EEG-fMRI INTEGRATION: A CRITICAL REVIEW OF BIOPHYSICAL MODELING AND DATA ANALYSIS APPROACHES

M. J. ROSA*, J. DAUNIZEAU**,† and K. J. FRISTON*

*Wellcome Trust Centre for Neuroimaging, Institute of Neurology, University College London, United Kingdom
†Laboratory for Social and Neural Systems Research, Institute for Empirical Research in Economics, University of Zurich, Switzerland
‡j.daunizeau@fil.ion.ucl.ac.uk

Received 1 August 2010
Accepted 17 September 2010

The diverse nature of cerebral activity, as measured using neuroimaging techniques, has been recognised long ago. It seems obvious that using single modality recordings can be limited when it comes to capturing its complex nature. Thus, it has been argued that moving to a multimodal approach will allow neuroscientists to better understand the dynamics and structure of this activity. This means that integrating information from different techniques, such as electroencephalography (EEG) and the blood oxygenated level dependent (BOLD) signal recorded with functional magnetic resonance imaging (fMRI), represents an important methodological challenge. In this work, we review the work that has been done thus far to derive EEG/fMRI integration approaches. This leads us to inspect the conditions under which such an integration approach could work or fail, and to disclose the types of scientific questions one could (and could not) hope to answer with it.

Keywords: Neuroimaging; information fusion; functional segregation; functional integration; event-related; neurophysiology; data-driven; model-based; Bayesian analysis; model comparison.

1. Introduction

The realization of any cognitive, motor or sensory process rests on cerebral dynamics and creates order in the bioelectric and hemodynamic signals measured with EEG and fMRI, respectively. To detect and interpret the relevant features of these signals, one typically describes processes at their own temporal and spatial scales. The main sources of scalp EEG signals are postsynaptic cortical currents associated with large pyramidal neurons, which are oriented perpendicular to the cortical surface [115]. EEG (and magnetoencephalography, MEG) is well suited to studying the temporal dynamics of neuronal activity, since it provides direct measurement of this activity with millisecond precision. However, the scalp topology of measured electrical

‡Corresponding author.
Rosa, Daunizeau & Friston

potentials does not, without additional (prior) information, uniquely specify the location of underlying bioelectric activity. This issue is referred to as the ill-posed EEG/MEG inverse problem [19]. Conversely, even though BOLD-fMRI discloses complementary features of neuronal activity [108, 116], it is only an indirect measure thereof, through metabolism, oxygenation and blood flow, where these slow mechanisms provide temporally smoothed correlates of neuronal activity (\(\sim 1\) sec). Arteriolar control of blood flow however is spatially well-matched to regional increases in neural activity, hence the high localization power of fMRI [144].

Standard unimodal EEG (respectively, fMRI) data analysis relies on the specificity of a given bioelectric (resp., hemodynamic) feature of neuronal activity. Given the complementary limitations and advantages of these techniques, the vast majority of existing EEG-fMRI integration strategies attempt to enhance the spatial or temporal resolution of the combined EEG-fMRI dataset. But can we exploit their complementary nature to infer the underlying neuronal activity and its dynamics?

This paper focuses on the alternative approaches to integrating EEG and fMRI information, both from a biophysical modeling and signal analysis perspective. We aim at providing a comprehensive review of the state-of-the-art in this very active neuroimaging subdiscipline. We also identify promising research directions and highlight the types of scientific questions this approach can address.

This paper contains three sections. First we summarize the potential limitations to any EEG-fMRI integration approach, which leads us to comment on what has been coined the “neuro-vascular coupling”. We then review the data analysis approaches to EEG-fMRI integration. We classify them on the basis of their experimental motivation, which turned out to have practical consequences in terms of the methods of implementation (c.f. “symmetrical” versus “asymmetrical” approaches). We emphasize the increasing amount of biophysical modeling that is now embedded into statistical data analysis, with the aim of identifying properties of the neuro-vascular coupling.

Finally, we discuss the relevance of EEG/fMRI integration as a whole, in the light of the divergent aims, limitations and efficiency of its variants.

2. EEG-fMRI Integration: Limitations

In the following section, we will try to list the potential physiological and experimental confounds that constitute the main limitations of any EEG-fMRI integration procedure. These were found to be critical in understanding the reliability of the EEG/fMRI data analysis strategies, in terms of its expected sensitivity to these issues.

2.1. “Neurovascular” coupling and decoupling

Neurovascular coupling refers to the relationship between local neuronal activity and subsequent changes in cerebral blood flow (CBF). Despite the increasing amount of
literature in this field this issue is still under intense debate (for a recent review, see [125]). This is because neuronal information processing within a cortical unit and its relationship to neurophysiological processes can be described in many different ways and on many different scales. For example, membrane potentials versus spiking activity, excitatory versus inhibitory postsynaptic potentials or different types of receptor activation [134].

Most observed mismatches between EEG and fMRI can be interpreted as:
(i) a decoupling between the electrophysiological and the hemodynamic activity or
(ii) a signal detection failure (i.e., false positive/negative results in either modality). This distinction is important, firstly because signal detection failure should and can, in principle, be corrected. But also, it is the case that observing (local) decoupling might actually be very informative [42]. For example, in clinical applications (e.g., neuroimaging investigations of epilepsy), evidence for decoupling between electrophysiological and metabolic activity can be understood as a fingerprint of the pathology itself [129]. The question we address here is whether one can reliably disambiguate between a neurovascular decoupling and a signal detection failure.

Most information regarding neurovascular coupling comes from sophisticated invasive animal studies that combine metabolic/vascular measurements (i.e., fMRI or optical imaging — see [53]), with invasive multi-electrode data, such as local field potentials (LFPs) and multiunit activity (MUA) recordings\(^a\) [60, 70, 95, 96, 100, 101, 118, 121, 122, 132, 147]. These studies showed significant correlations between the time courses of hemodynamic and electrophysiological signals. More precisely, it seems to be that input or sub-threshold synaptic activity, rather than output spiking activity or energy demand [8, 91], drives the local vascular response. The important role of interplay between neurons and glial cells has also been emphasized in recent studies (for a recent review see [27]). These cells are good candidates for generating vasodilatory substances, such as epoxyeicosatrienoic acids and nitric oxide, in response to glutamate (and perhaps other neurotransmitters) release. However, the quantitative contribution of inhibitory neuronal activity to energy consumption and its role in the generation of BOLD still remains unclear [28, 70].

The above results also suggest that if BOLD and LFPs are closely related at the mesoscopic scale, then BOLD and EEG/MEG sources might have a tight spatial correspondence at the macroscopic scale. Indeed, this has been observed by a number of studies comparing the locations of EEG and fMRI sources, in a given subject and task. To mention but a few studies, a good concordance has been shown for primary sensorimotor [77] and visual [103] sources. Similar conclusions emerge when using more complex cognitive tasks: e.g., the motor response to visual stimulation [71], face perception [65] and decision making tasks [136]. Using a bilateral auditory

\(^a\)Logothetis et al. [96] used microelectrode arrays that have a sub-millimetric spatial resolution. By definition, MUA is the high-frequency (above 200 Hz) component of the recorded signals. It measures the spiking (output) activity of a population of cells. On the contrary, low frequency (below 200 Hz) LFPs correspond primarily to synchronized sub-threshold synaptic signals.
stimulation with ten subjects, Stippich et al. [135] found an average distance of 14 mm between the MEG dipole and the centre of the fMRI activation. This can be taken as a rule-of-thumb for good spatial concordance between EEG and fMRI. More recently, Lachaux et al. [81] have found, using intra-cranial EEG (icEEG) recordings in a cohort of epileptic patients undergoing a semantic decision task, a close spatial correspondence between regions of fMRI activation and those icEEG sites that showed high frequency power in the gamma band (approximately 40 Hz). This is important, since it suggest that the (macroscale) neuro-vascular coupling might be frequency-dependent (but see below).

However, a similar number of studies have shown significant differences between the regions implicated by EEG and fMRI, respectively. Gonzales-Andino et al. [57] list many of these case reports (mostly involving sensorimotor and auditory cortices). In the context of multimodal identification of epileptogenic foci, an average distance larger than 3 cm between EEG sources and fMRI results has been frequently reported [17, 23, 85, 89]. Note that these studies could potentially have been confounded by the ill-posed EEG inverse problem. Using more elaborate probabilistic EEG source reconstructions and fMRI cluster-by-cluster analysis, Grova et al. [60] could estimate the expected signal detection failure contribution to such a discrepancy. On a cohort of nine epileptic patients (62 fMRI clusters), Grova et al. [60] found approximately (i) 25% of concordant EEG sources and fMRI clusters, (ii) 25% of EEG/fMRI false positives/negatives and (iii) 50% of unconclusive discrepancy (which is thus an upper-bound on the local neuro-vascular decoupling). In terms of frequency-band specificity, Muthukumaraswamy and Singh [112] report significant stimulus-induced increases in EEG gamma band activity (in the visual domain) without corresponding changes in the BOLD signal.

In brief, despite concordant experimental evidence on the nature of the neuro-vascular coupling at the mesoscopic scale, macroscopic physiological processes might lead to local discrepancy or decoupling between EEG and fMRI activity. This is important, since this is a principled limitation to any multimodal EEG/fMRI integration procedure, which might be confounded by a number of potential decoupling mechanisms. For example, the neuronal population whose electrical activity is generating the EEG signal is not necessarily co-localized with the vascular tree that provides the blood supply to these neurons and gives rise to BOLD [20]. Similarly, in addition to pre- and post-synaptic electrochemical dynamics, a number of physiological processes also require energetic support; for example, neurotransmitter synthesis [117], glial cell metabolism [87], maintenance of the steady-state transmembrane potential [67], etc. These phenomena may cause hemodynamic BOLD changes, without EEG correlates ([6], see also [133] for the potential implication of neuromodulatory signals). This differential sensitivity to neuronal activity and energetics can also arise whenever hemodynamic activity is caused by non-synchronized electrophysiological activity or if the latter has a closed source configuration that is invisible to EEG. Conversely, if the electrophysiological activity is transient, it might not induce any significant (i.e., detectable) metabolic activity changes [116].
2.2. Experimental limitations

Another important potential source of bias in EEG-fMRI integration is experimental variability. In some situations, it might be necessary to acquire the EEG and fMRI data in separate sessions. In this case, habituation effects, variations in the stimulation paradigm, or any other differences between sessions might lead to differential activity of neuronal networks [57, 128, 148].

Simultaneous EEG/fMRI acquisition techniques have been developed specifically to address these issues [85, 126]. Nevertheless despite advances in simultaneous EEG-fMRI hardware and software, the signal-to-noise ratio (SNR) of these signals is still significantly lower than the corresponding unimodal paradigms. This is mainly due to reciprocal electromagnetic perturbations [78, 80]. For the EEG signal, this SNR degradation can be catastrophic: the most important artefacts in the raw data can completely mask the signal of interest. This is due to a complex combination of factors, including the MR field strength (and frequency) and orientation/positioning of the EEG recording equipment relative to the RF coil and the MR gradients. All these unavoidable artefacts manifest themselves as induced voltages that add linearly to the EEG signal and obscure the biological signal of interest. Although de-noising algorithms have been reasonably successful in gradient artefact correction [3, 52], the ballistocardiogram artefact remains an issue for most currently available softwares [22, 44]. For a comprehensive and up-to-date review of the pros and cons of simultaneous EEG-fMRI acquisition see Laufs [85].

3. EEG-fMRI Integration: Solutions

Since the main potential limitations of EEG-fMRI integration are well-established, many data analysts have argued that dedicated modeling and signal processing tools should be used to combine the advantages of these two techniques (e.g., [29, 62, 94, 109]).

Reconstructing the spatial deployment of current density from EEG measurements is an intrinsically ill-posed problem. On the other hand, estimating neuronal activity from the hemodynamic response is a difficult temporal deconvolution problem. Critically, the dual fitting of the bioelectric and hemodynamic responses does not necessarily circumvent the difficulties of the inverse problems that attend each modality. Then, what exactly do we expect to gain, or learn about neuronal activity,
from fusing EEG and fMRI? In this section we discuss the scientific questions, which motivate EEG-fMRI integration as well as the tools employed to address them.

3.1. Asymmetrical versus symmetrical approaches

Due to the above physiological and experimental confounds, electromagnetic and metabolic activity, as detected by EEG and fMRI, are not necessarily caused by the same underlying neural processes. Nevertheless, one can define “neuronal activity” operationally as the state of nodes in a network responding to specific events (e.g., cognitive, sensorimotor or spontaneous changes in brain activity) [50]. This allows one to consider event-related (ER) EEG and fMRI data as measures of “neuronal activity”, since the ER response is a reproducible EEG or fMRI signature that can be elicited systematically [49].

“Neuronal activity” ζ can then be decomposed into two overlapping sub-spaces, ζ_{EEG} and ζ_{fMRI} that correspond to the parts of ζ that contribute to EEG and fMRI signals, respectively. The intersection ζ_1 (see Fig. 1) defines a “common substrate” of neuronal activity. Conversely, ζ_2 (respectively ζ_3) denotes the subspace of neuronal activity detected by EEG (respectively fMRI) that does not contribute to fMRI (respectively EEG) measurements. This decomposition formalizes the apparent coupling-uncoupling between bioelectric and hemodynamic responses. Since no information about ζ_2 (respectively ζ_3) is available from the fMRI (respectively EEG), no multimodal procedure will provide a better characterization of this activity subspace than a unimodal EEG (respectively fMRI) analysis. Having said this, a multimodal approach should benefit from the complementary nature of EEG and fMRI information about the common subspace, ζ_1 [120].

Fig. 1. Formalization of EEG/fMRI coupling-uncoupling (a) and EEG/fMRI integration approaches (b) (adapted from [32] and [75]). Any multimodal information integration approach will be beneficial for inferring common neuronal states, ζ_1. This means that asymmetrical EEG-fMRI approaches systematically bias their estimate of ζ_1 by introducing information from ζ_{EEG} ((i): EEG to fMRI approaches, i.e., integration through prediction) or ζ_{fMRI} ((ii): fMRI to EEG approaches, i.e., integration through constraints). In contradistinction, symmetrical EEG-fMRI fusion approaches rely on a joint EEG-fMRI generative model, which allows the estimation of ζ_1 to be derived from an optimal balance between EEG and fMRI-derived information ((iii): integration through forward models).
Most EEG/fMRI integration approaches are motivated by the need for importing some reliable information from one modality into the analysis of another. Typically, so-called asymmetric approaches treat one modality as a cause or predictor of the other. In contradistinction, symmetrical approaches depend bilaterally on multi-modal data. These approaches were first proposed to circumvent certain specific limitations of the asymmetrical approaches, in terms of allowing some form of reciprocal cross-validation of EEG and fMRI datasets. In fact, they turned out to be most useful for answering a completely different scientific question, which relates to the neuro-vascular coupling. We will now review these approaches.

(i) EEG to fMRI approaches
The objective of these techniques is to localize, using fMRI, brain regions whose response is temporally correlated with a given EEG-defined event or feature. In other words, temporal information from the EEG signal is used as a constraint or predictor variable in the fMRI time-series model. This type of EEG-fMRI integration is implemented within a simultaneous EEG-fMRI acquisition paradigm. Typical applications include neuroimaging investigations of epilepsy and sleep, where there is no experimental control over ongoing (spontaneous) brain activity, but electro-physiological events can be easily detected on the scalp [110].

This pioneering work has been pursued largely by functional imaging groups focusing on pre-surgical planning for pharmaco-resistant epilepsy [1, 4, 7, 24, 59, 60, 79, 88, 142]. After artefact correction, the epileptiform activity is identified by an expert on EEG traces. These events are then convolved with a hemodynamic response function (HRF), and used as a regressor of interest in standard voxel-wise general linear model (GLM) analysis [16]. These studies have also revealed that the hemodynamic correlates different epileptic events. As an example, Archer et al. [5] showed that increased slow wave activity is associated with decreases in BOLD (deactivations), while spike and wave discharges have been shown to correlate with BOLD activations [63].

Other EEG to fMRI asymmetrical approaches have investigated the neuronal correlates of spontaneous cerebral activity occurring when the subject is not exposed to any extrinsic stimulation (i.e., at “rest”, hence the name of “resting-state” or “default mode” network; for a recent review see [82]). After simultaneous EEG-fMRI acquisition, spontaneous fluctuations of power in specific frequency bands are quantified in the EEG traces. Time-dependent power in each of these frequency bands is convolved with the HRF and used to form a regressor in the GLM for fMRI analysis [111]. This type of exploratory analysis tends to confirm observations made using invasive recordings looking at frequency-dependent correlates in fMRI [15, 21, 56, 98]. The main conclusion of this body of work is that reductions in ongoing scalp EEG alpha power correlate with increases in BOLD activity in human occipital cortex [55, 82, 84, 107].

Some multivariate extensions of this type of approach have also been proposed. These were motivated by the somewhat arbitrariness of the EEG feature selection
in the above techniques. It was thus argued that looking for multivariate patterns in the fMRI signals might increase the sensitivity of the analysis. Essentially, these extensions try to find a linear decomposition of fMRI data which covaries with the time, or time-frequency, decomposition of the EEG data [40, 41, 105].

Note that the observed correlations between EEG features and BOLD signals in both event-related and resting-state experimental designs are expressions of the neurovascular coupling. Identified using explorative (non-informative) statistical detection techniques, these correlations cannot inform us about the underlying biophysical mechanisms through which the neurovascular coupling emerges from. We discuss the last point in the subsection “Identification of neurovascular coupling models” above, which relies on a symmetrical treatment of both EEG and fMRI datasets.

\((ii)\) fMRI to EEG approaches

The aim of these techniques is to finesse the study of fast dynamics of neuronal activity as measured by EEG by using fMRI-derived spatial priors in the EEG source reconstruction problem. Again, this has been the subject of extensive literature in the past decade. The conceptual framework that dominates in this context rests on functional integration or coupling among sources [92, 94]. Going beyond functional specialization [46], evoked responses are understood as arising from an interacting network of connected “nodes” (the localized regions); these interactions are referred to as “arcs” or “edges” in graph theory. Interactions are expressed in the temporal dynamics of neuronal activity, since they shape the influence of one neuronal population on another. EEG and MEG are the best suited techniques for characterizing these connections since they are the only neuroimaging modalities whose temporal resolution is similar to that of the underlying neuronal processes. In this context, fMRI spatial priors are used to circumvent the unavoidable uncertainty about the number and deployment of nodes in the network generating EEG scalp data.

This approach can be divided into two classes, associated with the EEG source model employed: (i) the equivalent current dipole (ECD) model and (ii) the distributed source model (see [51, 73] for recent approaches). Dipolar fMRI to EEG approaches simply associate each fMRI focus with an ECD, whose position lies \textit{a priori} at the centre of the activation [148]. This type of \textit{a priori} constraint is hard, in the sense that the results of the fMRI analysis are not questioned (e.g., the number of active regions). In addition, since the ECD model does not accommodate the spatial extent of underlying active regions, it is difficult to assess the relevance of the fMRI constraint [39]. For example, it has been shown that many ECDs are required to model spatially extended regions correctly [130].

Distributed fMRI to EEG approaches rely on “weighted regularization techniques”. In these techniques, the model uses the fMRI activation as a prior on the spatial profile of cortically-distributed sources. This is done by penalizing EEG sources whose fMRI-derived activation probability is low. This approach has been shown to estimate the position and extent of underlying sources robustly, whenever
the fMRI-derived constraints are veridical [2]. However, when decoupling occurs, the EEG source reconstruction is strongly biased, which is why many variants of the fMRI-penalty term have been proposed [14, 61, 93]. Furthermore, these techniques require the tuning of a “hyperparameter” which regulates the weight of the penalty term. This hyperparameter is critical because it determines the balance between the effects of the fMRI prior and the EEG data fit on the estimation. Therefore, some authors [31, 64, 102] have proposed principled probabilistic (Bayesian) extensions to the standard regularization techniques, whereby this hyperparameter is estimated from the EEG data. Finally, the plausibility of the inverse reconstruction should depend on the relevance of the prior information. However, given solutions constrained or unconstrained by fMRI, which should be chosen? In Daunizeau [31], authors proposed a Bayesian model comparison method to decide whether one should use the fMRI constraint or not. This approach has been applied successfully to clinical epilepsy data. In Grova [60], it was used in combination with other probabilistic techniques to determine whether any observed EEG-fMRI discrepancy is due to decoupling or signal detection failure (see Sec. 2.1). In Daunizeau [33], it was used to identify the different subsets of fMRI clusters that respectively generated the spike and the wave in Generalized Spike-Wave (GSW) complexes measured on the scalp EEG. More precisely, it was found that the GSW involved (i) bilateral temporo-parietal structures that were common to both the spike and the wave components, and (ii) a prefrontal region that was only active during the spike (i.e., that was inhibited during the slow wave). Effectively, this technique relied upon the EEG to temporally resolve the activity of fMRI clusters. Authors also discuss the relevance of making statistical inferences at the level of clusters (versus voxels) when solving the EEG inverse problem.

(iii) Towards symmetrical EEG-fMRI approaches
As noted above, neurovascular decoupling, signal detection failures and experimental sources of brain activity variability are frequent. This has insidious consequences for asymmetrical EEG-fMRI approaches, since the relative reliability of EEG and fMRI is not evaluated. For instance, in Dale [29], the authors recognized that when fMRI was considered as the “truth” for spatial information, serious biases might occur in fMRI-regularized EEG source reconstruction, when the actual electrophysiological activity did not induce significant variations in the BOLD signal. Therefore, it has been observed that “integration of functional modalities into the solution of the neuro-electromagnetic inverse problem should be cautiously considered until a tighter coupling between BOLD effects and electrophysiological measurements could be established” [57].

This has motivated researchers to develop symmetrical data analysis approaches that would allow reciprocal cross-validation of EEG and fMRI responses, thereby

\[d\]This was coherent with the putative involvement of prefrontal cortex during absence seizures [119].
allowing for the coined term “information fusion”. From a statistical perspective, information fusion should be cast within a probabilistic framework, which allows one to formalize the propagation of both information and uncertainty from observations (the data) to unknown causes. Typically, this requires a so-called \textit{generative model} (or forward model) that specifies the (possibly uncertain) relationships between the data and what caused it. In this context, data analysis entails specifying an appropriate model, with a set of unknown parameters, and then looking for parameter distributions that explain the data. This is called \textit{model inversion} and involves extracting information from data by quantifying the uncertainty associated with a model of the system generating the data. If the model can generate multimodal data, its inversion corresponds to fusion (see [146] for a generic Bayesian interpretation of EEG-fMRI information fusion).

However, in practice, very few fusion approaches (i.e., data analysis techniques) have relied on realistic neurophysiologic models. This is because the complexity of realistic neuronal-metabolic-hemodynamic cascades (see Sec. 3.2) renders their inversion, given EEG and fMRI data, difficult. As an alternative, a few approaches have somehow shunted the complex neurovascular coupling/decoupling processes by effectively restricting themselves to estimating common properties exhibited by “active” areas contributing to both event-related EEG and fMRI measurements.

For example, Moosmann \textit{et al.} [106] propose using a joint decomposition of EEG and fMRI data that employs independent component analysis for group inferences on trial-to-trial consistently expressed spatiotemporal patterns. Other approaches have assumed that local BOLD signal is a simple low-frequency (linear) convolution of cortical current density (this was already suggested in [140]). This allows using Kalman filtering-like techniques to estimate spatio-temporal patterns of activity with supposedly augmented time and space resolution [37]. In [32], the authors further restricted common features of electrophysiological and metabolic responses to the spatial profile (i.e., position and extent) of the EEG and fMRI sources. Intracranial EEG measurements allowed authors to validate the results of the non-invasive fusion procedure on one epileptic patient with absence seizures, namely (i) the temporal precedence of the hemodynamic responses within the network does not match that of the bioelectric responses, and (ii) an excitatory bioelectric response can be co-localized with a negative hemodynamic response, i.e., a local deactivation (see Fig. 2).

The above approaches, any of which is (or can be) framed in Bayesian terms, represent the first attempts to fuse EEG and fMRI information in a rigorous way. Despite their somewhat heuristic aspect, these approaches are not confounded by the lack of detailed information about the neurovascular coupling. However, they are clearly suboptimal, in the sense that they do not include realistic models embedding likely neurovascular coupling scenarios. We will review these in the next section.
Fig. 2. Bayesian spatiotemporal event-related EEG-fMRI fusion approach (adapted from [32]). Symmetrical multimodal EEG-fMRI information fusion has been applied to the analysis of event-related bioelectric and hemodynamic responses. In this work, the spatial profile of common EEG-fMRI sources is introduced as an unknown hierarchical prior on cerebral activity. A bespoke Variational Bayesian (VB) learning scheme is derived to infer sources from a joint EEG-fMRI dataset. This yields an estimate of the common spatial profile, which embodies a trade-off between information harvested from the EEG and fMRI data. Furthermore, the spatial structure of the EEG-fMRI coupling-uncoupling is learned from the data. The figure shows: (left) the application of the fusion approach to EEG-fMRI recordings from a patient with epilepsy, to identify areas involved in the generation of epileptic spikes and (right) intracranial EEG data. In all graphics, colors code respectively: green = right occipital region, blue = left occipital region, red = right post-central gyrus and turquoise = left post-central gyrus. Both the measured intracranial EEG and estimated cortical sources seem to exhibit a similar temporal response: the epileptiform activity starts in the right occipital region, and spreads to left occipital and right post-central regions, the latter being temporally coherent. In contradistinction, the estimated hemodynamic responses did not conform to the same chronology. Since hemodynamic responses are driven partly by biophysical processes that are independent of the underlying neuronal activity (e.g., glial cell processes), one might be inclined to favor EEG-related analysis in any inference regarding causal relations within the active network.

3.2. Identification of neurovascular coupling models

Ideally, an optimal EEG-fMRI fusion approach would entail building a model that encompasses our knowledge about the link between bioelectric and hemodynamic activities and being able to invert that model, given the joint EEG-fMRI data [124]. Figure 3 summarizes the key components of a neurophysiologically and biophysically grounded generative model of both EEG and fMRI data, considering the state-of-the-art modeling on both the anatomo-functional and the neurovascular coupling.
Fig. 3. EEG-fMRI fusion and neuro-vascular coupling. Recent advances in understanding physiological mechanisms at different spatiotemporal scales have provided a framework within which to develop sophisticated biophysical models for integrating different imaging modalities. More precisely, evolution and observation equations encoding the relationship between bioelectric and hemodynamic mesostates can be motivated using both physiological and physical evidence. An increasing amount of experimental evidence shows that the dynamical behavior of a source depends on both its intrinsic and extrinsic connectivity. Dynamic causal models [48] based on neural mass models, may be an appropriate framework for models of bioelectric and metabolic activity in neuronal populations [73]. In this type of model, the EEG scalp data is assumed to be an instantaneous measure of the electrical potential generated by the activity of a subpopulation or neuronal mass (e.g., pyramidal cells), which has been propagated through the head tissues [19]. On the other hand, the fMRI data are modeled as a temporally smoothed response to mostly pre-synaptic neuronal activity [95] that results from a slow cascade of metabolic-hemodynamic events [9, 47, 123]. Note that this illustrative picture does not account for the cellular mechanisms that have been recently proven to be involved in the control of blood flow (see [125]). In brief, current evidence suggests that blood flow is mediated by multi-cellular signals of vasoactive substances, rather than by local energy demand in the brain. In other words, the cellular mechanisms responsible for consuming most of the energy, accounted in terms of ATP molecules, during activation (such as the $\text{Na}^+/\text{K}^+$ pumps responsible for restoring ionic gradients in the post-synaptic and axonal membrane [8]) do not directly control the vascular response [24]. This response seems to be triggered by a cascade of neuronal and astroglial events that can unfold into two different pathways: (a) a tonic pathway, in response to brief stimulation, where blood flow is directly induced by easily diffusible vasodilatory substances of neuronal origin. For instance, GABAergic interneurons in certain networks are capable of synthesizing nitric oxide (NO) [43], whilst certain pyramidal cells can produce prostaglandin E$_2$ (PGE$_2$) molecules [149]; (b) a phasic pathway, in response to sustained stimulation, where blood flow might depend on astrocytes as intermediaries between neurons and the vasculature [76, 151]. The release of neurotransmitters from pre-synaptic terminals, in particular glutamate, can activate the propagation of $\text{Ca}^{2+}$ waves in gap-junction connected astrocytes, which can cause the synthesis of vasoactive substances such as epoxyicosatrienoic acid, NO and/or PEG$_2$ [150].
This model involves many levels of description, including (i) the macroscale, i.e., the relationship between active brain regions, shaping the dynamics of local (mesoscale) neuronal populations [48]; (ii) the mesoscale, i.e., the interplay of local excitatory and inhibitory neuronal populations (see e.g., [73]) and (iii) the microscale, where neurovascular coupling is mediated through cellular mechanisms (see e.g. [125] for a recent review).

An outstanding modeling effort has focused on designing neurophysiologically grounded forward models from neuronal activity to hemodynamics. This has been done for the healthy brain [8, 9, 10, 11, 13, 123, 131, 139, 140] and in the context of neurological pathology (see e.g., [67, 97]).

Many of these theoretical frameworks make use of neural mass models, to characterize neuronal activity underlying the BOLD signal. However, due to the lack of consensus on the exact nature of neurovascular coupling, these models employ different coupling mechanisms. For instance, in Sotero [140] this link is based on the incoming spikes to a neuronal population, while in Babajani [10, 11, 13], the input to BOLD is proportional to changes in membrane potential of post-synaptic neurons. In Riera [123] the hemodynamic signals are driven by the total concentration of nitric oxide in the cortical unit. In brief, the main link between electrophysiological activity and energy consumption is invariably modeled through glucose metabolism. This assumes a monotonic mapping between excitatory activity and energy budget, and thus the BOLD signal (through blood flow). Note that in addition to the direct pathway from excitatory activity to blood flow, Sotero and Trujillo-Barreto [139] have introduced an indirect pathway from both excitatory and inhibitory activity to BOLD. This new component of the model accounts for the amount of glucose, and consequently oxygen, respectively consumed by excitatory and inhibitory activities.

The last part of this metabolic-hemodynamic (MHM) cascade (i.e., the mesoscopic relationship between blood flow and measured BOLD signal) is well-established, and forms the basis of the so-called “balloon model”. In brief, the balloon model relates the relative amount of deoxyhemoglobin in the tissue (which causes the increase in MRI signal) to coupled dynamics of local blood flow and blood volume (see [26] for seminal work and [47] for its extension).

Lastly, some authors have proposed interesting suggestions to account for the frequency-dependent side of neuro-vascular coupling. In Riera and Sumiyoshi [125], authors identify a candidate cellular mechanism that would explain the correlation that has been frequently reported between the amplitude modulation of gamma-band frequency power in the EEG with the BOLD signal. As an alternative, Kilner et al. [75] applied dimensional analysis to relate hemodynamic changes (as monotonically mapped from rates of energy dissipation) to the spectral profile of

These are well-established models that describe the dynamical interplay between membrane potentials and firing rates of the different neural populations (i.e., excitatory and inhibitory) composing a cortical unit (see e.g., [35,104]).
EEG activity. The analysis suggests that increases in BOLD signal should be associated with a shift in the EEG spectral profile to higher frequencies. It has been somehow overlooked such that this heuristic model makes a prediction that is completely at odds with previous perspectives on the relationship between electrophysiological rhythms and BOLD. This is because BOLD changes are not predicted to correlate with the amplitude modulation within frequency bands of interest in the EEG spectrum (see, e.g., [111]), but rather to the time-dependent frequency modulation in the EEG spectrum.

To our knowledge, such biophysical models for neurovascular coupling have only been inverted on single-region EEG-fMRI data thus far [12, 124]. This means that their complexity a priori prevents to use these models in order to perform EEG-fMRI information fusion, as it was originally hoped. In other words, there is no critical parameter in these models thus far that would be conjointly informed by both EEG and fMRI datasets. Nevertheless, these models have proven very useful when statistically comparing different biophysically plausible hypotheses for neurovascular coupling.

An important step in this direction has been recently undertaken by Sotero et al. [140]. These authors compare different variants of the MHM framework described above using BOLD data. The models under comparison, considered the hemodynamic response to be coupled to (i) excitatory activity, (ii) inhibitory activity and (iii) both activities together, as described above. Their results revealed that the first (and less certainly, the third) mechanism best explained the observed data.

In Rosa et al. [127], authors have embedded the biophysical model proposed by Riera et al. [123] within a Bayesian framework for joint EEG-fMRI data to assess whether BOLD is best explained by changes in input synaptic activity or by the output firing-rate of a cortical unit. The results for one subject exposed to a flickering checkerboard suggest that BOLD is best explained by the input synaptic activity model in the visual cortex and interval of stimulus frequencies analyzed (4 to 16 Hz).

Also, the predictions of Kilner et al. [75] have been tested by Rosa et al. [127]. The authors used simultaneous EEG-fMRI data in humans, with a visual flicker stimulation task, to compare the BOLD signal in the visual cortex with the qualitatively different features of the EEG frequency spectrum, namely: (i) amplitude modulations in different frequency-bands of the EEG spectrum (including alpha and gamma power) and (ii) frequency modulation in the EEG spectrum. These features were convolved with a canonical hemodynamic response function and used as regressors in a standard GLM analysis of fMRI data. As predicted by Kilner et al. [75], the frequency modulation significantly explained more BOLD activity than the total.

\[^{4}\text{Conductance-based models of neural masses dynamics (see e.g., [104]) actually predict that increase in synaptic input induces increases in rate constants of synaptic transmission. This will ultimately shifts the frequency spectrum towards higher frequencies.}\]
time-varying spectral power and any linear combination of frequency-band amplitude modulations.

4. Discussion

We have reviewed the advances and issues attending EEG-fMRI information fusion. We have emphasized the importance of neurovascular coupling for generative models, which are the key to any balanced fusion procedure. This issue underlies many of the challenges for fusion and has been a source of much debate: “it is far from trivial to suppose, for instance, that a statistically significant Z-score in the left inferior frontal gyrus and a large left anterior negativity at 200 ms after stimulus presentation correspond to the same thing” [66]. In this paper, we tried to identify features of cerebral activity that could form the basis of models of electromagnetic and hemodynamic markers of neuronal activity and are required for EEG-fMRI multimodal fusion. We have also emphasized the importance of developing statistical methods for model inversion, given EEG and fMRI signals.

Optimal EEG-fMRI fusion has to rely on models of neurovascular coupling to exploit the complementarities of EEG and fMRI information. There is a subtle balance between the plausibility of the assumptions and the efficiency of any model to make precise inferences. The tighter the prior belief regarding the underlying causes of our observations, the more precise our interpretations of the data. However, these inferences become increasingly constrained by our priors and the space of models examined. Therefore, such a “model-based” procedure will suffer from the usual limitation of modeling: refutability. Whether the assumptions of the model are satisfied in a given experimental context will remain a question in itself.

Gathering the knowledge and know-how necessary for such EEG-fMRI fusion has proven to be a challenging exercise for the neuroimaging community. However, very few rational criticisms have questioned the intrinsic motivation of EEG-fMRI fusion. In short, what is the type of scientific question (apart from established diagnostic application, e.g., epilepsy) that really requires EEG-fMRI fusion?

Although the motivation for developing symmetric fusion tools emerges from the limitations of asymmetric approaches, it has perhaps mislead scientists into thinking that symmetrical fusion would be the optimal EEG-fMRI integration method, and hence would eventually replace asymmetrical approaches. In fact, most well-motivated questions targeted by the asymmetric approaches can, in practice, be readily answered with the adequate asymmetric tools. More precisely, using user-defined EEG events as a predictor of fMRI, BOLD signals have already proven very useful, whenever no experimental control over brain activity is available. This is because the simplicity of this type of approach makes it very reliable and robust against most established confounds to EEG-fMRI integration.

One has to admit though, that despite more than a decade of methodological research on fMRI to EEG asymmetrical approaches, very few established positive results has emerged thus far from introducing fMRI spatial priors into the EEG
source reconstruction problem. Recently, it has even been argued that state-of-the-art probabilistic source reconstruction might not have anything to gain, in theory, from sophisticated fMRI spatial information (see [64] for more details). Perhaps it is the time to reconsider, in the light of the accumulated evidence about its cost-benefit trade-off, the motivation for such approaches and their symmetric variants.\textsuperscript{8} Having said this, symmetric approaches can be seen as a complementary technique with its own research agenda. For instance, the generative models originally developed for asymmetrical fusion can be used to answer qualitatively different scientific questions. As has been argued in Riera et al. [125]: “In the light of the recent findings about blood supplying mechanisms during brain oscillations, we have to regrettably admit that further experiments need yet to be carried out to have a better understanding of the neurovascular coupling in the cerebral cortex of mammals”. In this context, the symmetrical fusion approach might prove useful as a statistical framework aimed at identifying mechanistic processes leading to observed neurovascular coupling and decoupling.

Acknowledgment

This work was supported by the Foundation for Science and Technology, Portugal (M. J. R.), the Wellcome Trust (K. J. F.) and the NEUROCHOICE project of the Swiss Systems Biology initiative SystemsX.ch (JD).

References


\textsuperscript{8}But see, e.g., [31,36], for recent numerical evidence of the relative superiority of symmetrical approaches in solving the EEG inverse problem.


