



# A Sparse EEG-Informed fMRI Model for Hybrid EEG-fMRI Neurofeedback Prediction

Claire Cury<sup>1,2\*</sup>, Pierre Maurel<sup>1</sup>, Rémi Gribonval<sup>2</sup> and Christian Barillot<sup>1</sup>

<sup>1</sup> University of Rennes, CNRS, Inria, Inserm, IRISA UMR 6074, Empenn Team ERL U 1228, Rennes, France, <sup>2</sup> University of Rennes, CNRS, Inria, IRISA UMR 6074, PANAMA Team, Rennes, France

## OPEN ACCESS

### Edited by:

Bertrand Thirion,  
Institut National de Recherche en  
Informatique et en Automatique  
(INRIA), France

### Reviewed by:

Dustin Scheinost,  
Yale University, United States  
Masaya Misaki,  
Laureate Institute for Brain Research,  
United States

### \*Correspondence:

Claire Cury  
claire.cury.pro@gmail.com

### Specialty section:

This article was submitted to  
Brain Imaging Methods,  
a section of the journal  
Frontiers in Neuroscience

**Received:** 15 July 2019

**Accepted:** 30 December 2019

**Published:** 31 January 2020

### Citation:

Cury C, Maurel P, Gribonval R and  
Barillot C (2020) A Sparse  
EEG-Informed fMRI Model for Hybrid  
EEG-fMRI Neurofeedback Prediction.  
*Front. Neurosci.* 13:1451.  
doi: 10.3389/fnins.2019.01451

Measures of brain activity through functional magnetic resonance imaging (fMRI) or electroencephalography (EEG), two complementary modalities, are ground solutions in the context of neurofeedback (NF) mechanisms for brain rehabilitation protocols. While NF-EEG (in which real-time neurofeedback scores are computed from EEG signals) has been explored for a very long time, NF-fMRI (in which real-time neurofeedback scores are computed from fMRI signals) appeared more recently and provides more robust results and more specific brain training. Using fMRI and EEG simultaneously for bi-modal neurofeedback sessions (NF-EEG-fMRI, in which real-time neurofeedback scores are computed from fMRI and EEG) is very promising for the design of brain rehabilitation protocols. However, fMRI is cumbersome and more exhausting for patients. The original contribution of this paper concerns the prediction of bi-modal NF scores from EEG recordings only, using a training phase where EEG signals as well as the NF-EEG and NF-fMRI scores are available. We propose a sparse regression model able to exploit EEG only to predict NF-fMRI or NF-EEG-fMRI in motor imagery tasks. We compared different NF-predictors stemming from the proposed model. We showed that predicting NF-fMRI scores from EEG signals adds information to NF-EEG scores and significantly improves the correlation with bi-modal NF sessions compared to classical NF-EEG scores.

**Keywords:** optimization, EEG, sparsity, machine learning, neurofeedback, EEG-fMRI

## 1. INTRODUCTION

Neurofeedback approaches (NF) provide real-time feedback to a subject about his or her brain activity and help him or her perform a given task (Hammond, 2011; Sulzer et al., 2013). The estimation of neurofeedback information is done through online brain functional feature extraction to provide this real-time feedback to the subject. NF appears to be an interesting approach for clinical purposes, for example, in the context of rehabilitation and psychiatric disorders (Birbaumer et al., 2009; Sulzer et al., 2013; Wang et al., 2017). Functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) are the most used non-invasive functional brain imaging modalities in neurofeedback. EEG measures the electrical activity of the brain through electrodes located on the scalp. EEG has an excellent temporal resolution (milliseconds) but a limited spatial resolution (centimeters), implying a lack of specificity. Furthermore, source localization in EEG is a well-known ill-posed inverse problem (Grech et al., 2008).

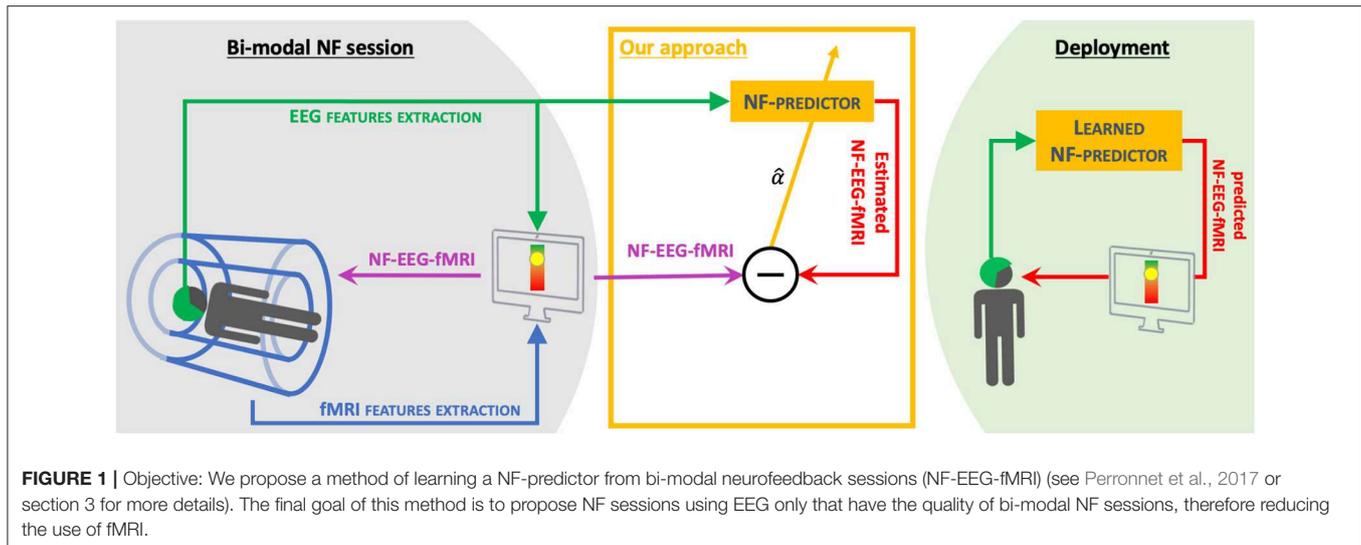
On the other hand, blood oxygenation level-dependent (BOLD) fMRI measures a delayed hemodynamic response to neural activity with a good spatial resolution and a temporal resolution of 1 or 2 s depending on the sequence used. Therefore, fMRI is more specific than EEG, making fMRI a suitable modality for neurofeedback (NF-fMRI) (Thibault et al., 2018). However, the use of the MRI scanner is costly, exhausting for patients (since staying perfectly still when suffering is challenging), and time-consuming, and hence NF-fMRI sessions cannot be repeated an excessive number of times for the same subject or patient.

During the past few years, simultaneous EEG-fMRI recording has been used to understand the links between EEG and fMRI in different states of brain activity and has received recognition as a promising multi-modal measurement of brain activity (Abreu et al., 2018; Perronnet et al., 2018). However, this bi-modal acquisition is cumbersome for subjects or patients due to the use of the fMRI scanner. The methodology to extract information from fMRI with EEG has been also intensively investigated (some methods involved in the process are reviewed in Abreu et al., 2018). Indeed, both modalities are sensitive to different aspects of brain activity with different speeds. EEG provides a direct measure of the changes in electrical potential occurring in the brain in real time, while fMRI indirectly estimates brain activity by measuring changes in the BOLD signal, reflecting neurovascular activity, which occurs, in general, a few seconds after a neural event (Friston et al., 1994; Logothetis et al., 2001). Several studies have investigated correlations between EEG signal and BOLD activity in specific and simple tasks (de Munck et al., 2007; Goncalves et al., 2008; Scheeringa et al., 2011; Engell et al., 2012; Magri et al., 2012) and have found different relationships of certain frequency bands of the EEG signal. All those studies reveal the existence of a link between EEG and fMRI, but this relationship varies considerably with the task, location in the brain, and frequency bands considered.

In the literature, the term EEG-informed fMRI refers to methods extracting features from EEG signals in order to derive a predictor of the associated BOLD signal in the region of interest under study. A recent review (Abreu et al., 2018) gives a good overview of the principal EEG-informed fMRI methods and their limitations. Different strategies have been investigated, depending on the type of activity under study (epilepsy, resting state, open/closed eyes, relaxation): either selecting one channel of interest or using multiple channels before extracting features of interest. For example, in Formaggio et al. (2011) and Leite et al. (2013), the authors used a temporal independent component analysis to select the channel that best reflected epileptic seizures. In Schwab et al. (2015), the authors used a spatial, spectral, and temporal decomposition of the EEG signals to map EEG onto BOLD signal changes in the thalamus. Using a more symmetrical approach, Noorzadeh et al. (2017) proposed a method for the estimation of brain source activation, improving the spatio-temporal resolution compared to EEG or BOLD fMRI only. However, in the context of neurofeedback, using simultaneous recording of EEG-fMRI to estimate neurofeedback scores computed in real time from features of both modalities (NF-EEG-fMRI) is a recent application that was first introduced and its feasibility demonstrated by Zotev et al. (2014), Mano

et al. (2017), and Perronnet et al. (2017). The recent methodology synchronizing both signals for real-time neurofeedback (Mano et al., 2017) allows the construction of a new kind of data named NF-EEG-fMRI data, such as the dataset presented by Perronnet et al. (2017), which we used for the present study. Furthermore, it has been shown in Perronnet et al. (2017) that the quality of a neurofeedback session is improved when using both modalities simultaneously in NF-EEG-fMRI sessions. Thus, being able to reproduce an NF-EEG-fMRI session in real time when using EEG only would reduce the use of fMRI in neurofeedback while increasing the quality of NF-EEG sessions. To export fMRI information outside the scanner, most of the methods intend to predict the fMRI BOLD signal activity in a specific region of interest by learning from an EEG signal recorded simultaneously inside the fMRI scanner. Indeed, the method proposed in Meir-Hasson et al. (2014) uses a ridge regression model with a  $\ell_2$  regularization, based on a time/frequency/delay representation of the EEG signal from a single channel. Results show a good estimation of the BOLD signal in the region of interest, but the use of the fMRI neurofeedback in this study is only to serve the paradigm. The method aims at better targeting the amygdala in NF-EEG sessions.

Our challenge here is to learn EEG activation patterns (see section 2.2) from hybrid (or bi-modal) NF-EEG-fMRI sessions (Perronnet et al., 2017) and improve the correlation with NF-EEG-fMRI of NF scores using the EEG signal only. The motivations of this are multiple. Since we are considering a new kind of data, we want to provide a simple method of characterizing NF-EEG-fMRI in EEG, leading to an understandable model to confirm existing relations between EEG and fMRI in neurofeedback scores or to discover new relationships. Neurofeedback features in fMRI come from BOLD activation in one or more regions of interest. We propose an original alternative to source reconstruction in the context of neurofeedback by taking advantage of a recent and unique dataset of NF-EEG-fMRI that we have access to. Indeed, we intend to predict NF scores directly without dealing with source reconstruction or spatial filtering to estimate the BOLD-fMRI signal first in a specific region of interest, as proposed by a previous study (Noorzadeh et al., 2017). To our knowledge, this prediction of hybrid neurofeedback scores (without source reconstruction) is new and has not yet been explored in the literature. Also, we want the activation pattern to be applicable in real time when using new EEG data. The main objective of this paper (**Figure 1**) is to design a method able to exploit EEG only and predict an NF score of a quality comparable to the NF score that could be achieved with a simultaneous NF-EEG-fMRI session. The approach is based on a machine learning mechanism. During a training phase, both EEG and fMRI are simultaneously acquired and used to compute and synchronize, in real time, NF-EEG and NF-fMRI scores, both being combined into a hybrid NF-EEG-fMRI score (Mano et al., 2017). EEG signals and NF scores are used to learn activation patterns. During the testing phase, the learned NF-predictor (also called the activation pattern) is applied to unseen EEG data, providing simulated NF-EEG-fMRI scores in real time. Sparse regularization is exploited to build a model called the



NF-predictor. The model used for the NF-predictor uses an adapted prior for brain activation patterns, using a mixed norm giving a structured sparsity, to spatially select electrodes and then select the most relevant frequency bands.

In section 2, we present the proposed model and the methods used to solve it. We then detail experiments with our learning model in neurofeedback sessions with a motor imagery task, from which unique data are presented in section 3. Section 4 presents results for a dataset of 17 healthy subjects with three NF sessions of motor imagery each; one is used to learn the model, and the other two are used to test the model. Section 5 provides a discussion of the proposed framework.

## 2. PROBLEM AND METHOD

Considering that, during a learning phase, we have access to reference scores  $y(t)$  and a temporal representation (potentially non-linear) of EEG signals  $X$  (called a design matrix, presented in section 2.1), the approach consists of choosing a vector of parameters  $\alpha$  such that  $y(t) \approx q(X(t), \alpha)$  for all  $t$ , where  $q$  is some parametric function.  $\alpha$  is a matrix matching the size of  $X(t)$ . Here we consider

$$q(X(t), \alpha) := \langle X(t), \alpha \rangle = \sum_{i=1}^E \sum_{j=1}^F X_{i,j}(t) \alpha_{i,j}. \quad (1)$$

To determine this parameter  $\alpha$ , a regularization is used to select an optimal parameter  $\hat{\alpha}$  that fits the training data (see **Figure 2** and section 2.3) while avoiding over-fitting, as detailed in section 2.2.

Only a few brain regions are expected to be activated by a given cognitive task, so the electrode configuration is said to be spatially sparse. However, the frequency bands of each electrode are not necessarily sparse and might even be smooth, depending on the frequency band sampling.

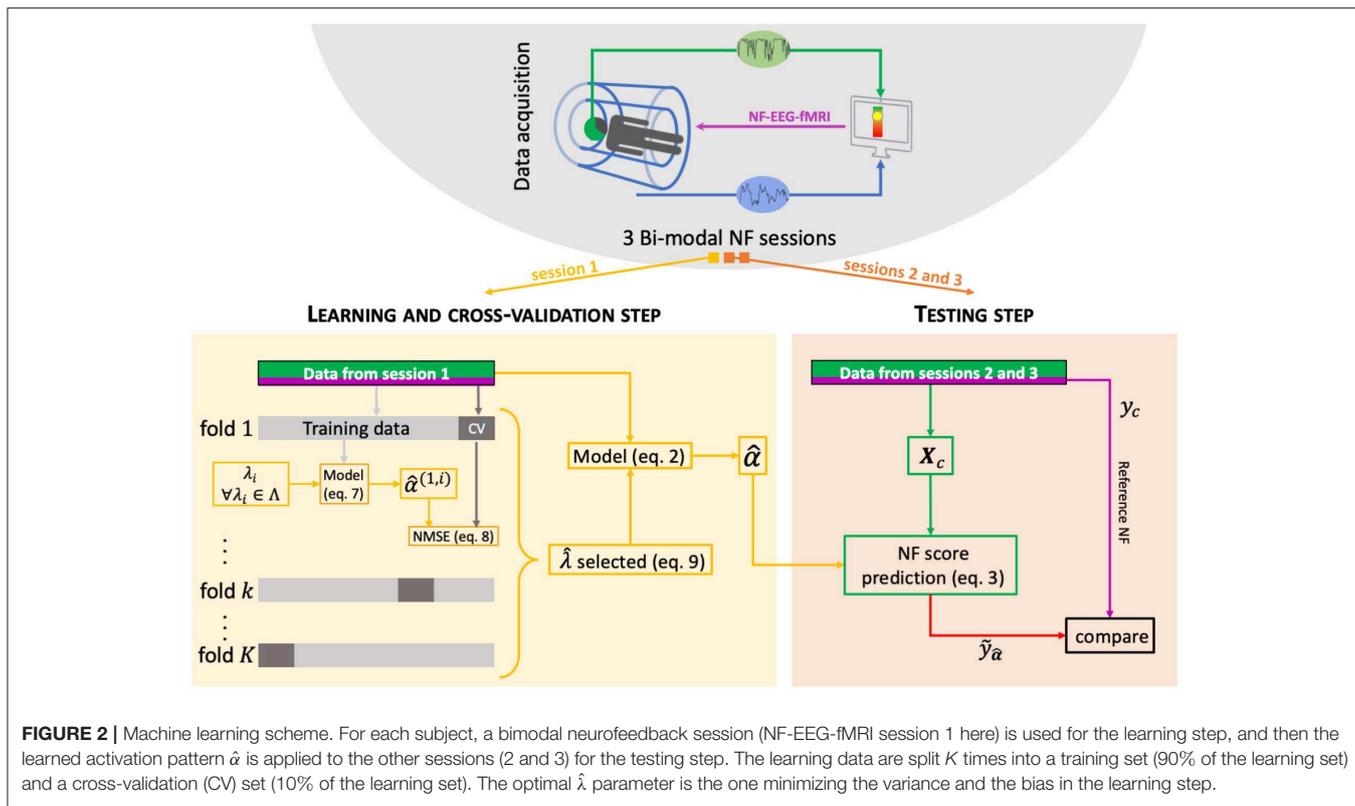
From here, we will use the following notations:

- $y_e(t) \in \mathbb{R}, \forall t \in \{1, \dots, T\}$  are the  $T$  neurofeedback scores estimated from EEG signals (noted  $S_{\text{EEG}} \in \mathbb{R}^{E \times T_{\text{EEG}}}$ ) measured from  $E$  electrodes during  $T_{\text{EEG}}$  samples of time.
- $y_f(t) \in \mathbb{R}, \forall t \in \{1, \dots, T\}$  are the  $T$  neurofeedback scores extracted from the Blood Oxygen Level Dependent imaging (BOLD) signal of functional-MRI acquisitions  $S_{\text{fMRI}} \in \mathbb{R}^{V \times T_{\text{fMRI}}}$ , with  $V$  the number of voxels and  $T_{\text{fMRI}}$  the number of acquired volumes.
- $y_c(t) = y_e(t) + y_f(t) \in \mathbb{R}, \forall t \in \{1, \dots, T\}$ , a combination of both NF scores (more details are provided in section 3).
- $y(t) \in \mathbb{R}, \forall t \in \{1, \dots, T\}$  is a set of neurofeedback scores that can be  $y_e, y_f$  or  $y_c$ .

First, to build our predictor, relevant information needs to be extracted from EEG data and organized to form what we call a design matrix.

### 2.1. Structured Design Matrices From the EEG Signal

The design matrix  $X_0 \in \mathbb{R}^{T \times E \times B}$ , where  $E$  is the number of electrodes and  $B$  the number of frequency bands, contains relevant information extracted from the EEG signal. Each temporal matrix of  $X_0$ ,  $X_0(t) \in \mathbb{R}^{E \times B} \forall t \in \{1; \dots; T\}$  is a frequency decomposition corresponding to the past 2 s of  $S_{\text{EEG}}$ . We used a Hamming time window of 2 s to estimate the average power of each frequency band  $b \in \{1; \dots; B\}$  (defined below) on each channel  $e \in \{1; \dots; E\}$ . Each time window of the EEG signal overlapped by 1.75 s (0.25-s shift) to match with the 4 Hz sample of the  $y$  values. The  $B$  frequency bands have an overlap of 1 Hz with the next band and are defined between a minimum frequency  $b_{\min}$  Hz and a maximum frequency  $b_{\max}$  Hz (see section 3). We chose to use several relatively narrow frequency bands to let the model select the relevant bands for each electrode. Furthermore, it has been suggested (de Munck et al., 2009; Rosa et al., 2010) that different frequency bands be used when working with coupled EEG-fMRI data.



**FIGURE 2 |** Machine learning scheme. For each subject, a bimodal neurofeedback session (NF-EEG-fMRI session 1 here) is used for the learning step, and then the learned activation pattern  $\hat{\alpha}$  is applied to the other sessions (2 and 3) for the testing step. The learning data are split  $K$  times into a training set (90% of the learning set) and a cross-validation (CV) set (10% of the learning set). The optimal  $\hat{\lambda}$  parameter is the one minimizing the variance and the bias in the learning step.

The model also has to be able to predict  $y_f$  scores derived from the BOLD signal (see section 3). There is no linear relationship between the BOLD signal and the average power on frequency bands from the EEG signal. Therefore, to better match  $y_f$  scores, we decided to apply a non-linear function to  $X_0$  used in fMRI to model BOLD signals (Lindquist et al., 2009; Pedregosa et al., 2013), the canonical Hemodynamic Response function (HRF). We convolved  $X_0$  on its temporal dimension with the HRF, formed by two gamma functions, for a given delay of the first gamma function to compensate the response time of the BOLD signal, as suggested in Moosmann et al. (2008) and Meir-Hasson et al. (2014). The HRF will temporally smooth and give a BOLD-like shape to the design matrix and increase a potential linear relationship between  $y_f$  and the design matrix. Since HRF is known to vary considerably across brain regions and subjects (Handwerker et al., 2004), it is recommended to consider different delays but also to chose a range of values corresponding to the task asked. For the type of task addressed in the experimental part, the observed delay is around 4 s, so we convolved  $X_0$  with three different HRFs, leading to three new design matrices,  $X_3$ ,  $X_4$ , and  $X_5$ , with respectively peak locations of 3, 4, and 5 s. We started from the canonical HRF and changed the peak parameter (respectively using 3, 4, and 5) to induce the three different HRFs used to induce a time delay in the initial design matrix  $X_0$ . By doing so, we kept the total of design matrices to a reasonable size.

These design matrices are concatenated in their 2nd dimension to form the  $X_c \in \mathbb{R}^{T \times M \times B}$  matrix, with  $M = 4 * E$ .

Therefore, for each time  $t$ ,  $X_c(t) = [X_0(t); X_3(t); X_4(t); X_5(t)]$ . We also denote  $X_d(t) = [X_3(t); X_4(t); X_5(t)]$  the design matrix of the different delays.

### 2.2. Optimization

The EEG data are now represented by a structured design matrix  $X_c$ , and we can search for a weight matrix  $\hat{\alpha} \in \mathbb{R}^{M \times B}$ , such that  $\sum_{m,h} \hat{\alpha}(m, h) X_c(t, m, h)$  estimates the NF score  $y(t), \forall t \in \{1; \dots; T\}$  as well as possible. Note: the methodology is presented for the design matrix  $X_c$  but can be used for  $X_0$  or  $X_d$ .

To identify  $\hat{\alpha}$ , called the activation pattern, we propose the following strategy, which consists of learning, for a given subject and NF session, the optimal  $\hat{\alpha}$  by solving the following problem:

$$\hat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \sum_{t=1}^T \frac{1}{2} (y(t) - q(X_c(t), \alpha))^2 + \phi_{\lambda}(\alpha) \quad (2)$$

with  $\phi_{\lambda}$  a regularization term, and  $\lambda$  a weighting parameter for the regularization term. This  $\hat{\alpha}$  is then applied to a design matrix  $X_c^{test}$  from a new session to predict its NF scores.

$$\tilde{y}_{\hat{\alpha}}(t) = q(X_c^{test}(t), \hat{\alpha}) \forall t \in \{1; \dots; T\} \quad (3)$$

Equation 2 is of the form  $\operatorname{argmin}(g_1(\alpha) + g_2(\alpha))$ , and its resolution can be performed using the Fast Iterative Shrinkage Thresholding Algorithm (FISTA) (Beck and Teboulle, 2009), which is a two-step approach of the Forward-Backward algorithm (Combettes and Wajs, 2005), making it faster. FISTA requires the same

conditions as the Forward-Backward algorithm, meaning a convex differentiable with Lipschitz gradient term  $g_1$  and a convex term  $g_2$  that is not necessarily differentiable but smooth enough to make its proximal map computable.

Here,  $g_1(\alpha) = \sum_t \frac{1}{2}(y(t) - q(X_c(t), \alpha))^2$  is a sum of convex and differentiable functions with

$$\nabla g_1(\alpha) = \sum_t -X_c(t)(y(t) - q(X_c(t), \alpha)) \quad (4)$$

since  $\forall i \in \{1; \dots; M\}, j \in \{1; \dots; B\}, [\frac{\partial q(X_c(t), \alpha)}{\partial \alpha(i,j)}]_{i,j} = X_c(t, i, j)$ . By representing  $X_c(t)$  and  $\alpha$  as vectors of size  $M * B$ , we can easily note that  $\frac{\partial g_1}{\partial \alpha}$  is a sum of Lipschitz functions. Therefore, the Lipschitz constant of  $\frac{\partial g_1}{\partial \alpha}$  is  $L = \|X_V^T X_V\|$ , with  $X_V \in \mathbb{R}^{T * M * B}$  the vectorized version of  $X_c$ .

The NF-predictor uses the structured design matrix to achieve better control over the interpretation of results and to better optimize the weights  $\hat{\alpha}$ . Therefore, we have to adopt an optimization strategy that is coherent with this structure. The activation pattern of the NF-predictor:

1. has to be spatially sparse to regulate the model, as EEG signals are noisy, and to select the most relevant electrodes on each frequency bands,
2. has to be smooth across different overlapped frequency bands,
3. has to allow non-relevant frequency bands to be null.

The term  $g_2$  is the prior term. Here, for  $g_2(\alpha) = \phi_\lambda(\alpha)$ , we chose to use a  $\ell_{21}$  mixed norm (Ou et al., 2009) followed by a  $\ell_1$ -norm (noted  $\ell_{21+1}$ -norm in Gramfort et al., 2011) to fit all the structure conditions mentioned above.

$$\phi_\lambda(\alpha) = \lambda \|\alpha\|_{21} + \rho \|\alpha\|_1 \quad (5)$$

with  $\rho \in \mathbb{R}^+$  and  $\lambda \in \mathbb{R}^+$ . We chose not to estimate the parameter  $\rho$  to keep computation time reasonable. Indeed,  $\rho$  weights the induced spatial sparsity over EEG channels, and we chose to fix this parameter for all subjects, as we hypothesize that there is no reason for the number of electrodes involved in the activation pattern to significantly change between subjects. However, the estimation of the  $\lambda$  parameter is needed (since we do not have a hypothesis on its behavior) and is presented in the next section. The  $\ell_{21}$  mixed norm that writes  $\|\alpha\|_{21} = \sum_m \sqrt{\sum_b \alpha_{m,b}^2}$  satisfies conditions (1) and (2). The  $\ell_1$  norm defined as  $\|\alpha\|_1 = \sum_{m,b} |\alpha_{m,b}|$  satisfies condition (3) since  $\ell_p$  norms with  $p \leq 1$  are known to promote sparsity. The last key point of FISTA algorithm is the proximal map associated to the  $\ell_{21+1}$  norm  $\text{Prox}_{\ell_{21+1}}: \mathbb{R}^{M \times B} \rightarrow \mathbb{R}^{M \times B}, \beta \mapsto \text{argmin}_\alpha (\phi_\lambda(\alpha) + 1/2 \|\beta - \alpha\|^2)$ , defined as

$$\begin{aligned} & (\text{Prox}_{\ell_{21+1}}(Y))_{m,b} \\ &= \frac{Y_{m,b}}{|Y_{m,b}|} (|Y_{m,b}| - \rho)^+ \left(1 - \frac{\lambda}{\sqrt{\sum_b (|Y_{m,b}| - \rho)^2}}\right)^+ \quad (6) \end{aligned}$$

with operator  $(\cdot)^+ = \max(\cdot, 0)$ . One can note that by canceling either the  $\lambda$  parameter or the  $\rho$  parameter, we retrieve the proximal map associated with the  $\ell_{21}$  (when  $\rho = 0$ ) and the  $\ell_1$

(when  $\lambda = 0$ ) norms, demonstrations of which can be found in the appendix of Gramfort et al. (2012). For the stopping criterion of FISTA, a large enough number of iterations has been used to allow the model to converge before reaching the last iteration. All elements and conditions are gathered to run the FISTA algorithm.

### 2.3. $\lambda$ Parameter Selection

The parameter  $\lambda$  is important in the optimization problem, and we decided to estimate it automatically. The following process chooses the best  $\lambda$  among a list of  $\Lambda = \{\lambda_1; \dots; \lambda_I\}$  sorted in increasing order. First of all, the data must be split into two sets. In our case, subjects have three NF-EEG-fMRI sessions: one session is used as the learning set, and the other two NF-EEG-fMRI sessions are used as the testing sets (see **Figure 2**). For each value  $\lambda_i$  of  $\Lambda$ , the learning set, formed by  $T$  neurofeedback scores with their associated design matrices, is divided  $K = 50$  times into a training set of indices  $R_k$ , representing 90% of the  $T$  data, and a cross-validation set  $CV_k$  composed of the remaining 10% of the learning set. A model  $\hat{\alpha}^{(k,i)}$  is estimated on the training dataset  $k$  composed by  $R_k$  neurofeedback scores  $y(j)$  and the associated design matrices  $X_c(j)$  with  $\lambda_i$ , i.e.:

$$\begin{aligned} \hat{\alpha}^{(k,i)} = \text{arg min} & \sum_j \|y(j) - q(X_c(j), \hat{\alpha}^{(k,i)})\|^2 \\ & + \lambda_i \|\alpha^{(k,i)}\|_{21} + \rho \|\alpha^{(k,i)}\|_1 \quad (7) \end{aligned}$$

with  $j \in R_k$  and  $X_c(j) \in \mathbb{R}^{M \times B}$ . For the current  $\lambda_i$  evaluation, we stop the process when  $\sum_k |\hat{\alpha}^{(k,i)}|_0 / K < 2$ . There is no need to investigate the next  $\lambda_i$ ; the current one is sparse enough, and the next one might lead to null models.

We then apply the model  $\hat{\alpha}^{(k,i)}$  to the corresponding cross-validation set of  $CV_k$  NF scores to obtain estimated values of  $y(s), \tilde{y}(s) = q(X_c(s), \hat{\alpha}^{(k,i)})$  with  $s \in CV_k$ . For each one of the 50 partitioned into training and cross-validation sets, we computed the normalized mean squared error NMSE for a given set of data  $\{y, X_c\}$  for the training sets and the cross-validation sets.

$$\text{NMSE}(\{y, X_c\}, \hat{\alpha}^{(k,i)}) = \frac{\sum_s (y(s) - q(X_c(s), \hat{\alpha}^{(k,i)}))^2}{\sum_s (y(s) - \bar{y})^2} \quad (8)$$

with  $\bar{y} = 1/n \sum_s y(s)$ . The optimal  $\hat{\lambda}$  parameter is defined as the one minimizing the error during training and cross-validation. Considering that only the errors from the training set  $\text{NMSE}(\{y(R_k), X_c(R_k)\}, \hat{\alpha}^{(k,i)})$  would introduce bias, and considering only the error of the cross-validation set  $\text{NMSE}(\{y(CV_k), X_c(CV_k)\}, \hat{\alpha}^{(k,i)})$  would introduce variance. Then, the optimal  $\hat{\lambda}$  is the  $\lambda_i$  that minimizes

$$\begin{aligned} & \sum_{k=1}^K [\text{NMSE}(\{y(R_k), X_c(R_k)\}, \hat{\alpha}^{(k,i)}) \\ & + \text{NMSE}(\{y(CV_k), X_c(CV_k)\}, \hat{\alpha}^{(k,i)})] \quad (9) \end{aligned}$$

The  $\hat{\lambda}$  parameter is the optimal parameter used for the model estimation. If there are several candidates, to favor sparsity, the larger of these candidates is chosen.

### 3. DATA ACQUISITION AND PRE-PROCESSING

We used an existing dataset presented in Perronnet et al. (2018) composed of 17 healthy subjects who were scanned using the hybrid Neurofeedback platform from Neurinfo (Rennes, France) coupling EEG and fMRI signals (Mano et al., 2017). Data are now available online in BIDs format on OpenNeuro: <https://openneuro.org/datasets/ds002338> (Lioi et al., 2019). A 64-channel MR-compatible EEG solution from Brain Products (Brain Products GmbH, Gilching, Germany) was used, the signal was sampled at 5 kHz, and FCz was the reference electrode and AFz the ground electrode. For the fMRI scanner, we used a 3T Verio Siemens scanner with a 12-channel head coil (repetition time (TR)/echo time (TE) = 2,000/23 ms, FOV = 210 × 210 mm<sup>2</sup>, voxel size = 2 × 2 × 4 mm<sup>3</sup>, matrix size = 105 × 105 with 16 slices, flip angle = 90°). All subjects were healthy volunteers, were right-handed, and had never taken part in any neurofeedback experiments before. They all gave written informed consent in accordance with the Declaration of Helsinki, as specified in the study presenting the data (Perronnet et al., 2017). They all underwent three NF motor imagery sessions of 320 s each after a session dedicated to the calibration. One session consists of eight blocks alternating between 20 s of rest, eyes open, and 20 s of motor imagery of the right hand. The neurofeedback display was uni-dimensional (1D) for nine subjects (Figure 3 left) and bi-dimensional (2D) for eight subjects (Figure 3 middle). For both, the goal was to bring the ball into the dark blue area (Perronnet et al., 2018).

NF scores  $y_e$  and  $y_f$ , being from different modalities, were standardized before summing to form  $y_c$ . In this study, NF scores refer to standardized scores except when they are predicted. For this study,  $y_e$  have been computed from the Laplacian operator (commonly used in neurofeedback) centered around the region of interest, channel C3 here. For each time interval  $I_t$ , the spatial filtering is noted as  $\text{Lap}(C3, I_t)$ . The temporal segments  $I_t$  are spaced at 250 ms apart and have a length of 2 s (therefore, there is an overlap of 1.75 s), as in the design matrix construction. The power of the frequency band [8–30 Hz] is then extracted via the Power Spectral Density function PSD:

$$y_e(t) = -\text{PSD}_{[8-30]}(\text{Lap}(C3, I_t))$$

One may note the presence of the minus operator, which is used here for the sake of coherence with  $y_f$  (Figure 3 right).

The neurofeedback scores  $y_f$  have been computed from the maximal intensity of BOLD signal covering the right-hand motor area and the supplementary motor area. One score is computed per volume acquired (i.e., 1 per second). Scores  $y_f$  are then re-sampled and smoothed (using a Savitzky-Golay filter, known to avoid signal distortion) to fit the 4 Hz  $y_e$  scores ( $T = 1, 280$ ).

An active set has been selected on the design matrices to avoid potentially correlated noise due to head movement during resting blocks obstructing the signal from channels of interest. Indeed in coupled EEG-fMRI acquisitions, subjects are lying in the MRI scanner, so the outer electrodes may be in contact with the bed or holds. We excluded the outer

electrodes and kept 28 electrodes: the three central lines have seven electrodes (FCz is the reference), three frontal electrodes, and three posterior electrodes.

Potential outliers in the design matrices (i.e., observations  $>$  mean  $\pm$  3std) were thresholded in the NF-EEG-fMRI session used as the learning set, and bad observations from annotations on the EEG signal were removed as their corresponding NF scores. For the frequency dimension of the design matrix  $X_0$  construction (cf. section 2) and therefore for the other design matrices, we chose  $b_{min} = 8$ ,  $b_{max} = 30$  to cover the alpha and beta frequency bands involved in motor tasks. We considered  $B = 10$  frequency bands, leading to bands of 3 Hz wide with an overlap of 1Hz.

As mentioned in the previous section, for the regularization of the NF-predictor, we used 15 values of  $\lambda_i$  from 100 to 3,000 and fixed the parameter  $\rho = 1, 500$ .

### 4. EXPERIMENTS AND RESULTS

As stated in the previous section, there were three neurofeedback sessions per subject. For each subject, we will consider one session as a learning set, and the two others as testing sets. Thus, there are three different learning sets and six different testing sets per subject.

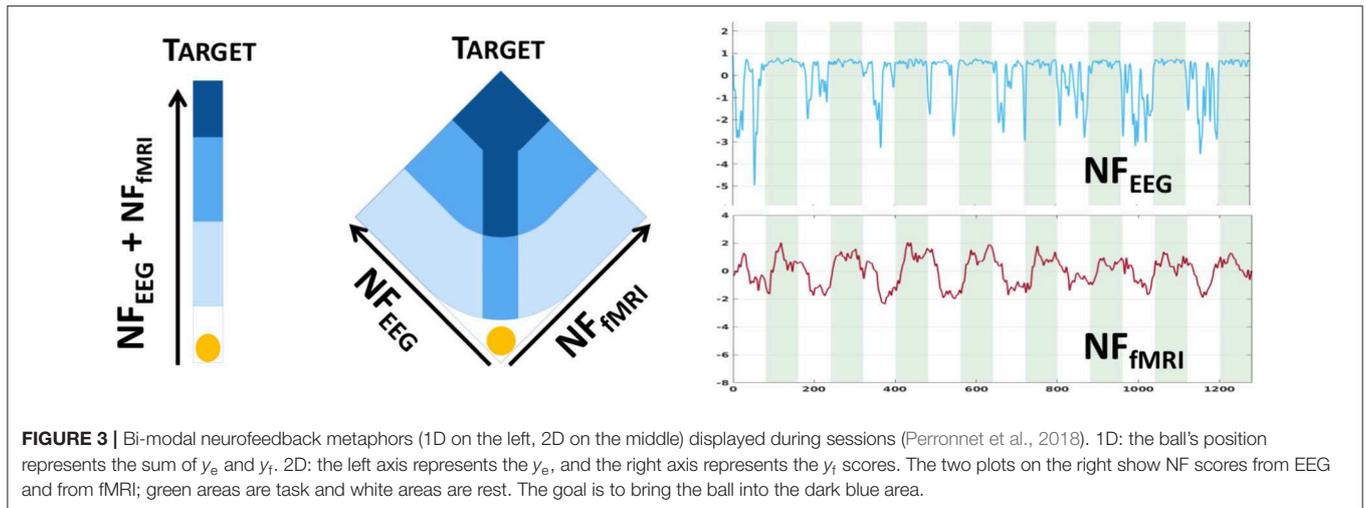
#### 4.1. Experiments and Validation

We tested different NF-predictors for the prediction of different NF scores,  $\forall t \in \{1; \dots; T\}$ :

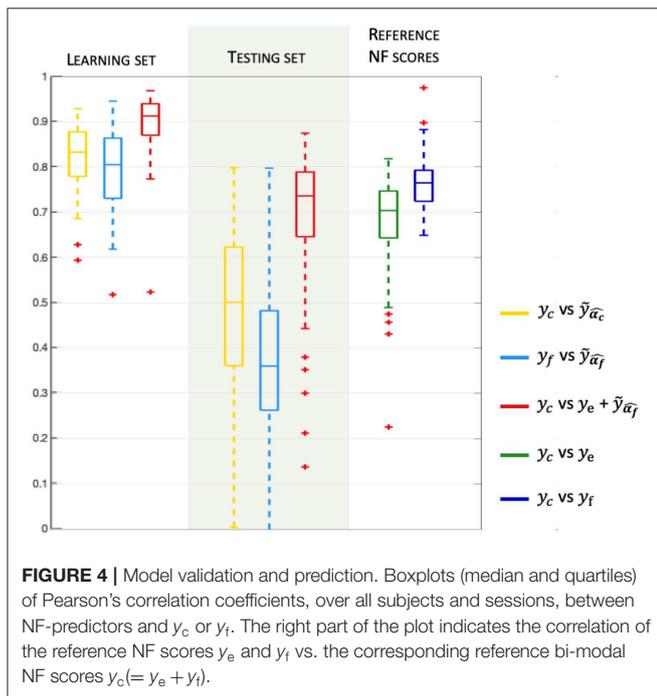
1.  $\tilde{y}_{\hat{\alpha}_c}(t) = q(X_c^{test}(t), \hat{\alpha}_c)$  with  $\hat{\alpha}_c$  (Equation 2), learned from  $X_c$  and  $y_c$
2.  $\tilde{y}_{\hat{\alpha}_f}(t) = q(X_d^{test}(t), \hat{\alpha}_f)$  with  $X_d = [X_3; X_4; X_5]$  and  $\hat{\alpha}_f$  (Equation 2), learned from  $X_d$  and  $y_f$
3.  $y_e(t) + \tilde{y}_{\hat{\alpha}_f}(t)$  using NF-predictor 2.

The advantage of the last NF-predictor is the possibility of using the 2D score visualization (Figure 3) to display NF scores. Also in this case, only NF-fMRI scores have to be learned. We run the following experiments to test our different NF-predictors:

- Model validation: we used the learning set to assess whether the different NF-predictors could accurately model the NF scores. For each subject and each NF-predictor, we estimated correlations with the reference NF score to quantify the quality of prediction. As a reference, we compared to the correlation of  $y_e$  to  $y_c$ , which is part of the  $y_c$  score, to assess whether, in the validation process, the model can better predict  $y_c$  scores than  $y_e$ .
- Model prediction: we applied the learned activation patterns to a new NF-EEG-fMRI recorded session (testing set, Figure 2). For each subject and each pair of sessions (six learning/testing pairs per subject), we compared correlations between NF-predictors  $\tilde{y}_{\hat{\alpha}_c}$  and  $y_e + \tilde{y}_{\hat{\alpha}_f}$  to reference score  $y_c$  and between the prediction of  $\tilde{y}_{\hat{\alpha}_f}$  to  $y_f$ .
- To observe the captured structure in the activation patterns  $\hat{\alpha}_c$  and  $\hat{\alpha}_f$ , respectively learned to predict  $y_c$  and  $y_f$ , we re-shaped the average activation patterns of a subject over sessions into



**FIGURE 3** | Bi-modal neurofeedback metaphors (1D on the left, 2D on the middle) displayed during sessions (Perronnet et al., 2018). 1D: the ball's position represents the sum of  $y_e$  and  $y_f$ . 2D: the left axis represents the  $y_e$ , and the right axis represents the  $y_f$  scores. The two plots on the right show NF scores from EEG and from fMRI; green areas are task and white areas are rest. The goal is to bring the ball into the dark blue area.



**FIGURE 4** | Model validation and prediction. Boxplots (median and quartiles) of Pearson's correlation coefficients, over all subjects and sessions, between NF-predictors and  $y_c$  or  $y_f$ . The right part of the plot indicates the correlation of the reference NF scores  $y_e$  and  $y_f$  vs. the corresponding reference bi-modal NF scores  $y_c (= y_e + y_f)$ .

matrices corresponding to the design matrices defining  $X_c$ , respectively  $X_d$  (see section 2.1) and displayed the results of the first dimension (electrodes) and of the second dimension (frequency bands).

## 4.2. Results

### 4.2.1. Model and Prediction

The results of the model validation of the NF-predictors (i.e., the learned activation patterns  $\hat{\alpha}$  are applied to the learning set) are shown on the left side of **Figure 4**.

Pearson's correlation coefficients between the prediction and the ground truth are computed for each of the three learning sessions and for all subjects. Correlations are very high ( $> 0.8$ ) for the two NF-predictors of  $y_c$  scores, indicating that the model

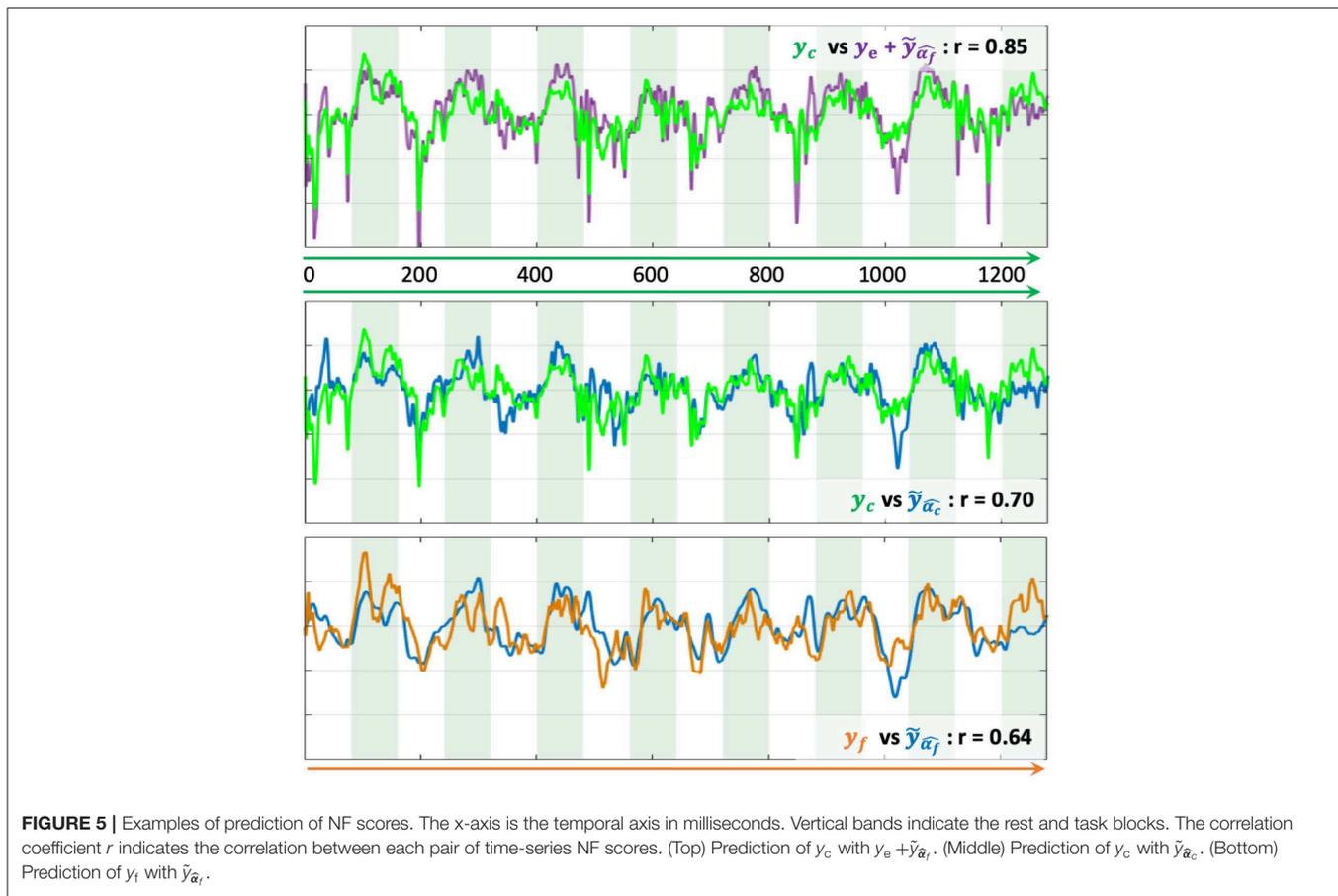
**TABLE 1** | Model prediction: Pearson's correlation coefficients, over all subjects and sessions, between NF-predictors and  $y_c$ .

Correlation ( $y_{c, \cdot}$ )	Paired t-test		
	$\tilde{y}_{\hat{\alpha}_c}$	$y_e + \tilde{y}_{\hat{\alpha}_f}$	$y_e$
$\tilde{y}_{\hat{\alpha}_c}$	N.A.	$t = -23.17$ $p = 1$	$t = -12.76$ $p = 1$
$y_e + \tilde{y}_{\hat{\alpha}_f}$	$t = 23.17$ $p = 2.6 \cdot 10^{-42}$	N.A.	$t = 3.52$ $p = 6.6 \cdot 10^{-4}$
$y_e$	$t = 12.76$ $p = 9 \cdot 10^{-23}$	$t = -3.51$ $p = 1$	N.A.

The table shows  $p$ -values of a  $t$ -test in which the hypothesis is: models in columns correlate better with  $y_c$  than models in rows.

is adapted to the problem of NF prediction. Indeed, the model could fit NF-fMRI scores using only EEG signal information with a median correlation of  $r = 0.81$  with  $y_f$ .

For the evaluation of the model prediction, the learned activation patterns are applied to the testing sets, i.e., the two other unseen NF sessions for each one of the three NF sessions. The significance of the differences between the correlations to the reference NF scores is assessed by applying a Fisher transformation to the correlation coefficients. The  $p$ -values obtained via Student's  $t$ -test after this transformation indicate the significance of the differences between the correlation coefficient distributions. **Table 1** shows that  $y_e + \tilde{y}_{\hat{\alpha}_f}$ , whose median correlation is  $r = 0.74$ , better predicts  $y_c$  than  $y_e$  only, despite the fact that  $y_e$  is part of  $y_c$ . The paired  $t$ -test gives a  $p$ -value  $p = 6.6 \cdot 10^{-4}$  ( $t = 3.52$ ), meaning that the prediction of NF-fMRI scores by  $\tilde{y}_{\hat{\alpha}_f}$  significantly adds information to  $y_e$ . Predicting  $y_c$  scores with  $y_e + \tilde{y}_{\hat{\alpha}_f}$  is also better than  $\tilde{y}_{\hat{\alpha}_c}$  ( $p = 2 \cdot 10^{-42}$ ,  $t = 23.17$ ), which is expected since the model predicts NF-fMRI scores and directly uses NF-EEG scores (with all the potential noise coming from the EEG measures), which are part of the reference score, as  $y_e$  is in  $y_e + \tilde{y}_{\hat{\alpha}_f}$ , and therefore correlates well with  $y_c$ . Furthermore, the proposed model is able to predict  $y_f$  scores with a fair correlation of 0.36 in the median and



0.35 in the average. Thus, predicting NF-fMRI scores instead of predicting bi-modal NF scores seems to be the best way of predicting those bi-modal NF scores. **Figure 5** gives examples of predictions of  $y_c$  with  $y_e + \tilde{y}_{\hat{\alpha}_f}$  and  $\tilde{y}_{\hat{\alpha}_c}$ . It also illustrates that, even if the correlations of  $\tilde{y}_{\hat{\alpha}_c}$  are lower than  $y_e + \tilde{y}_{\hat{\alpha}_f}$  (**Table 2**),  $\tilde{y}_{\hat{\alpha}_c}$  can correctly predict the reference score  $y_c$ .

During the learning session (i.e., the NF session used to learn the predictor), subjects might unevenly focus on their fMRI or EEG feedbacks, and when subjects focus more on NF-fMRI than on NF-EEG, the EEG-signals might lose coherence with respect to the NF-fMRI scores. Despite this, the EEG signals could predict NF-fMRI scores with a correlation of 0.36 in median and 0.35 in mean (cf **Figure 4**, which is a fair correlation between such different modalities). Examples of NF prediction are given at the bottom of **Figure 5**, the plot shows the prediction  $\tilde{y}_{\hat{\alpha}_f}$  of NF-fMRI scores on a testing set.

#### 4.2.2. Activation Patterns

We now focus on the learned models to evaluate the sparsity over sessions and subjects and observe the dispersion of the learned patterns over the sessions and frequency bands of a subject. For each subject and each learning session, we computed the proportion of zero coefficients of the activation patterns (**Table 2**). By construction of the design matrix  $X_c$ ,  $\hat{\alpha}_c$ , and  $\hat{\alpha}_f$  can be split into different activation patterns, as shown in the columns

**TABLE 2 |** Sparsity of the learned models  $\hat{\alpha}_c$  and  $\hat{\alpha}_f$ .

	$\hat{\alpha}_0$	$\hat{\alpha}_3$	$\hat{\alpha}_4$	$\hat{\alpha}_5$	All
$\hat{\alpha}_f$	x	$0.92 \pm 0.03$	$0.96 \pm 0.03$	$0.91 \pm 0.03$	$0.93 \pm 0.04$
$\hat{\alpha}_c$	$0.87 \pm 0.05$	$0.92 \pm 0.03$	$0.96 \pm 0.02$	$0.93 \pm 0.03$	$0.92 \pm 0.05$

Proportion of zero coefficients in the learned models.

of **Table 2**. The models could select relevant coefficients of the design matrices to predict reference NF scores. In average, there are 87 non-zero coefficients on  $\hat{\alpha}_c$  to predict  $y_c$  and 57 non-zero coefficients on  $\hat{\alpha}_f$  to predict  $y_f$ .

To display the activation patterns of a subject over their three sessions, we denote  $\hat{\zeta}_c = \sum_{s=1}^3 |\hat{\alpha}_c^{(s)}| \in \mathbb{R}^{M \times B}$  the absolute activation pattern of  $\hat{\alpha}_c$  (respectively,  $\hat{\zeta}_d^f = \sum_{s=1}^3 |\hat{\alpha}_d^{(s)}| \in \mathbb{R}^{M \times B}$  for  $\hat{\alpha}_f$ ) and  $\hat{\gamma}_c = \sum_{s=1}^3 \hat{\alpha}_c^{(s)} \in \mathbb{R}^{M \times B}$  the average activation pattern (respectively  $\hat{\gamma}_d^f = \sum_{s=1}^3 \hat{\alpha}_d^{(s)} \in \mathbb{R}^{M \times B}$ ). We can display heat maps for each of the four absolute activation patterns at each electrode  $\sum_{b \in B} \hat{\zeta}_0$ ,  $\sum_b \hat{\zeta}_3$ ,  $\sum_b \hat{\zeta}_4$ , and  $\sum_b \hat{\zeta}_5$  (**Figure 6**, top line of each panel) and color maps of the four average patterns  $\sum_b \hat{\gamma}_0$ ,  $\sum_b \hat{\gamma}_3$ ,  $\sum_b \hat{\gamma}_4$ , and  $\sum_b \hat{\gamma}_5$  (**Figure 6**, bottom line of each panel) showing the sign of the strongest and most stable coefficients across all subjects and sessions. The activation patterns displayed at **Figure 6** represent the dispersion of the



comparable learned activation patterns to predict  $\tilde{y}_{\hat{\alpha}_f}$ , but more spread over sessions than for the model on panel A, even if the model has around 93% of null coefficients (Table 2). This suggests that, even for a single subject, the relation between EEG and fMRI changes over sessions, probably depending on the strategy used by the subject.

At last, it is also possible to display the frequency profile of each average activation patterns. The 4 frequency profiles are  $\sum_{e \in \{1, \dots, E\}} \hat{\mathcal{Y}}_0$ ,  $\sum_e \hat{\mathcal{Y}}_3$ ,  $\sum_e \hat{\mathcal{Y}}_4$  and  $\sum_e \hat{\mathcal{Y}}_5$ ,  $e \in \{1, \dots, E\}$ , summing weights over electrodes. Figure 7 shows that the most used frequency bands over all sessions and electrodes are the alpha band ([8–12] Hz) and lower beta band ([13–17] Hz); the last four frequency bands are not displayed as they only have null coefficients. We can observe the effect of  $\ell_2$  regularization, which allows continuity in the frequency bands, and the effect of the subsequent  $\ell_1$  regularization, which removed the smaller coefficients located in the high frequencies. When considering all activation patterns (Figure 7 right side), there is a change of sign between alpha and lower beta. Each activation pattern shows a different frequency profile, which, together with Figure 6, tends to demonstrate that patterns have complementary information.

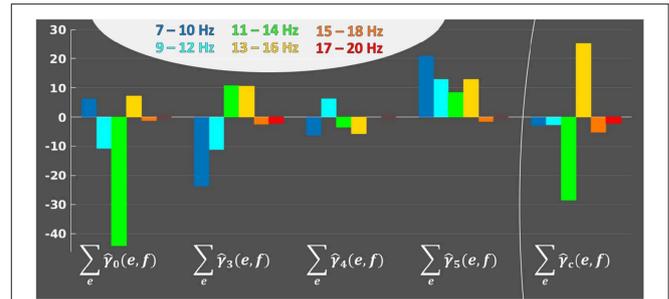
### 4.3. Ablation Study

We run an ablation study to understand the impact of the non-linearity part of the model, obtained by the convolution of different HRFs inducing different delays for the prediction of NF-fMRI scores, by analyzing the results of the model without HRF convolutions. Therefore, in this ablation study, we used the exact same processes as described in section 4 (with a learning step and a testing step), except that we used the design matrix  $X_0$  alone to learn the model parameters to predict  $y_c$  scores and  $y_f$  scores.

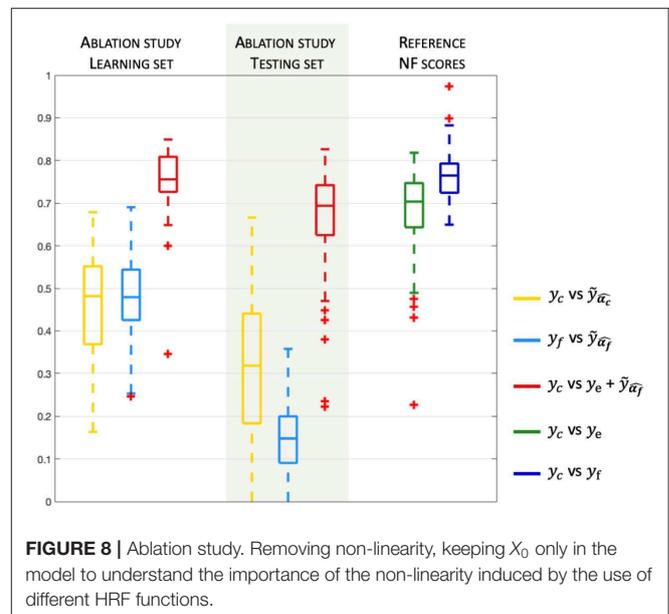
The results of the ablation study are displayed at Figure 8 and are to be compared with the results of the proposed model presented at Figure 4. Figure 8 shows that during the learning step, the prediction of  $\tilde{y}_{\hat{\alpha}_f}$  using  $X_0$  only correlates with  $y_f$  with a correlation of only 0.48 in mean and median. This prediction is significantly improved by the use of the different HRF functions (paired  $t$ -test,  $t = -14, 64, p = 1e-26$ ), as, in the proposed model,  $\tilde{y}_{\hat{\alpha}_f}$  correlates with  $y_f$  with a correlation of 0.81 in median and mean. Therefore, it is natural to observe that the prediction of the bi-modal NF score by  $y_e + \tilde{y}_{\hat{\alpha}_f}$  is also significantly improved by the use of non-linearity (paired  $t$ -test,  $t = -11, 87, p = 4e-21$ ). Figure 8 shows that, for the testing step, the prediction of  $\tilde{y}_{\hat{\alpha}_f}$  using  $X_0$  only correlates with  $y_f$  with a correlation of 0.14 in mean and 0.15 in median, which is significantly lower than with the proposed model (paired  $t$ -test,  $t = -9.56, p = 2e-18$ ). The prediction of the bi-modal NF scores by  $y_e + \tilde{y}_{\hat{\alpha}_f}$  is also significantly lower than with the proposed model (paired  $t$ -test,  $t = -2.74, p = 3e-3$ ). Also, when using  $X_0$  only, the prediction of the bi-modal NF scores using  $y_e + \tilde{y}_{\hat{\alpha}_f}$  does not significantly improve over  $y_e$  alone (paired  $t$ -test,  $t = -0.69, p = 0.5$ ).

## 5. DISCUSSION AND CONCLUSION

The model validation supports that the optimization strategy we chose for our problem is adapted to the model, as is the choice of



**FIGURE 7** | Frequency profiles, across subjects and sessions, of each average activation pattern, to represent the implication of each frequency band in the activation patterns.  $e \in \{1, \dots, E\}$ . For each average activation pattern, the y-axis indicates the sum of weights over electrodes for each frequency band  $f \in \{1, \dots, B\}$  on the horizontal axis.



**FIGURE 8** | Ablation study. Removing non-linearity, keeping  $X_0$  only in the model to understand the importance of the non-linearity induced by the use of different HRF functions.

the different design matrices. The ablation study supports the use of different non-linear delays to improve the prediction of NF-fMRI scores using the EEG signal. The evaluation of the model prediction strongly suggests that predicting only NF-fMRI scores from an EEG signal while applying a Laplacian on the EEG signal appears to be the best solution. Indeed, for  $\tilde{y}_{\hat{\alpha}_c}$ , when predicting the bi-modal NF scores, the variability between EEG signals induces a decreasing correlation with the reference NF score  $y_c$ . Also, NF-EEG scores can always be computed from available EEG signals, which leaves only NF-fMRI scores to be estimated from EEG signals. However, one might want to improve the selection of features for the computation of NF-EEG scores, but this raises different questions about the validation and the reference score. Here, we assumed that the given NF scores are relevant to the task and good enough to be considered as reference scores. As expected, for the prediction of the bi-modal NF scores by  $\tilde{y}_{\hat{\alpha}_c}$ , when decomposing the activation pattern  $\hat{\alpha}_c$  (from  $\tilde{y}_{\hat{\alpha}_c}$ ) into the four matrices corresponding to the different design matrices, the

weights corresponding to  $X_0$  (0 s of delay) are mainly located above C3, which is the center of the Laplacian used for the computation of the NF-EEG scores. This seems to support the fact that with a delay of 0 s, only the part of the NF coming from the EEG provides information.

A possibility to improve the prediction of NF-fMRI scores would be to use more NF sessions as learning sessions, since, as observed, EEG signals bring variability into the prediction. Each new bi-modal neurofeedback session could be added to the subject-specific model to better adapt the model to the subject or patient. This will be investigated in the next study. Given the improved correlation of the proposed NF predictor with bimodal NF scores, it would be interesting in future work to validate its improved performance in actual NF sessions compared to classical NF-EEG scores. In particular, to assess the response of subjects to the predicted bi-modal NF scores and particularly the predicted NF-fMRI scores learned by the proposed model over a standard NF-EEG neurofeedback session, a new and large enough study is needed, as subjects can learn at different paces to regulate their own brain activity.

Presently, the proposed model learns an individual or specific model for each subject, which allows a personalized model for adapted neurofeedback sessions. Also, a change in strategy for the task (here, no specific strategy for imagining moving their right hand was given to the subjects) might impact the learned model, as the relation between the EEG and fMRI signals may change. However, in a forthcoming work, we are investigating an adaptation of the methodology for the extraction of a common model, taking into account the differences between sessions and subjects, allowing the prediction of NF-EEG-fMRI scores on new subjects who did not participate in the model construction. The model might be less specific, but this would give access to neurofeedback sessions of bi-modal quality using EEG only for subjects with MRI contraindications and/or drive a subject-specific model estimation, respecting the strategy used by the subject to progress in their neurofeedback task. The learned NF predictor is, of course, specific to the particular task considered during the learning session. One can expect to generalize the approach to other tasks where spatial sparsity is relevant. Its extension beyond such tasks is more challenging.

Other ways to improve the method proposed here would be to investigate the use of dynamic functional connectivity, a relatively recent field in BOLD fMRI, the use of which along with EEG data requires further investigation (Tagliazucchi and Laufs, 2015). Dynamic functional connectivity would allow the different networks (reward, control, and learning) involved during neurofeedback sessions to be taken into account (Sitaram et al., 2017). Dynamic functional connectivity studies the temporal fluctuations of the BOLD signal across the brain and appears to be a promising approach in the EEG-fMRI research

field (Allen et al., 2018). However, one should be careful with the potential unknown remaining noise coming from the MRI during simultaneous EEG-fMRI recording, which might unsettle the EEG signal coherence between electrodes.

The long-term objective of our project is to learn from EEG-fMRI NF sessions to provide, outside the MRI scanner, enhanced NF-EEG sessions (Figure 1). Future work will investigate the portability of the learned model (on EEG-fMRI neurofeedback data), outside the MRI scanner, introducing new challenges, such as dealing with the remaining noise in the MRI after artifact correction and the absence of ground truth once the EEG is measured outside the MRI scanner.

To conclude, the model proposed here is able to provide a good enough prediction of the NF-fMRI scores to overcome the absence of NF-fMRI and allows the estimation of NF-EEG-fMRI scores when using EEG only to be significantly improved.

## DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <https://openneuro.org/datasets/ds002338>.

## ETHICS STATEMENT

The studies involving human participants were reviewed and approved by Comité de Protection des Personnes Ouest V Rennes. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

CC was the guarantor of integrity of the entire study. All authors contributed to the study concepts and design, data analysis and interpretation, and manuscript drafting or manuscript revision for important intellectual content. All authors gave approval of the final version of the submitted manuscript.

## ACKNOWLEDGMENTS

Data acquisition was supported by the Neurinfo MRI research facility at the University of Rennes I. Neurinfo receives grants from the European Union (FEDER), the French State, the Brittany Council, Rennes Metropole, Inria, Inserm, and the University Hospital of Rennes. This work has received French government support granted to the CominLabs excellence laboratory and managed by the National Research Agency in the Investissements d'Avenir program under reference ANR-10-LABX-07-01. It was also funded by the Brittany region under the HEMISFER project.

## REFERENCES

Abreu, R., Leal, A., and Figueiredo, P. (2018). EEG-informed fMRI: a review of data analysis methods. *Front. Hum. Neurosci.* 12:29. doi: 10.3389/fnhum.2018.00029

Allen, E. A., Damaraju, E., Eichele, T., Wu, L., and Calhoun, V. D. (2018). EEG signatures of dynamic functional network connectivity states. *Brain Topogr.* 31, 101–116. doi: 10.1007/s10548-017-0546-2

- Beck, A., and Teboulle, M. (2009). A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* 2, 183–202. doi: 10.1137/080716542
- Birbaumer, N., Ramos Murguialday, A., Weber, C., and Montoya, P. (2009). Chapter 8: neurofeedback and brain-computer interface. *Int. Rev. Neurobiol.* 86, 107–117. doi: 10.1016/S0074-7742(09)86008-X
- Combettes, P. L., and Wajs, V. R. (2005). Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.* 4, 1168–1200. doi: 10.1137/050626090
- de Munck, J., Gonçalves, S., Huijboom, L., Kuijter, J., Pouwels, P., Heethaar, R., et al. (2007). The hemodynamic response of the alpha rhythm: an EEG/fMRI study. *Neuroimage* 35, 1142–1151. doi: 10.1016/j.neuroimage.2007.01.022
- de Munck, J., Gonçalves, S., Mammoliti, R., Heethaar, R., and Lopes da Silva, F. (2009). Interactions between different EEG frequency bands and their effect on alpha-fMRI correlations. *Neuroimage* 47, 69–76. doi: 10.1016/j.neuroimage.2009.04.029
- Engell, A. D., Huettel, S., and McCarthy, G. (2012). The fMRI BOLD signal tracks electrophysiological spectral perturbations, not event-related potentials. *Neuroimage* 59, 2600–2606. doi: 10.1016/j.neuroimage.2011.08.079
- Formaggio, E., Storti, S. F., Bertoldo, A., Manganotti, P., Fiaschi, A., and Toffolo, G. M. (2011). Integrating EEG and fMRI in epilepsy. *Neuroimage* 54, 2719–2731. doi: 10.1016/j.neuroimage.2010.11.038
- Friston, K. J., Jezzard, P., and Turner, R. (1994). Analysis of functional MRI time-series. *Hum. Brain Mapp.* 1, 153–171. doi: 10.1002/hbm.460010207
- Goncalves, S. I., Bijma, F., Pouwels, P. W. J., Jonker, M., Kuijter, J. P. A., Heethaar, R. M., et al. (2008). A data and model-driven approach to explore inter-subject variability of resting-state brain activity using EEG-fMRI. *IEEE J. Select. Top. Signal Process.* 2, 944–953. doi: 10.1109/JSTSP.2008.2009082
- Gramfort, A., Kowalski, M., and Hämäläinen, M. (2012). Mixed-norm estimates for the M/EEG inverse problem using accelerated gradient methods. *Phys. Med. Biol.* 57, 1937–1961. doi: 10.1088/0031-9155/57/7/1937
- Gramfort, A., Strohmeier, D., Hauelsen, J., Hamalainen, M., and Kowalski, M. (2011). “Functional brain imaging with M/EEG using structured sparsity in time-frequency dictionaries,” in *IPMI*, Vol. 6801, eds G. Székely and H. K. Hahn (Berlin; Heidelberg: Springer Berlin Heidelberg), 600–611.
- Grech, R., Cassar, T., Muscat, J., Camilleri, K. P., Fabri, S. G., Zervakis, M., et al. (2008). Review on solving the inverse problem in EEG source analysis. *J. Neuroeng. Rehabil.* 5:25. doi: 10.1186/1743-0003-5-25
- Hammond, D. C. (2011). What is neurofeedback: an update. *J. Neurother.* 15, 305–336. doi: 10.1080/10874208.2011.623090
- Handwerker, D. A., Ollinger, J. M., and D’Esposito, M. (2004). Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *Neuroimage* 21, 1639–1651. doi: 10.1016/j.neuroimage.2003.11.029
- Leite, M., Leal, A., and Figueiredo, P. (2013). Transfer function between EEG and BOLD signals of epileptic activity. *Front. Neurol.* 4:1. doi: 10.3389/fneur.2013.00001
- Lindquist, M. A., Meng Loh, J., Atlas, L. Y., and Wager, T. D. (2009). Modeling the hemodynamic response function in fMRI: efficiency, bias and mis-modeling. *Neuroimage* 45, S187–S198. doi: 10.1016/j.neuroimage.2008.10.065
- Lioi, G., Cury, C., Perronnet, L., Mano, M., Bannier, E., Lécuyer, A., et al. (2019). Simultaneous MRI-EEG during a motor imagery neurofeedback task: an open access brain imaging dataset for multi-modal data integration. *bioRxiv [Preprint]*. doi: 10.1101/862375
- Logothetis, N. K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412, 150–157. doi: 10.1038/35084005
- Magri, C., Schridde, U., Murayama, Y., Panzeri, S., and Logothetis, N. K. (2012). The amplitude and timing of the BOLD signal reflects the relationship between local field potential power at different frequencies. *J. Neurosci.* 32, 1395–1407. doi: 10.1523/JNEUROSCI.3985-11.2012
- Mano, M., Lécuyer, A., Bannier, E., Perronnet, L., Noorzadeh, S., and Barillot, C. (2017). How to build a hybrid neurofeedback platform combining EEG and fMRI. *Front. Neurosci.* 11:140. doi: 10.3389/fnins.2017.00140
- Meir-Hasson, Y., Kinreich, S., Podlipsky, I., Hendler, T., and Intrator, N. (2014). An EEG finger-print of fMRI deep regional activation. *Neuroimage* 102, 128–141. doi: 10.1016/j.neuroimage.2013.11.004
- Moosmann, M., Eichele, T., Nordby, H., Hugdahl, K., and Calhoun, V. D. (2008). Joint independent component analysis for simultaneous EEG-fMRI: principle and simulation. *Int. J. Psychophysiol.* 67, 212–221. doi: 10.1016/j.ijpsycho.2007.05.016
- Noorzadeh, S., Maurel, P., Oberlin, T., Gribonval, R., and Barillot, C. (2017). “Multi-modal EEG and fMRI source estimation using sparse constraints,” in *Medical Image Computing and Computer Assisted Intervention MICCAI 2017*, Vol. 10433, eds M. Descoteaux, L. Maier-Hein, A. Franz, P. Jannin, D. L. Collins, and S. Duchesne (Springer International Publishing), 442–450.
- Ou, W., Hämäläinen, M. S., and Golland, P. (2009). A distributed spatio-temporal EEG/MEG inverse solver. *Neuroimage* 44, 932–946. doi: 10.1016/j.neuroimage.2008.05.063
- Pedregosa, F., Eickenberg, M., Thirion, B., and Gramfort, A. (2013). “HRF estimation improves sensitivity of fMRI encoding and decoding models,” in *2013 International Workshop on Pattern Recognition in Neuroimaging* (Philadelphia, PA: IEEE), 165–169.
- Perronnet, L., Lécuyer, A., Mano, M., Bannier, E., Lotte, F., Clerc, M., et al. (2017). Unimodal versus bimodal EEG-fMRI neurofeedback of a motor imagery task. *Front. Hum. Neurosci.* 11:193. doi: 10.3389/fnhum.2017.00193
- Perronnet, L., Lécuyer, A., Mano, M., Clerc, M., Lotte, F., and Barillot, C. (2018). Learning 2-in-1: towards integrated EEG-fMRI-neurofeedback. *bioRxiv [Preprint]*. doi: 10.1101/397729
- Rosa, M. J., Kilner, J., Blankenburg, F., Josephs, O., and Penny, W. (2010). Estimating the transfer function from neuronal activity to BOLD using simultaneous EEG-fMRI. *Neuroimage* 49, 1496–1509. doi: 10.1016/j.neuroimage.2009.09.011
- Scheeringa, R., Fries, P., Petersson, K.-M., Oostenveld, R., Grothe, I., Norris, D. G., et al. (2011). Neuronal dynamics underlying high- and low-frequency EEG oscillations contribute independently to the human BOLD signal. *Neuron* 69, 572–583. doi: 10.1016/j.neuron.2010.11.044
- Schwab, S., Koenig, T., Morishima, Y., Dierks, T., Federspiel, A., and Jann, K. (2015). Discovering frequency sensitive thalamic nuclei from EEG microstate informed resting state fMRI. *Neuroimage* 118, 368–375. doi: 10.1016/j.neuroimage.2015.06.001
- Sitaram, R., Ros, T., Stoekel, L., Haller, S., Scharnowski, F., Lewis-Peacock, J., et al. (2017). Closed-loop brain training: the science of neurofeedback. *Nat. Rev. Neurosci.* 18, 86–100. doi: 10.1038/nrn.2016.164
- Sulzer, J., Haller, S., Scharnowski, F., Weiskopf, N., Birbaumer, N., Blefari, M., et al. (2013). Real time fMRI neurofeedback: progress and challenges. *Neuroimage* 76, 386–399. doi: 10.1016/j.neuroimage.2013.03.033
- Tagliazucchi, E., and Laufs, H. (2015). Multimodal imaging of dynamic functional connectivity. *Front. Neurol.* 6:10. doi: 10.3389/fneur.2015.00010
- Thibault, R. T., MacPherson, A., Lifshitz, M., Roth, R. R., and Raz, A. (2018). Neurofeedback with fMRI: a critical systematic review. *Neuroimage* 172, 786–807. doi: 10.1016/j.neuroimage.2017.12.071
- Wang, T., Mantini, D., and Gillebert, C. R. (2017). The potential of real-time fMRI neurofeedback for stroke rehabilitation: a systematic review. *Cortex* 107, 148–165. doi: 10.1016/j.cortex.2017.09.006
- Zotef, V., Phillips, R., Yuan, H., Misaki, M., and Bodurka, J. (2014). Self-regulation of human brain activity using simultaneous real-time fMRI and EEG neurofeedback. *Neuroimage* 85, 985–995. doi: 10.1016/j.neuroimage.2013.04.126

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 Cury, Maurel, Gribonval and Barillot. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.