: Evidence for a Generic Neural System for Error Detection," *Journal of Cognitive Neuroscience*, vol. 9, no. 6, pp. 788–798, 1997.

[7] E. López-Larraz, I. Iturrate, L. Montesano, and J. Minguez, "Real-Time Recognition of Feedback Error-Related Potentials during a Time-Estimation Task," in *32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, 2010.

[8] E. López-Larraz, M. Creatura, I. Iturrate, L. Montesano, and J. Minguez, "EEG Single-Trial Classification of Visual, Auditive and Vibratory Feedback Potentials in Brain-Computer Interfaces," in *33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, 2011.

[9] J.d.R. Millán, "On the need for On-line learning in Brain-Computer Interfaces," in *Int. Joint Conference on Neural Networks*, 2004.

[10] M. Falkenstein, J. Hoormann, S. Christ, and J. Hohnsbein, "ERP components on reaction errors and their functional significance: A tutorial," *Biological Psychology*, vol. 51, pp. 87–107, 2000.

[11] J.M. Antelis and J. Minguez, "DYNAMO: Dynamic multi-model source localization method for EEG and/or MEG," in *32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)*, 2010.

[12] C.M. Michel, M.M. Murray, G. Lantz, S. Gonzalez, L. Spinelli, and R.G. de Peralta, "EEG source imaging," *Clinical Neurophysiology*, vol. 115, no. 10, pp. 2195–2222, 2004.

[13] J.R. Wolpaw, N. Birbaumer, D.J. McFarland, G. Pfurtscheller, and T.M. Vaughan, "Brain Computer Interfaces for Communication and Control," *Clinical neurophysiology*, vol. 113, no. 6, pp. 767–791, 2002.

[14] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain computer interfaces," *Journal of Neural Engineering*, vol. 4, June 2007.